BIT ALLOCATION METHOD FOR AC-3 ENCODER

Chi-Min Liu, Member, IEEE, Szu-Wei Lee and Wen-Chieh Lee
Department of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, 30050, Taiwan
E-Mail: cmliu@csie.nctu.edu.tw

Abstract — **Dolby AC-3 is an audio coder selected in the United State high definition television (HDTV) standards and widely adopted for the audio codec in DVD films. For audio coder, the complexity of the audio encoders is always higher than that of the decoders. A main module leading to the higher complexity is the bit allocation process. The objective of the bit allocation is to allocate restricted bits to encoded information. Since that AC-3 coder has adopted a floating point representation which the magnitude of each spectral line is represented with an exponent and a mantissa, the bit allocation process not only has to decide the suitable quantization method for the mantissa similar to the process in other coding standards such as MPEG1, but also the exponent strategies and hence the parameters in psychoacoustic models. This coding process has been referred to as a hybrid coding [5] and has led to immense complexity for an encoder. This paper proposes an efficient method for the bit allocation process.**

I. INTRODUCTION

THE Dolby AC-3 [1] is currently the audio standards for the United States Grand Alliance HDTV system audio coding standard and widely adopted for DVD films. The Dolby AC-3 encoding process can be illustrated in Fig. 1. The audio sequences are transformed into a domain referred to as spectral domain. Each spectral line in the spectral domain is represented as floating point consisting of exponent and mantissa. The exponents are encoded by suitable coding strategy and fed into psychoacoustic model. The psychoacoustic model calculates the perceptual resolution according to the encoded exponents and the proper perceptual parameters. Finally, the information of the perceptual resolution and the available bits are used to decide the appropriate quantization manner to quantize the mantissa of the spectral lines under restricted bits. The bit allocation process is to determine the suitable exponent coding strategies, the proper perceptual parameters, and the appropriate quantization manners in the encoding process with restricted bit number.

Consider the exponent coding process in Fig. 1. The difficulties of the exponent coding are on the efficient search for the large number of strategies and the criterion deciding the best strategies. In AC-3, it provides four exponent coding strategies for each audio block referred to as D15, D25, D45 and REUSE [1]. Except for the first audio block, the remaining audio blocks can use the REUSE coding strategy. Hence, there are 3*4*4*4*4*4=3072 possible strategies for the six blocks in a frame. The search space is large and there needs an efficient search method. Furthermore, even an exhaustive search is executed there needs a criterion for selecting the strategies. Since that there is no analytic relation between the final audio quality and these exponent strategies, an optimum solution is to follow an analysis-by-synthesis method. That is, all the candidate strategies for exponent coding are tried and hence provide the necessary information for the remaining encoding process. Then, the optimal coding strategy is selected from the associated coded or synthesis audio having the best quality. However, the complexity for the process is again too high to be practical. In this paper, we propose a selection criterion and an efficient search method for exponent strategies.

Consider next the psychoacoustic model in Fig. 1. The psychoacoustic model calculates the perceptual resolution according to the encoded exponents and perceptual parameters. The difficulty of the process is the way to adapt the perceptual parameters to the current audio content. The AC-3 standard draft [1] suggests that the perceptual parameters are fixed to simplify the complexity of bit allocation process. However, for low bit rate system such as that below 64 Kbit/s for a channel, these parameters are quite critical for audio quality. This paper presents the method to adapt the parameters to the audio contents.

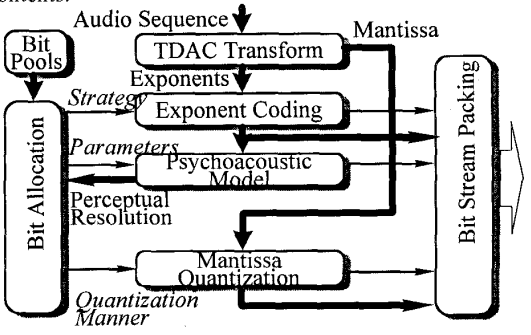


Fig. 1 Dolby AC-3 encoding process.

Manuscript received June 17, 1998

The third difficulty is on the mantissa quantization. The major problem arising from the mantissa quantization process is on the efficient search for the value of quantization parameter provided by the AC-3 to fit the available bits. In AC-3, the mantissa quantization process is to quantize the mantissa of each spectral line according to the perceptual resolution and the values of quantization parameter. There are 1024 selections for the parameter in AC-3 and there needs a vehicle searching for the optimal value fitting the restricted bits. The problem is that there is no direct relation among the values of the parameter, the perceptual resolution, and the available bits. That is, there is no way finding the suitable quantization value directly from the perceptual resolution and the available bits. This paper proposes the efficient algorithm for searching the optimal value of the quantization parameter in AC-3.

The rest of this paper is organized as follows: Section II illustrates the efficient searching algorithm and selection criteria for exponent coding process. Section III provides the method to adapt the perceptual parameters to current audio content. Section IV gives the efficient searching algorithm for the quantization parameter. Section V shows the experiment results. Section VI gives a brief conclusion.

II. EXPONENT CODING METHOD

In AC-3, each spectral line is represented by an exponent and a mantissa. All the exponents are coded by the exponent coding process. The coding strategies available in AC-3 are referred to as D15, D25, D45 and REUSE. The coding strategy D15 provides the finest frequency resolution and hence requires a large number of bits. On the contrary, the strategy D45 gives the coarsest frequency resolution and hence consumes a less number of bits. Especially, the strategy REUSE indicates that the exponents of current block are the same as the previous block and hence there is no bit requirement for the exponent of current audio block.

As described in last section, two difficulties on the exponent coding are the large combinational space of the exponent coding strategies and the selection criterion. This section proposes a selection criterion and the associated efficient search method for the exponent strategies. The block diagram of the exponent coding process is illustrated in Fig. 2. The process consists of three steps. First, the available bits of the exponents are determined from the current bit rate. A ratio of 20% of the overall bit rate has been adopted to select the exponent strategies. The ratio has been determined through immense experiments. On the ratio, the second step is to list all the exponent strategies which consume a bit number less than the available bits. For music sequence adopting a fixed frame rate, the candidates are fixed and will not vary with frames. Finally, all the candidates are used to encode the exponents. On all the associated encoded exponents, the strategy which minimize the error criterion is selected. The error criterion is listed as follows:

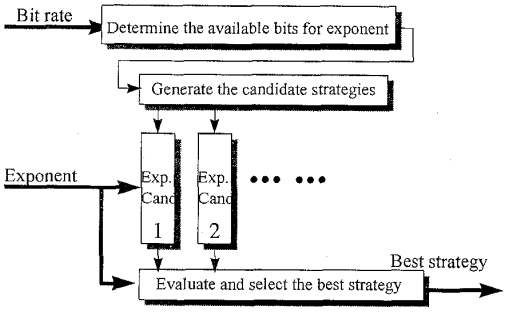


Fig. 2 Block diagram of exponent coding process

$$E=\sum_{k=0}^{5}\sum_{b=0}^{255}\left[\exp_o(k,b)-\exp(k,b)\right] \tag{1}$$

where $exp_o(k,b)$ is the original exponent of block k and spectral bin b before encoding, and $exp(k,b)$ is the corresponding exponent encoded by a candidate strategy. In a frame defined by AC-3, there are six blocks and 256 spectral bins in a block. The criterion is reasonable in the sense that the formula indicates the error between the coded and the original exponents. The overall process can find the best fitted exponent strategy under the bit rate constraint.

III. PERCEPTUAL PARAMETERS

In audio coding, the psychoacoustic model gives the information on the perceptual resolution of audio signals. The perceptual resolution is the key information to compress an audio sequence without losing audio quality. The perceptual resolution is calculated from the masking effects of signals. Masking effects demonstrate the perceptual resolution of spectral lines when various types of audio contents exist. Especially, two types of masking effects are considered in audio compression. The first type is the masking effects from the existing of narrow band noise. The other is the masking effects from tonal signals. The two types of masking result in different masking effects and hence different perceptual resolution. This section presents a method to detect the two types of masking effects from the audio exponents. The parameters in the psychoacoustic model of AC-3 are determined according to the detection results.

A. Parameters in the Psychoacoustic Model of AC-3

The psychoacoustic model in AC-3 calculates the masking threshold from the following three steps: First, the encoded exponents are transformed into power spectral density (PSD) through

$$\text{psd(k,b)} = 3072 - \text{exp (k,b)}*128 \tag{2}$$

Then, the bins of the PSD are combined into bands according to the perceptual bandwidth. At low frequencies, the band size is 1, and at high frequencies the band size is 16. Third, the masking threshold of a band is computed by summing the masking effects from other bands. The

masking effect of a band from the signals in other band is illustrated through the spreading function in Fig. 3. For a signal existing at band i with energy E, the spreading function indicates the resultant masking threshold of the bands above band i. The spreading function is approximated by two curves: a fast decaying curve and a slow decaying curve. The fast gain is the signal-to-mask ratio, that is the ratio between the energy of the masking sound and the masking threshold in band i. The gain can be chosen according to the audio contents. In AC-3 standard draft, the value is fixed and selected as -30dB. However, in [7] the voluminous experiments demonstrate that the corresponding parameter is selected from -10dB to -20dB for tonal signal and -5dB to -10dB for narrow band noise. This section shows the method to determine the values of the fast gain.

B. Selection on the Fast Gain

Due to the limit on AC-3, the fast gain is transmitted once per audio block rather than for each spectral bin.

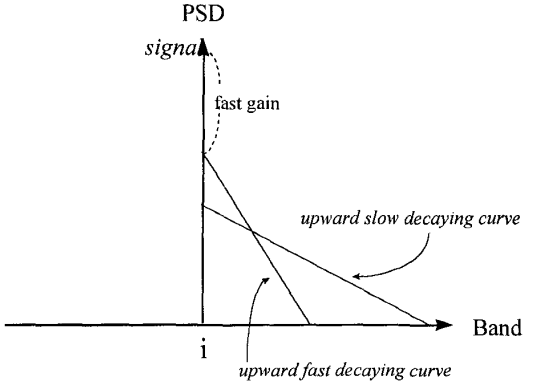


Fig. 3 Modeling spreading function

Hence, a simple method for the parameter selection is that the parameters are adopted according to the information of audio block rather than single spectral line. That is, if the audio block is tone-like, the conservative value -30dB is retained. On the contrary, if the audio block is noise-like, the value -10dB is selected. However, the difficulty is the tonality measure for an audio block.

Two properties of the tonal signals are the spectral peaks and the spectral similarity between blocks. Since that the exponent strategies decided in (1) has considered both the spectral and temporal similarity, the tonality can be selected directly through the exponent strategies. Since that the tonal signal has higher spectral peak than other frequency components near it, if the audio block is tone-like, it implies that the exponents of the block have to be encoded through the highest spectral resolution strategy, that is the D15 mode. In addition, since that the tonal signal can be determined from the likeness of a spectrum band through several audio blocks, those blocks using REUSE are also tone-like. Furthermore, if the exponent strategy is D45, the audio block is considered to be a noise-like block.

Now the information of the exponent coding process is used to decide the psychoacoustic parameters. As mentioned above, the perceptual parameters are transmitted once per audio block rather than per spectral bin. Hence, the conservative value of the fast gain is retained. If the result of the exponent coding process gives that the block is in the D15 mode and the following blocks are in the REUSE mode, the block is tone-like and the fast gain is selected as -24dB. If the exponent strategy is D45, the associated block is noise-like and the fast gain is selected as -12dB. For the D25 mode, the average value -18dB is adopted.

IV. MANTISSA QUANTIZATION

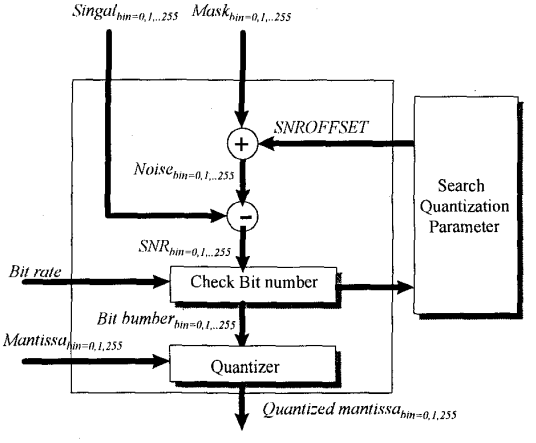


Fig. 4 Flowchart of mantissa quantization

Consider the flowchart of mantissa quantization shown in Fig. 4. The mantissa quantization retrieves the masking threshold $Mask_{bin}$ from the psychoacoustic model. The masking curve is added with the parameters *SNROFFSET* to produce the noise curve. The signal to noise ratio can then obtained for each spectral bin. The bit number of the mantissa can then be determined from the ratio of the signals and noise. In the flowchart, the problem is on the selection of the optimal value of *SNROFFSET*. There are 1024 selections for *SNROFFSET* and there needs a vehicle searching for the optimal value to fit the available bits. This section considers the efficient searching algorithm for the values of *SNROFFSET*.

Since that there are 1024 selections of *SNROFFSET*, therefore, at least ten iterations are needed to find the optimal quantization parameter if the binary searching algorithm is performed. To further reduce the complexity, we propose a new searching algorithm. Our experiments demonstrated that the new algorithm is more efficient than the binary searching algorithm.

The proposed searching algorithm consists of two phases: (1) iterative phase and (2) searching phase. The block

diagram of the quantization parameter search is shown in Fig. 5. Initially, the proposed searching algorithm is in the iteration phase. In this phase, the quantization parameter, *SNROFFSET*, is predicted in each iteration. The predictive equation is given as follows:

$$SNROFFSET_i = SNROFFSET_{i-1} + \frac{R_{i-1} - R_{av}}{nBIN_{i-1}} \times \mu \tag{3}$$

where $SNROFFSET_i$ is the quantization parameter at iteration i, $nBIN_i$ is the number of spectral lines with positive bit number and R_i is the allocated bit number in the *i-th* iteration. R_{av} is the current available bit number and μ is step size. In our experiments, we choose the step size μ as 128.

In AC-3, the psychoacoustic model is performed on the PSD domain [5]. The PSD is derived by the encoded exponent expressed in (2). Hence, the PSD-decibel has the following relation:

$$128 \text{ units PSD} = 6 \text{ dB} \tag{4}$$

Since that additional one bit resolution increases the signal-to-noise ratio by 6dB for uniform quantizers, the signal-to-noise ratio is increased by 128 units PSD. Therefore, the step size μ is chosen as 128. In the low bit rate system, the symmetric quantizers are often used. In the condition, the step size μ has to be decreased to avoid over-prediction.

The iteration terminates when the following two conditions are met: (a) $R_i \leqq R_{av}$, $R_{i-1} > R_{av}$ or (b) $R_{i-1} \leqq R_{av}$, $R_i > R_{av}$. The search phase then searches the optimal value from the range between $SNROFFSET_i$ and $SNROFFSET_{i-1}$ by the binary search algorithm. Since that the optimal quantization parameter is bounded by $SNROFFSET_i$ and $SNROFFSET_{i-1}$ which is the sub-region of 0 to 1024, the binary searching algorithm takes less than ten iterations to find the optimal value of *SNROFFSET*.

V. EXPERIMENT RESULTS

This section considers the efficiency of the encoding algorithm. In the following experiments, each audio channel is encoded at the bit rate of 64 Kbit/s with sampling frequency of 44.1 KHz. The bit number of the exponents is 435 in one frame. The exponents coding strategies which consume less than 20% frame bit rate are listed in Table 2. The three audio sequences illustrated in Fig. 6 can provide a typical example for the experiments. The decided exponent coding strategy also decides the tonality of the block. Fig. 6 illustrates three examples of the tonality decision. The decisions are quite consistent with audio contents.

For the experiments on searching the values of the *SNROFFSET*, a total of ten 20 sec stereo audio songs

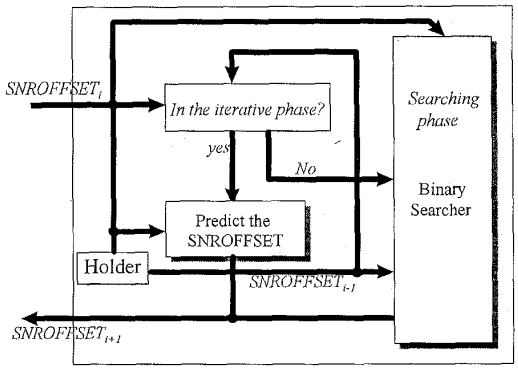


Fig. 5 Block diagram of the quantization parameter search

including vocal, symphony, piano and so on are taken as the materials. Table 1 lists the average iteration numbers per frame of mantissa quantization for above materials. The iteration numbers demonstrate that the proposed method provides an iteration number much lower than ten which is the iteration counts of binary searches for 1024 values.

VI. CONCLUDING REMARKS

In AC-3 encoder, the bit allocation is quite computation intensive and there is no article analyzing the problem. This paper has analyzed the problem and presented efficient methods of the bit allocation through three aspects: (1) the exponent coding, (2) the psychoacoustic model, and (3) the mantissa quantization. For the exponent coding, the problem is on the selection criterion and the efficient

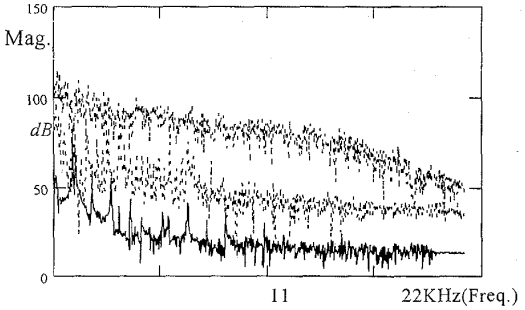


Fig. 6 Frequency responses of three typical audio sequences, where the lowest curve is encoded by D15, the middle

search method for the exponent strategies. For the psychoacoustic models, the difficulty is on the selection of the perceptual parameters adapting to audio contents. For the mantissa quantization, the issue is on the efficient search methods for the optiaml value of the quantization parameter. On the three aspects, this paper has presented methods to achieve efficient bit allocation.

Table 1 Average iteration counts per frame

source	butter	tsai	dance	flute	heart1	memory	second	march	Russian	Chinese
count	5.18	5.81	4.94	4.91	6.02	5.78	6.09	5.40	5.86	4.25

Table 2 Candidates of exponent coding strategies

(1)	[D15,REUSE,REUSE,REUSE,REUSE,REUSE]
(2)	[D25,REUSE, REUSE,D25,REUSE,REUSE]
(3)	[D25,REUSE,REUSE,D45, REUSE,D45]
(4)	[D25,REUSE,D45,REUSE,D45, REUSE]
(5)	[D45,D45,REUSE,D45, REUSE,D45]

REFERENCES

[1] J. C. McKinney and R. Hopkins, "Digital audio compression standard (AC-3)," *Advanced television system committee*, Dec. 1995.

[2] S. Shlien, "Guild to MPEG-1 audio standard," *IEEE Trans. Broadcasting*, vol. 40, no. 4, pp. 206-218, Dec. 1994.

[3] C. A. Serantes, A. S. Pena and N. G. Prelcic ,"A fast noise-scaling algorithm for uniform quantization in audio coding schemes," *IEEE Conf. Acoustic, Speech and Signal Processing*, pp. 339-342.1997.

[4] Y. Mahieux and J. P. Petit, "Transform coding of audio signals at 64kbits/s," *IEEE Conf. Acoustic, Speech and Signal Processing*, pp.405.2.1-405.2.4.1990.

[5] C. C. Todd, G. A. Davidson, M. F. Davis, L. D. Fielder, B. D. Link and S. Vernon ,"AC-3 flexible perceptual coding for audio ...," *AES 96th Convention*, Feb. 1994.

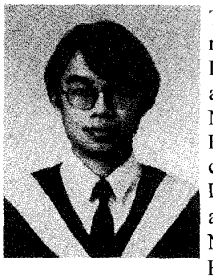
[6] G. A. Davidson, L. D. Fielder and B. D. Link, "Parametric bit allocation in a perceptual Audio coder," *AES 97th Convention*, Nov. 10-13 1994.

[7] J. B. Allen, "Speech and hearing in communication," The Acoustical Society of America by the American Institute of Physics.

Chi-Min Liu received the B.S. degree in electrical engineering from Tatung Institute of Technology, Taiwan, R.O.C. in 1985, and the M.S. degree and Ph. D. degree in electronics from National Chiao Tung University, Hsinchu, Taiwan, in 1987 and 1991, respectively.

He is currently an Associate Professor of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include video/audio compression, speech recognition, radar processing, and application-specific VLSI architecture design.

Szu-Wei Lee was born in Taichung, Taiwan, on March 11, 1974. He received the B. S. degree from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1996. He is currently a Ph. D. Student of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests are audio compression and real-time software development.



Wen-Chieh Lee was born in Toayuan, Taiwan in Oct. 1972. He received the B. S. degree from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1995. He is currently a Ph. D. Candidate of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests are audio compression and real-time computer architecture.