A FAST AUDIO BIT ALLOCATION TECHNIQUE BASED ON A LINEAR R-D MODEL

Chun-Yi Lee, Yao-Chun Fang, Hsiao-Chiang Chuang, Chung-Neng Wang, and Tihao Chiang
Dept. of Electronics Engineering, National Chiao-Tung University
1001 Ta Hsueh Rd., Hsinchu, 30010, Taiwan, R.O.C.
Email: tchiang@cc.nctu.edu.tw

ABSTRACT

The MPEG-1 Layer III audio coding is most widespread digital audio format on the Internet. Its encoder is more complicated than the other layers in the MPEG-1 audio specification, where the bit allocation constitutes significant part of the total computational load. This paper describes a fast bit allocation technique using a novel linear model to compute the bitrate and the global gain of the inner loop (rate control loop) for the classic MP3 bit allocation problem. Estimated quantization noise is approximated with a binomial expansion for the adjustment of the scalefactors. This procedure reduces the outer loop iterations (distortion control loop) to one loop. Our bit allocation technique is about five times faster compared to the well-known LAME MP3 encoder. Based on the objective quality measure, total number of distorted bands, and average total noise-to-masking ratio (NMR), our experimental results show that there is significant improvement in speedup of the encoding process without noticeable audio quality degradation.

1. INTRODUCTION

MPEG-1 audio specifications consists of three coding algorithms called layers I, II, and III [1]. Among all layers, Layer III provides the highest quality around 128 kbps and is the most popular format to deliver the CD quality music on the Internet. MPEG-1 Layer III (MP3) audio coding is an asymmetric codec where the encoder has much higher complexity than the decoder because of the bit allocation module. Thus, it is changeling that the highest computational burden can be reduced to provide an efficient implementation for portable MP3 devices familiar in consumer market.

Fig. 1 illustrates the block diagram of an MP3 encoding process. Using the filter bank, the input audio signals are transformed into scalefactor bands in a frame-by-frame manner. For each frame with a given target bit budget, the bit allocation process distributes bits to code each scalefactor band. The bit allocation process is based on information from the psychoacoustics model to provide the best listening quality for a given bit budget. In practice, the bit allocation needs to find the optimum gain and scalefactors for each block such that the perceptual distortion is least according to the psychoacoustics model for a target bit rate. Thus, the issue of an MP3 encoding process is on how to select the gain and scalefactors for each block within the audio sequence.

To derive the optimal gain and scalefactors for each block, typical bit allocation methods including the MPEG Audio specification [1] and LAME algorithm [2], [3] employ two nested iteration loops. As shown in Fig. 2, the two nested iteration loops are the outer loop and the inner loop. The outer

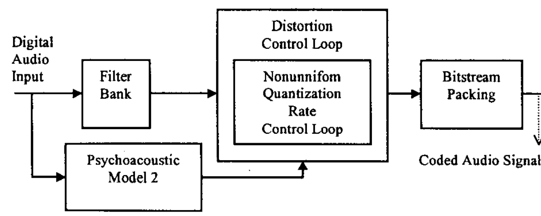


Fig. 1. The block diagram of MPEG-1/Audio Layer III encoder.

Table 1 The major percentage of the function times from the profile in the MP3 encoding

Function name	Percentage of function time
Psychoacoustic Model	13.3%
Filter bank	11.4%
Bit allocation	38.2%
Others	37.1%

loop is the distortion control loop that copes with the quantization noise of spectral signals and derives the optimal scalefactors. The inner loop is the rate control loop that computes the optimal global gain to fit the target bit budget. Both the MPEG committee and LAME algorithms use step-by-step search to find the bits satisfying the given bitrate. The iterative approaches cause high computational costs, which is consistent with the observations in Table 1. Table 1 shows an example of the profiling in the MP3 encoding, among which the bit allocation takes almost 38.2% of the overall execution time. The iteration nature makes it difficult to obtain a low complexity MP3 encoder. Thus, the remaining issue will be focused on how to efficiently allocate bits without using iterative approaches and still maintain the listening quality.

Several fast algorithms [4] and [5] are proposed to speed up the derivation of the optimal gain and scalefactors in coding the audio signals. For estimating the quantization noise, Liu et al [5] proposed a noise shaping approach to reduce the complexity. With the estimated noise and psychoacoustics model, the outer iteration loop selects the scalefactors that make the noise shape parallel to the masking thresholds from psychoacoustics model. It is advantageous to use single loop to estimate the error. However, the human ears are sensitive to the audio signals of low frequency. The outer iteration loop using parallel noise shape causes frequent usage of the scalefactors, which costs

Manuscript received June 24, 2002 0098 3063/00 $10.00 © 2002 IEEE

more bits for transmitting the scalefactors and may degrade the audio quality. As for the rate control loop, both approaches [4], [5] adopt binary search technique to decrease the iterations in getting the bits fit with the bit budget for each frame. The binary search is much efficient but still has a complexity of $O(\alpha \log_2 \beta)$ for the final global gain, where α means the overall complexity for each inner loop and β indicates the dynamic range of the global gain. Thus, high speed and better noise shaping are essential issues for the optimal bit allocation in MP3 encoding process.

To address these issues, we improve the noise estimation method as referred to in [5] and propose a new rate control algorithm based on a linear model. The quantization noise is estimated with the first order approximation of a binomial expression. With estimated noise and Noise-to-Masking ratio (NMR), the scalefactor can be adjusted to minimize the actual distortion, especially for the low frequency bands. For the rate control, a fast approach of complexity $O(\alpha)$ is proposed based on linear relationship between scalefactor, bitrate, and global gain. Our experimental results show that the estimated noise energy through the propose approach is close to the real noise energy. In terms of computational speed, the proposed fast rate control algorithm just needs less than 2.5 times for encoding audio signals at 128kbps and is more robust to the variation of bit rates.

In Section 2, we will introduce the basic of MPEG/Audio layer III encoding, and point out why we need a fast bit allocation algorithm. In Section 3, a new fast bit allocation model is proposed to replace the ISO/MPEG iterative bit allocation algorithm. In Section 4, some experimental results are provided for the comparisons of LAME and proposed algorithm. The conclusions and future works are given in Section 5.

2. MPEG/AUDIO BIT ALLOCATION ALGORITHM

Fig. 1 illustrates the block diagram of an MP3 encoding process. Among all, the bit allocation algorithm is the most critical for facilitating the listening quality of the reconstructed audio sequences. Additionally, the optimal bit allocation is based on the two important parameters including the global gain and scalefactors for each block such that the perceptual distortion is least according to the psychoacoustics model for a target bit rate.

As summarized in Table 2, in order to derive these two parameters the two most well known techniques are the MPEG committee [1] and the LAME approaches [2]-[3], where both employ two nested iteration loops. However, the derivation of the optimal gain and scalefactors is much time consuming using the iterative approaches [1]-[3]. Thus, deterministic fast algorithms are useful to improve the speed in deriving the optimal gain and scalefactors in coding the audio signals.

2.1 THE ITERATION LOOPS (DISTORTON AND RATE CONTROL LOOP)

The bit allocation algorithm adopted at ISO/MPEG specification is based on an iterative approach [1]. The ISO/MPEG bit allocation processes are illustrated in Fig. 2. To have an optimal bit allocation, two nested loops are employed including the outer loop (distortion control loop) and inner loop (rate control loop).

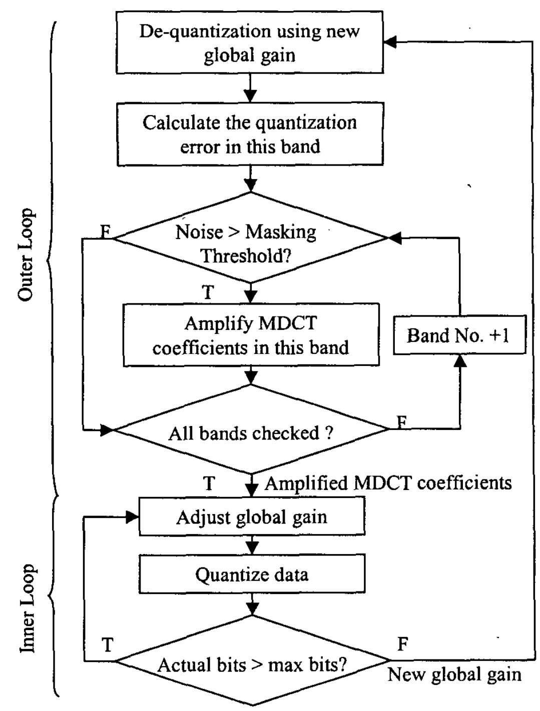


Fig. 2. The block diagram of MPEG-1/Audio layer III bit allocation iteration loops

The outer iteration loop controls the quantization noise, which comes from the quantization of the spectral signals within the inner iteration loop. The noise spectrum is computed by multiplying the values within the scalefactor bands with the actual scalefactors before quantization. After quantization, the calculation of the quantization noise is processed band-by-band iteratively. If the noise exceeds the specified threshold, which is the maximally allowed distortion that is derived based on the masking thresholds quantified by psychoacoustics model, the spectral values of the scalefactor bands are amplified by increasing the scalefactor by one [1].

In the inner loop, the encoder iteratively employs the non-uniform quantization, which has three major steps including the quantization of the spectral values, the calculation of actual number of bits using Huffman tables, and the computation of the resulting noise. By applying these three steps iteratively for each frame, the bit allocation algorithm provides the actual number of bits needed to encode the source with number of bits less than the assigned bit budget. Based on the iterative approach, ISO/MPEG can achieve the optimal audio quality for a given bit rate [1]. However, there is a possibility that an infinite loop occurs when the psychoacoustics model requires a very small quantization step size but the inner loop (rate control loop) still increases the quantization step size to fit the bit budget. To avoid such an infinite loop, the ISO/MPEG adopts three extra conditions to stop the iterative process as follows:

1. None of the scalefactor bands has more than the allowed distortion.
2. The next iteration would cause the amplification for any of the bands to exceed the maximum allowed value.
3. The next iteration would require all the scale-factor bands to be amplified.

Even though the committee approach can exploit the optimal audio quality for any bitrate, its iterative approach make it inefficient for real-time audio coding.

2.2 LAME BIT ALLOCATION ALGORITHM

LAME algorithm [2]-[3], which is an open source implementation, has improved two of the key modules including the psychoacoustics model and noise shaping in the MP3 encoding process as shown in Fig. 1. Thus, the LAME algorithm has high speed and superior quality that rivals most commercial competitors [3]. However, the iterative nature makes it difficult to obtain a low complexity MP3 encoder.

There are two major differences between the MPEG/Audio committee approach [1] and the LAME algorithm [2]-[3]. The first one is the committee approach uses exhaustively iterative searches while the LAME approach [2] adopts a binary search in the inner loop for speedup. The other difference is the selection of the quantization results, which is designed to allocate the bits to the bands that provide the best rate-distortion behavior for encoding each frame. When the inner loop is done, the committee approach chooses the last of the quantization stepsizes visited within the outer loop. However, the LAME algorithm chooses the best of the quantization stepsizes [2]-[3]. The best quantization means the least quantization distortion within the outer loop. The default criteria in LAME algorithm are arranged in order of importance as in the following:

1. Less distorted scalefactor bands,
2. Lower sum of noise over the thresholds,
3. Lower total noise.

A larger step size search based on the global gain and NMR information of the previous frame was proposed to speed up the search in the inner iteration loop [4], [5]. However, both approaches still employ step-by-step search.

2.3 NOISE SHAPING FOR MPEG/AUDIO

The noise-shaping scheme is used to estimate the noise for each frame before the bit allocation. The optimization of noise-shaping scheme for bit allocation is usually based on the following two criteria [6].

1. Quantization error is the least in terms of mean square error (MSE), or the signal to quantization noise ratio (SNR) is the largest. However, the NMR is different over the frequency range. The noise is audible at some frequencies.
2. Quantization noise spectrum parallels the masking threshold and the summation of all NMR values is the least at each frame. This criterion concerns about the result of psychoacoustics model and the noise is equally audible in different frequency bands [6], [7].

The second criterion enables us to have better listening quality because it considers the psychoacoustics effect of human ears. However, to optimize the bit allocation under these criteria requires iterative calculation, which makes the encoding procedure less efficient. To avoid iterative calculation, we propose a new bit allocation model.

Table 2 The algorithm comparisons among MPEG committee, LAME, and our approaches.

	MPEG committee	LAME	Proposed
Outer loop	Selects the *last* iteration visited	Selects the *best* iteration visited	Compute the scalefactor with Eq.(6) and Eq.(7)
Inner loop	Exhaustive step-by-step search for the global gain given the target bitrate	Initially binary and then step-by-step search for the global gain	Compute the global gain with Eq. (8) and occasionally fine-tuning

3. NEW BIT ALLOCATION MODEL

To speed up the bit allocation process we propose a new and efficient model-based approach to replace the iterative process and retain the listening quality simultaneously.

3.1 SINGLE LOOP DISTORTION CONTROL ALGORITHM

Similar to the algorithm by Liu et al. [5], our distortion control algorithm employs the second criterion for shaping the noise spectrum. Consequently, the estimated noise is used to decide the value of scalefactors that can parallelizes the quantization noise spectrum with the masking threshold. Since the noise estimation is a deterministic process, single loop distortion control algorithm can provide a significant reduction of execution time.

In the MP3 coding specification [1], the quantization value of the spectral signals $xr_i(j)$ and its approximation can be derived from

$$ix_i(j) = n\operatorname{int}\left(\left(\frac{xr_i(j)\cdot scale_i}{2^{gain}}\right)^{0.75} - 0.0946\right) = \left((xr_i(j))^{3/4} + e_i(j)\right)\cdot\left(\frac{scale_i}{2^{gain}}\right)^{3/4} \tag{1}$$

, where *gain* is 0.25*(global_gain-210) and $scale_i$ is a function of the scalefactor for the *i*-th scalefactor band. $e_i(j)$ is the estimation error for the quantized spectral signals $ix_i(j)$ at the *i*-th scalefactor band and the *j*-th spectral line. Therefore, the dequantization value of $ix_i(j)$ equals to

$$\overline{x}r_i(j) = sign(xr_i(j))\cdot\left((xr_i(j))^{3/4} + e_i(j)\right)^{4/3} \tag{2}$$

with the sign information transmitted by $sign(xr_i(j))$. The estimation error using the specified gain value can be measured by the mean square error (MSE) between $xr_i(j)$ and its reconstruction $\overline{x}r_i(j)$. However, to compute the MSE as in Eq. (2) is time consuming. To save the computational cost, the following two steps covering the expanding of the MSE using binomial theorem and the reduction of the approximation order can approximate the distortion in a MSE manner. Thus, the distortion function can be represented by

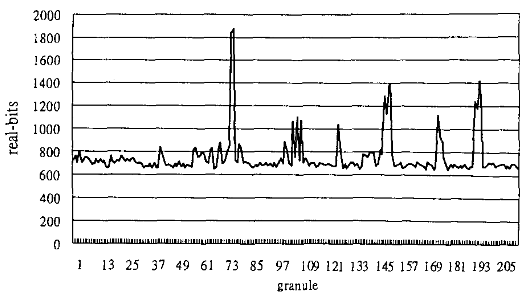


Fig. 3. The real bits used in the encoding processing. The audio sequence is "Guitar" at sampling rate of 44.1kHz at 128Kbps.

$$N(Noise)_i = E\left((xr_i(j) - \bar{x}r_i(j))^2\right) = \frac{16}{9}E\left(xr_i^{\frac{1}{2}}(j)\cdot e_i^{2}(j)\right) + \frac{16}{27}E\left(xr_i^{\frac{-1}{4}}(j)\cdot e_i^{3}(j)\right) + \frac{4}{81}E\left(xr_i^{-1}(j)\cdot e_i^{4}(j)\right) \quad (3)$$

To reduce the approximation order, it is assumed that the random variables $xr_i(j)$ and $e_i(j)$ are independent and $e_i(j)$ is uniform distributed. Consequently, we can derive the first and second order approximation of Eq. (3) as

$$N(Noise)_i \approx 0.1640578 \cdot E\left(xr_i^{\frac{1}{2}}(j)\right)\cdot\left(\frac{2^{gain}}{scale_i}\right)^{3/2}, \quad (4)$$

and

$$N(Noise)_i \approx 0.1640578\cdot E\left(xr_i^{\frac{1}{2}}(j)\right)\cdot\left(\frac{2^{gain}}{scale_i}\right)^{3/2} - 0.0145165\cdot E\left(xr_i^{\frac{-1}{4}}(j)\right)\cdot\left(\frac{2^{gain}}{scale_i}\right)^{9/4} + 8.4221\times10^{-4}\cdot E\left(xr_i^{-1}(j)\right)\cdot\left(\frac{2^{gain}}{scale_i}\right)^{3} \quad (5)$$

, respectively. The first order approximation is employed here since its simplicity and ignorable approximated error in addition to the error contributed from the second order approximation.

In addition, the $N(Noise)_i$ can be derived from the psychoacoustics model by

$$N(Noise)_i = M_i \times NMR_i \quad (6)$$

By substituting Eq. (6) into Eq. (4), the $scale_i$ can be calculated by

$$scale_i \cong 2^{gain}\left(\frac{D_i(xr_i(j))}{M_i\cdot NMR_i}\right)^{2/3} \quad (7)$$

with

$$D_i(xr_i(j)) = 0.1640578\cdot E\left[xr_i^{1/2}(j)\right] \quad (8)$$

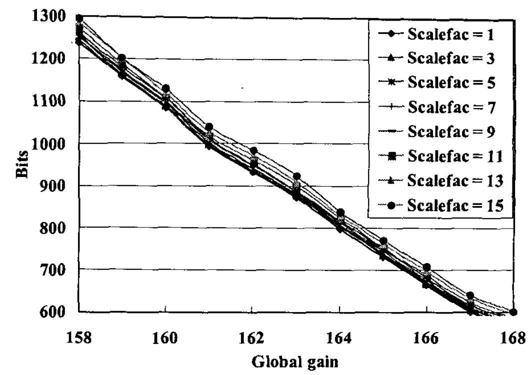


Fig. 4. A linear model between bitrate and global gain parameterized by scalefactors. The audio sequence is "Guitar" at sampling rate of 44.1kHz for the sixth band.

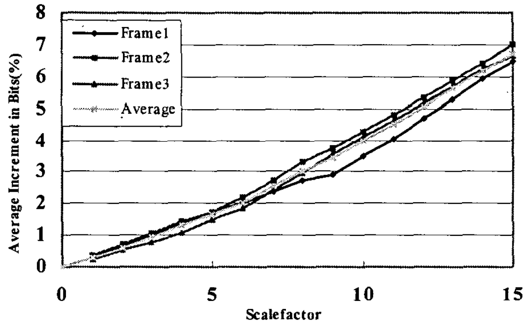


Fig. 5. Illustration of monotonically increasing function that denotes the average percentage of the incremental bit rate under the various scalefactors in coding the 6th scalefactor band of the Guitar audio sequence.

,where the $scale_i$ controls the magnitude of the quantization distortion for each scalefactor band i and the $gain$ controls the bit rate of the quantization for each granule. NMR_i is the noise-to-masking ratio of band i. $E[X]$ denotes the expectation value of random variable X and M_i is the masking threshold from signal energy by signal-to-masking ratio. From Eq. (7) when we change the value of $gain$ by fixing the magnitude of $scale_i$, the value of NMR_i will be changed simultaneously. Because the $gain$ is the same for all scalefactor bands, we can also get the same NMR_i for all scalefactor bands. To determine the scalefactors, which are used to parallelize the estimated noise spectrum and the masking threshold energy, we let the $scale_i$ equal to unity at the band i, where the ratio of the masking threshold energy M_i to $D_i(xr_i(j))$ is the maximum. Thus, we can determine the other scalefactors in the remaining bands with the value of $D_i(xr_i(j))$ and the masking threshold energy M_i is defined by

$$scale_i = \left[\frac{\max\left(M_i/D_i(xr_i(j))\right)}{\left(M_i/D_i(xr_i(j))\right)}\right]^{2/3} \quad (9)$$

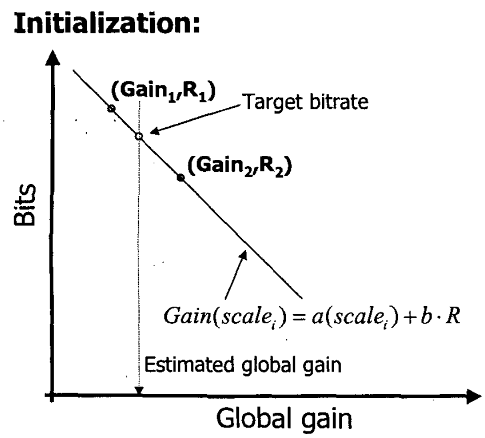


Fig. 6. Initialization of proposed fast rate control algorithm

After choosing the scalefactor in each band, we can quantize the spectral signals that are amplified by scalefactors $scale_i$ derived by Eq. (9).

Our proposed algorithm is similar to the algorithm of Liu et al. [5]. However, we improve the noise estimation with more accuracy by considering the bias term 0.0096 in Eq. (1). In addition, we enhance the listening quality by modifying the scalefactor selection method based on the following two considerations.

1. The importance for different scalefactor band.

2. The extra overhead for coding enlarged scalefactors at low frequency bands

Because even the masking threshold is usually large at low frequency bands, the large scalefactors may use more bits that can force the inner loop to increase the global gain for each granule. The increased global gain produces more qunatization noise and thus decreases the quality. Thus, in the outer iteration loop, we select the scalefactors for different bands by rescaling the values of $D_i(xr_i(j))$.

3.2 FAST RATE CONTROL ALGORITHM

To improve the iterative approaches, we propose a linear model-based algorithm for replacing inner iteration loop. The model-based algorithm is to employ a linear relationship among the following three factors including the global gain, the used bits after quantization, and the scalefactor. Theoretically, the relationship between the global gain (distortion) and actually used bits is non-linear. However, if the target bit rate is specified as, for example, 128kbps, the required bits for each temporally successive frame are ranged from about 600 bits to 2000 bits according to the statistics as shown in Fig. 3. In this specified range of bits, the correlation of the required bits and global gains is close to 0.99. Consequently, we can view the dependence between the actual bits and global gains as a piecewise linear function and adopt a linear model to map to and from the global gain and required bits at a fixed bit rate. When each spectral line is amplified by a scalefactor for each band, the actual bits needed for each frame is increased monotonically as illustrated in Fig. 4 and Fig. 5. Therefore, a shift of the line occurs when any scalefactor of each band is increased simultaneously. In

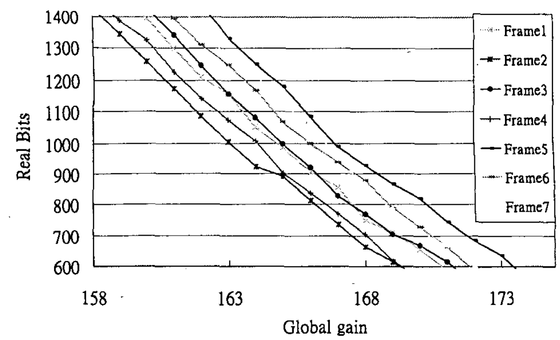


Fig. 7..Relationship of the slopes in successive frames

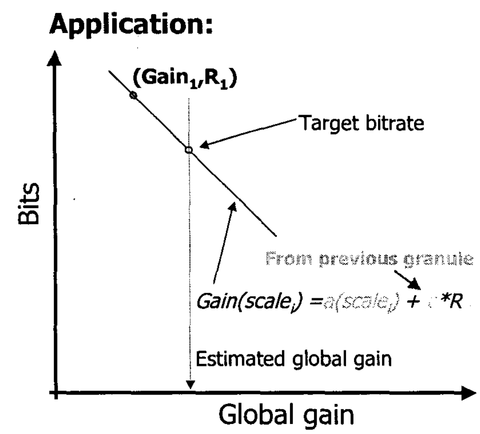


Fig. 8. Application of proposed fast rate control algorithm

summary, the dependence among the required bits, global gains and scalefactors can be modeled as

$$Gain(scale_i)=a(scale_i)+b\mathrm{R} \tag{8}$$

Where the $a(scale_i)$ is the offset of the global gain with the set of scalefactors, $\{scale_i\}$. The symbol b denotes the ratio of the varying magnitude of the global gain versus the actual bits. That is, we can estimate a global gain for a specified bit budget based on this linear function. Initially, we use two global gains, which are obtained from the temporally preceding frames, to quantize the audio spectral lines and to get the real bits as shown is Fig. 6. If the difference between the real bit and the target bit is less than a specified threshold, which is small enough to ensure the quality, the actual global gain can be found by Eq. (8). Because of the similar relationships of the slope in different frames or granules as shown in Fig. 7, for the next frame or granule we can use only one iteration of encoding and the slope b of previous frame or granule to get a new $a(scale_i)$ and get a global gain through Eq. (8) as shown in Fig. 8. Otherwise, we proceed to a fine-tuning stage that calculates new $a(scale_i)$ and b of this model and finds the final global gain with extra calculations for quantization. Without entering the fine-tuning stage, we spend only one iteration except that the first two iterations on quantization procedure is needed to derive the global gain. The extra overhead is necessary for model refinement, which decreases the performance of the proposed bit allocation algorithm. Since the strong similarity between the temporally successive frames, the percentage of the frames that employ the fine-tuning is small. Consequently, the averaged iteration numbers are reduced since most of the frames or granules take only single iteration to compute the global gain.

3.3 SUMMARY OF NEW BIT ALLOCATION ALGORITHM

The flow chart of our proposed bit allocation algorithm is shown in Fig. 9. In our proposed algorithm, there is no iterative computation. The distortion control can be done by a single loop calculation and the rate control loop can compute the appropriate global gain easily for a given target bitrate with the fine-tuning processing.

4. EXPERIMENTAL RESULTS

In our simulations, we use the code from the LAME project [2] as the platform because has higher quality and faster performance. The functions employed for the outer and inner iteration loops are denoted as "cal_noise" and "count_bits" in LAME encoder, respectively.

There are three primary experiments in this subsection. The first one is simulation results about the precision of the noise estimation. The second one is for the improvement in speed from the new bit allocation algorithm. The third one is to measure the listening quality with a subjective measure and the listening test.

The performance of the proposed model-based bit allocation approach is compared with that of the LAME algorithm [2], which has rapid execution and superior quality that rivals most commercial competitors [3]. The comparisons are based on the four factors including the average number of hits per frame (granule), reconstructed quality, audio sequences, and target bits. To compare the reconstructed quality of both LAME and our proposed bit allocation approaches, two objective measures including the total number of distorted bands and the total average Noise-to-Masking ratio (NMR) for each segment are employed. The NMR for each segment is an objective quality measure for the coded/decoded audio signals. The definition and features of the test audio sequences are shown as Table 3.

4.1 NOISE ESTIMATION

We use Eq. (4) to estimate the quantization noise and compare it with real quantization noise. Fig. 11 shows the noise spectrum in the band without amplifying the input spectral signals by the scalefactors. For a fair comparison, the same global gain from the final quantization result is used to calculate the estimated noise and real quantization noise. The estimated noise is calculated by Eq. (4) and the real noise is computed within the quantization processes. Fig. 11 shows that the energy of the estimated noise is almost identical to the energy of the real noise. Therefore, the noise estimation approach defined in Eq. (4) can estimate the actual noise for each frame based on the first order approximation of a binomial expression.

4.2 PERFORMANCE COMPARISONS

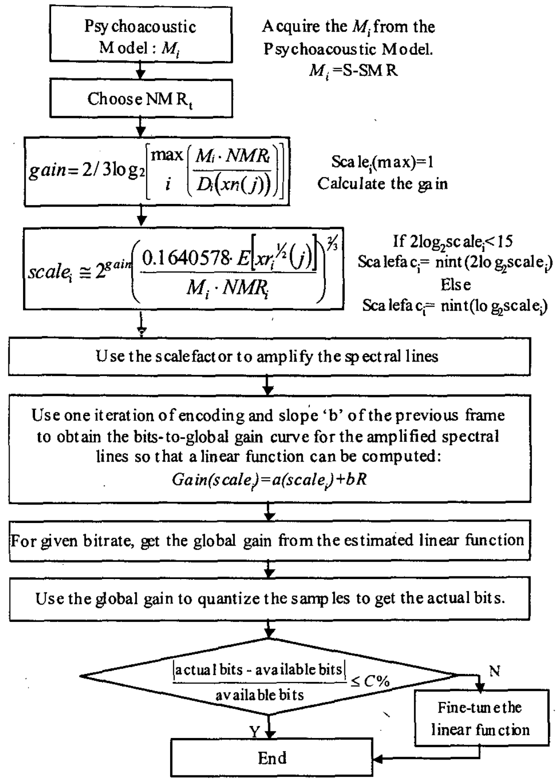


Fig. 9. The flow chart of the model-based bit allocation algorithm.

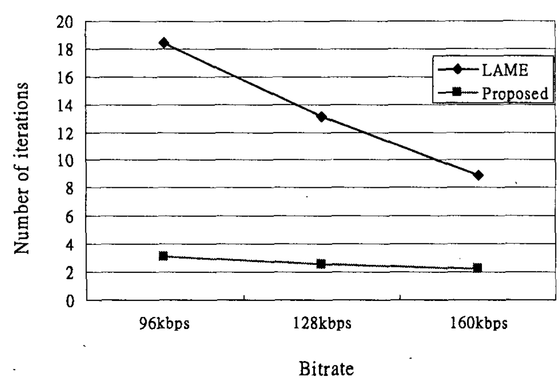


Fig. 10. The trend of number of iterations when encoding bitrate is decreased for the different rate distortion control algorithms. The Guitar audio sequence is used for simulations.

Table 4 demonstrates the hits per granule of the LAME and the proposed new bit allocation algorithm. For each simulation, the specified threshold for deciding the global gain using the proposed linear model is set as ±5% of the target bit rate. In

Table 3. The definition and features of audio sequences.

	Audio Sequences	File Size (KB)	Number of Frames	Number of the granules *1			
				Frame or Granule Type			
				Normal	Start	Short	Stop
1	Female Human	49514	40767	989	1264	990	11005
2	Organ	3446	2247	74	77	74	767
3	Guitar	2757	1359	126	231	126	614
4	Violin	4996	3811	7	9	7	1112
5	Piano Quintet in A	43276	38193	89	97	89	9618
6	String Quartet in D	37511	33249	31	33	31	8337

*1 The entry is total number of the blocks with the specific type contained in the audio sequence. The type of block is selected by the window switching decision based on psychoacoustics model.

terms of computational speed, we have about 3.28 to 7.11 times improvement by changing the LAME's outer iteration loop and achieve about 3.09 to 6.33 times speedup by changing the LAME's inner iteration loop at 128kbps. The magnitude of hits per granule for finding the scalefactors by our algorithm is unity. This is because we use the estimated noise to select the values of scalefactors without using any iterative calculation as the approach in [5].

Table 5 and Fig. 10 show the variation of hits per granule when the bitrate changes. The LAME algorithm uses different lowpass filter when the bit rate is changed. The lowpass operation will affect the number of the outer and inner iterations. In order to observe the variation of iterations, we use identical bandwidth of 128kbps to encode six sequences. This is why the numbers of hits per granule are different in Table 5 and Table 4 for the Guitar test sequence. We found the proposed algorithm ('NEW') has better performance when the target bitrate is low and the LAME algorithm is more sensitive to the bitrate variation.

4.3 OBJECTIVE EVALUATION

To compare the reconstructed quality through both bit allocation approaches, subjective measures are employed. Additionally, in order to measure the quality, we employ two kinds of common objective criteria to evaluate the perceived audio quality.

1. Total number of distorted bands.
2. Average total NMR.

Based on the LAME algorithm [2], there are several criteria to measure the listening quality objectively. The most important one is the total number of distorted bands. The distorted bands are the scalefactor bands that have noise energies larger than the corresponding masking thresholds. The other objective quality measure is defined as the total average NMR by

$$NMR_{tot} = 10 * \log_{10} \frac{1}{N} \sum_{n} \left(\frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k,n]}{M[k,n]} \right) \tag{9}$$

Table 5. Performance comparison through the number of hits per granule by encoding the Guitar audio sequence with the LAME and the proposed algorithm ('NEW') at various bitrates

Bitrate (kbps)	Outer loop function (cal_noise) hits / granule			Inner loop function (count_bits) hits / granule		
	LAME	New	Speedup	LAME	New	Speedup
96	7.74	1.00	7.74	18.41	3.05	6.04
128	5.91	1.00	5.91	13.15	2.52	5.23
160	3.92	1.00	3.92	8.89	2.27	3.92

Table 4 Performance comparison through the number of hits per granule using various audio sequences for the LAME [2] and the new algorithm ('NEW') at 128kbps.

Audio sequences	Outer loop function (cal_noise) hits / granule			Inner loop function (count_bits) hits / granule		
	LAME	New	Speedup	LAME	New	Speedup
1	7.11	1.00	7.11	14.32	2.37	6.05
2	3.28	1.00	3.28	6.85	2.22	3.09
3	5.61	1.00	5.61	12.24	2.46	4.97
4	5.15	1.00	5.15	10.85	2.32	4.67
5	6.32	1.00	6.32	13.94	2.36	5.92
6	6.93	1.00	6.93	14.76	2.33	6.33

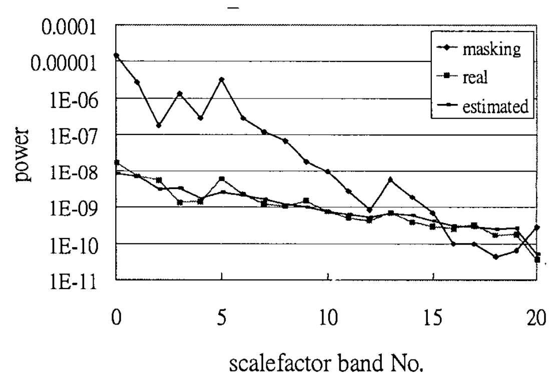


Fig. 11. Energy curves of the masking threshold, real quantization noise, and estimated noise of the 100^{th} frame from the Guitar audio sequence.

,where $P_{noise}[k,n]$ means the energy of the noise for the n-th frame. The symbol k is the index of FFT lines. The symbol $M[k,n]$ indicates the energy of the masking threshold. The symbol Z is the total number of FFT lines and the symbol N is the total number of frames in the current audio sequence. If the value of NMR is negative, it means the noise is inaudible.

Table 6 and Table 7 show the results of objective quality evaluation. In the Table 6 and Fig. 12, we can find the number of distorted bands using the proposed algorithm is less than that by the approach in [5] and is close to that with the LAME algorithm, which uses the iterative approach for distortion control algorithm. We can take it as an optimum case. In the Table 7, the values of total average NMR are almost the same for the three algorithms.

4.4 SUMMARY OF THE EXPERIMENTAL RESULTS

From the experiment results, we find that the estimated noise energy is close to the real noise energy. The proposed fast rate control algorithm just needs less than 2.5 times evaluations at 128kbps and is more robust to the variation of bit rates. In addition, the objective quality of our proposed single loop distortion control algorithm is better than that using the approach in [5].

5. CONCLUSION

In this paper, we present a single loop distortion control algorithm at the outer loop and a new rate control algorithm based on the linear model to improve the speed of the iterative approaches for MP3 coding. As to the distortion control algorithm, based on the modified scheme, we can get better objective quality.

For more accurate noise shaping, we estimate the energy of quantization noise and calculate the relationship of scalefactors from the different scalefactor bands to form the noise spectrum that is almost parallel to the masking threshold. With the derived noise spectrum, the listening quality is further improved with an accurate noise shaping method and by assigning less scalefactors for the low frequency bands and zero scalefactors for the bands having negative SMR. Experimental results show that we spend about 1/6 of calculation time for the distortion control loop in LAME algorithm.

The other innovation is about rate control scheme. Based on the linear function of the required bits, global gains and scalefactors, the new rate control algorithm can find the sub-optimal global gain for at most 2.5 times searches. Since the similar relationship among the required bits, global gains and scalefactors can be found at the MPEG-1, 2/Audio encoders, the efficient rate control approaches can be applied into these audio coding standards for speeding up their bit allocation algorithm. The proposed algorithm can also be applied to the other iterative approaches such as MPEG-1, 2 and AAC specifications.

Table 6. Number of distorted bands for the three algorithms. The small value indicates the less distortion bands and thus implies the better quality.

Algorithm / Audio Sequences	LAME	Liu et al.	New
1	45607	114864	78181
2	95	233	112
3	1109	2501	1691
4	2345	2573	1676
5	31898	64580	41527
6	22083	41527	25745

Table 7. Total average NMR (in dB) for the three algorithms. The small value indicates the less distortion bands and thus implies the better quality.

Algorithm / Audio Sequences	LAME	Liu et al.	New
1	-9.40	-8.71	-9.18
2	-11.52	-10.32	-10.12
3	-10.00	-9.44	-9.69
4	-10.68	-9.38	-9.51
5	-7.65	-7.30	-7.35
6	-7.98	-7.72	-.7.73

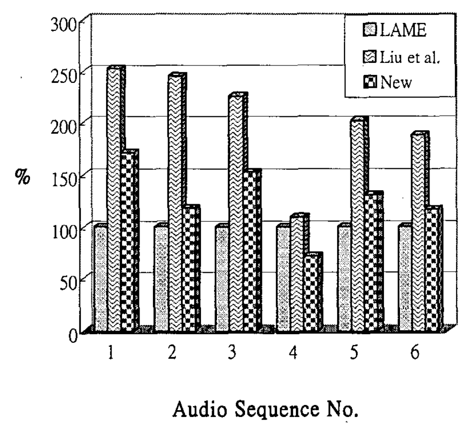


Fig. 12. The practical numbers of the distorted bands using the three algorithms covering the LAME., Liu et al, and our new approach.

6. REFERENCES

[1] ISO/IEC IS 11172-3, "Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s. – part3 : audio, " 1992.

[2] http://www.mp3dev.org/mp3/

[3] K. Brandenburg and G. Stoll, "ISO-MPEG-1 audio: a generic standard for coding of high-quality digital audio, " *J. Audio Eng. Soc.*, vol. 42, pp.780-792, Oct. 1994.

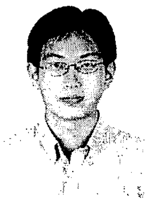
[4] K.-T. Fung, Y.-L. Chan, and W.-C. Siu, "A fast bit allocation algorithm for MPEG audio encoder, " in *Proc. of International Symposium on Intelligent Multimedia, Video and Speech Processing*, 2001, pp. 5–8.

[5] C. M. Liu, C. C. Chen, W. J. Lee, and S. W. Lee, "A fast bit allocation method for MPEG layer III, " in *Proc. of ICCE'99*, 1999, pp. 22-23.

[6] X. Wei, M. J. Shaw, and Martin R. Varley, "Optimum bit allocation and decomposition for high quality audio coding," in *Proc. of ICASSP'97*, 1997, pp.315-318.

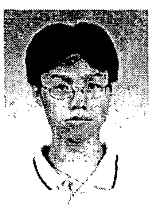
[7] H. Najafzadeh, and P. Kabal, "Perceptual bit allocation for low rate coding of narrowband audio," in *Proc. of ICASSP'00*, 2000, pp.893-896.

BIOGRAPHIES



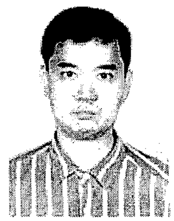
Chun-Yi Lee was born in I-Lan, Taiwan in 1977. He received the B.S. degree in electronics engineering from National Chiao-Tung University in 2000. He is now pursuing his Master degree in the Institute of Electronics Engineering, NCTU. His current research interest is digital audio compression.

Yao-Chun Fang was born in Tainan, Taiwan in 1979. He is an undergraduate in electronics engineering from National Chiao-Tung University. His current research interest is in the area of digital audio compression.



Hsiao-Chiang Chuang was born in Kee-Lung, Taiwan, on December 8, 1979. He is currently pursuing his B.S. degree in electronics engineering from National Chiao-Tung University, Hsin-Chu, Taiwan. His research interest is digital audio compression.

Chung-Neng Wang was born in PingTung, Taiwan, in 1972. He received the B.S. degree in computer engineering from the National Chiao-Tung University (NCTU), HsinChu, Taiwan, in 1994. He is now workint toward his Ph. D degree in the Institute of Computer Science and Information Engineering, NCTU. His current research interests are video/image compression, motion estimation, video transcoding and streaming.



Tihao Chiang was born in Cha-Yi, Taiwan, Republic of China, 1965. He received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1987, and the M.S. degree in electrical engineering from Columbia University in 1991. He received his Ph.D. degree in electrical engineering from Columbia University in 1995. In 1995, he joined David Sarnoff Research Center as a Member of Technical Staff. Later, he was promoted as a technology leader and a program manager at Sarnoff. While at Sarnoff, he led a team of researchers and developed an optimized MPEG-2 software encoder. For his work in the encoder and MPEG-4 areas, he received two Sarnoff achievement awards and three Sarnoff team awards.

Since 1992 he has actively participated in ISO's Moving Picture Experts Group (MPEG) digital video coding standardization process with particular focus on the scalability/compatibility issue. He is currently the co-editor of the part 7 on the MPEG-4 committee. He has made more than 50 contributions to the MPEG committee over the past 10 years. His main research interests are compatible/scalable video compression, stereoscopic video coding, and motion estimation. In September 1999, he joined the faculty at National Chiao-Tung University in Taiwan, R.O.C. Dr. Chiang is currently a senior member of IEEE and holder of 9 US patents and 26 European and worldwide patents. He was a co-recipient of the 2001 best paper award from the IEEE Transactions on Circuits and Systems for Video Technology. He published over 30 technical journal and conference papers in the field of video and signal processing.