Low Complexity Bit Allocation Algorithms for MP3/AAC Encoding

Conference Paper · May 2008

Suresh Kumaraswamy
College of Engineering Trivandrum

Tv Sreenivas
Indian Institute of Science

Audio Engineering Society

Convention Paper

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

Low complexity Bit allocation algorithms for MP3/AAC encoding

S Nithin[1], K Suresh[2],and T.V. Sreenivas[2]

[1] *Dept. of ECE, National Institute of Technology, Surathkal, INDIA 575025*

[2] *Dept. of ECE, Indian Institute of Science, Bangalore, INDIA 560012*

Correspondence should be addressed to T.V. Sreenivas (tvsree@ece.iisc.ernet.in)

ABSTRACT

We have developed two reduced complexity bit-allocation algorithms for MP3/AAC based audio encoding, which can be useful at low bit-rates. One algorithm derives optimum bit-allocation using constrained optimization of weighted noise-to-mask ratio and the second algorithm uses decoupled iterations for distortion control and rate control, with convergence criteria. MUSHRA based evaluation indicated that the new algorithm would be comparable to AAC but requiring only about $1/10^{th}$ the complexity.

1. INTRODUCTION

Audio/video signal compression has caused a revolution in multimedia representation/storage/distribution. A typical user now expects high quality program material, from a large selection, in a portable device, even on the move. Generally, users are more sensitive to the audio quality and hence there is great demand for high quality audio distribution through band limited mobile channels. Traditionally, audio coding has addressed mainly transparent quality reconstruction through standardized low-complexity decoders. Since the encoder is not standardized, there has been many improvements proposed for quantization, psycho-acoustic models, etc. Also, since the encoding algorithm is much more complex than the decoder, there has been a lot of literature on reducing complexity of psycho-acoustic modeling, MDCT computation, quantization, etc. Another issue with traditional audio coding is that of variable rate encoding, to achieve perceptual transparency. Although the rate is moderated by a bit-reservoir, the variable rate is not well suited for certain fixed rate communication channels. The variable rate characteristic also results in variable delay, due to buffering. The variable delay or even fixed large delay, is not well suited for audio/video synchronization as well as audio/audio synchronization. Therefore, the problem of low-delay, low-complexity, fixed-rate, high quality audio coding is still a challenging issue.

State of the art audio coders can achieve almost transparent quality reconstruction of CD quality audio at $64 Kbits/sec/channel$ bit-rate; i.e., $128 Kbits/sec$ for stereo programs. This corresponds to $1.5 bits/samp$ representation of the CD audio data. For lower band-width signals, such as 8-16 kHz (16-32 KSamp/sec), it is more difficult to achieve transparent quality at the same rate of bits/samp. Often a higher rate is required, such as 2 bits/samp. Currently, there is emphasis for transparent quality reconstruction of such low bandwidth signals as well as their graceful degradation of quality at lower bit-rates. These schemes would be attractive for mobile communication channels, whose bit rates are limited. The lower bit rate representation, coupled with low delay and low complexity are typical of speech coders, which are not effective when used directly for music or general audio signals. Therefore, there is need to develop new techniques of audio compression which are more effective than the classical techniques.

There is much literature [1-8] addressing the issue of reducing complexity of MP3/AAC encoders. Many of these are aimed towards efficient DSP implementation of transparent quality reconstruction of high quality audio (CD). However, there are few attempts towards efficient bit-allocation for lower than transparent quality. Perceptual issues of lower bit-rate quantization noise are not adequately addressed. The recent results in [1, 2] illustrate that least complexity can be achieved by omitting the iterative quantization of MP3/AAC, but instead use an upper bound on the quantization error to determine bit-allocation; naturally this leads to over estimation of the source rate, although the reconstruction quality may be satisfactory. Another approach is to decouple the nested quantization loops of the ISO recommendation to sequential loops [3]. This has been shown to require only 20% complexity, again with comparable quality to AAC. This shows that with slight increase in bit rate we may significantly reduce complexity. In other attempts to complexity reduction, the quantization noise shaping is sacrificed using uniform quantization [4] or combining the left and right channels of the stereo signal into a single stream for bit allocation [5]. This approach requires careful evaluation of psycho-acoustic masking criteria and spatialization properties, in addition to complexity reduction. In [6] a significant speed up over the LAME bit-allocation algorithm is shown by converting the nested loops into a single iterative algorithm. The inner rate control loop is replaced by a linear model of global gain and bit-rate. In [7] the emphasis is on reducing the bit-rate and not complexity; hence, two nested loops structure is retained, but a moving average noise to mask ratio (NMR) is proposed to exploit inter-frame redundancies. In [8, 9], while complexity reduction is the goal, it is realized through efficient psycho-acoustic model, smoothed scale factors as well as loop free bit-allocation, again by over estimation of the noise level.

In this paper we address jointly the three closely related issues of low bit rate quantization (lower than transparent quality), lower complexity than the standard MP3/AAC algorithms, as well as a fixed rate representation. All the three issues are motivated by new applications of audio coding in mobile communication. We recognize that the quantization noise may not be completely masked, but its perceptual effect may be controlled through psycho-acoustic criteria. We propose two new algorithms for bit allocation, (i) single-pass algorithm of optimum bit allocation using weighted NMR and (ii) limited iteration algorithm of decoupled iteration loops. We evaluate the algorithms through a simplified implementation of the audio coder and the reconstructed audio quality is evaluated using MUSHRA and PEAQ scores. It is found that the new algorithms have selective advantages over those reported in the literature.

2. SINGLE-PASS ALGORITHM OF OPTIMUM BIT ALLOCATION USING WEIGHTED NMR

In the bit allocation process of an audio coder, the objective function which represents the audibility of the quantization noise, plays major role in deciding the quality of the encoded audio. If the objective function can track closely the actual audibility of the quantization noise, the performance of the coder will be better. Most of the existing objective functions are based on the NMR.

The NMR predicts the extent to which this noise signal would be masked by the original audio signal if both noise and signal are presented to a listener at the same time. The NMR is widely used in audio coding and audio-quality assessment.

In our algorithm, we use weighted NMR as the objective function. A closed form expression is derived to minimize the objective function under a fixed bit-rate constraint using Lagrange multiplier.

The problem is to find a bit distribution such that the weighted NMR,

$$D = \sum_{j=1}^{L} w_j \times \frac{\sigma_N^2(j)}{\sigma_M^2(j)} \tag{1}$$

is minimized subject to the constraint $\sum_{j=1}^{L} R(j)B(j) = R$

where

j scalefactor band index.

$\sigma_N^2(j)$ denotes the quantization noise energy in j^{th} scalefactor band.

$\sigma_M^2(j)$ denotes the masking energy associated with j^{th} scalefactor band.

$B(j)$ number of coefficients in j^{th} scalefactor band.

$R(j)$ bit-rate (bits/sample) to be used in the j^{th} scalefactor band.

w_j weight

L Number of scalefactor bands.

R Number of bits available for the frame of audio data.

$\sum_j B(j) = N$ the transform length.

Using Lagrange multiplier,

$$\frac{\partial}{\partial R(j)} \left[\sum_i w_i \frac{\sigma_N^2(i)}{\sigma_M^2(i)} + \lambda \left\{ \sum_i R(i)B(i) - R \right\} \right] = 0 \tag{2}$$

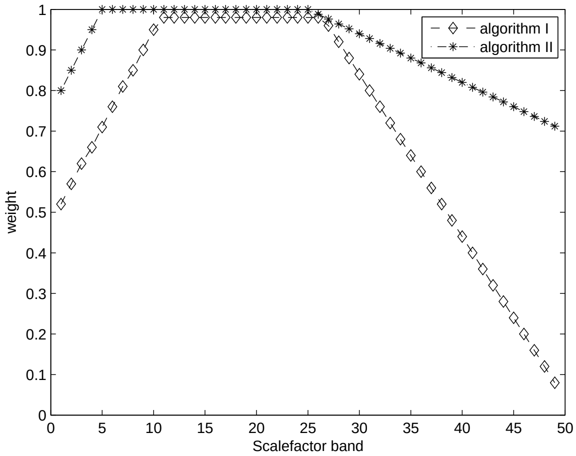

Fig. 1: Weights used for NMR (algorithm I - WNMR) and scalefactors (algorithm II - Limited Iterative algorithm)

From [10], we have

$$\sigma_N^2(j) = \epsilon \times 2^{-2R(j)} \times \sigma_x^2(j) \tag{3}$$

where $\sigma_x^2(j)$ represents the signal energy and ϵ is a input pdf dependent parameter.
Eqn(2) $\Rightarrow$

$$\left\{ -2log2 \times \epsilon \times 2^{-2R(j)} \times w_j \times \frac{\sigma_x^2(j)}{\sigma_M^2(j)} \right\} + \lambda \times B(j) = 0 \tag{4}$$

$\Rightarrow$

$$2^{-2R(j)} = \frac{\lambda \times B(j)}{2 \times log2 \times \epsilon \times [w_j \times \frac{\sigma_x^2(j)}{\sigma_M^2(j)}]} \tag{5}$$

$$R(j) = \frac{1}{2} \left[log_2 \left\{ \frac{\epsilon(2log2)\left(w_j \frac{\sigma_x^2(j)}{\sigma_M^2(j)}\right)}{B(j)} \right\} - log_2 \lambda \right] \tag{6}$$

$$\begin{aligned} R &= \sum_m B(m) \times R(m) \\ &= \sum_m B(m) \times \end{aligned} \tag{7}$$

$$\frac{1}{2}\left[log_2\left\{\frac{2log2\times\epsilon\times w_m\times\frac{\sigma_x^2(m)}{\sigma_M^2(m)}}{B(m)}\right\}-log_2\lambda\right]$$

Simplifying,

$$log_2\lambda = \frac{-2R}{N}+\frac{1}{N}\sum_m B(m)\times log_2\left\{\frac{2log2\times\epsilon\times w_m\times\frac{\sigma_x^2(m)}{\sigma_M^2(m)}}{B(m)}\right\} \quad (8)$$

Substituting (8) in (6)

$$R(j) = \frac{R}{N}-\frac{1}{2N}\times\sum_m B(m)\times log_2\left\{\frac{(2log2)\times\epsilon\times w_m\times\frac{\sigma_x^2(m)}{\sigma_M^2(m)}}{B(m)}\right\} + \frac{1}{2}log_2\left[\frac{(2log2)\times\epsilon\times\left(w_j\times\frac{\sigma_x^2(j)}{\sigma_M^2(j)}\right)}{B(j)}\right] \quad (9)$$

This gives R(j) in bits/sample to be used in the j^{th} scalefactor band. Some of the R(j) values may turn out to be negative, those values are made zero. Redistribution of bits is done by taking out bits from higher values of R(j) such that bit-rate constraint is satisfied. The first two terms in the above equation are constants. The bit-rate for the j^{th} scalefactor band is directly proportional to the weighted Signal to Mask ratio (SMR), $\left(w_j\times\frac{\sigma_x^2(j)}{\sigma_M^2(j)}\right)$ of that band. That is, if the SMR is higher, it can tolerate less quantization noise, requiring more number of bits and vice versa.

The quantization step sizes are calculated as follows.

In each scalefactor band, the absolute maximum of the spectral coefficients, X_{max} is chosen. Then the step size is,

$$\Delta[j] = \frac{2\times X_{max}[j]}{2^{R(j)}} \quad (10)$$

Δ values are rounded to nearest integers. The spectral coefficients are quantized using

$$X_q[i] = int\left(\frac{|X[i]|^{\frac{3}{4}}}{\Delta[j]}\right) \qquad i=1:1024;\ j=1:49 \quad (11)$$

We have used the inverse of the absolute threshold of hearing in quiet as the weight function, which is plotted in Fig 1.

3. LIMITED ITERATION ALGORITHM OF DECOUPLED ITERATION LOOPS

In this algorithm, the bit allocation is done in three steps. In the first step, the quantization step sizes are initialized based on the estimated error variance. In the second step, the distortion is controlled such that the quantization noise is less than the masking thresholds for all the scalefactor bands. If the target bit-rate is not met, reshaping of the quantization noise is done in the third step to meet the bit-rate constraint.

Algorithm:
1.Initialization : The initial quantizer step sizes (i.e., the values of *global_gain* and *scalefactors*) are calculated based on the estimated quantization error variance, as explained below.
The quantization equation used in mp3/AAC encoding is given by

$$X_q[i] = int\left(\left[\frac{|X[i]|}{2^{\frac{1}{4}(global\_gain-sf[b])}}\right]^{\frac{3}{4}}+0.4054\right) \qquad i=1:1024;\ \ b=1:49 \quad (12)$$

where $X_q[i]$ denotes the quantized value of MDCT coefficient X[i]; sf[b] denotes the value of scalefactor for the b^{th} scalefactor band. *global_gain* is constant for the entire frame and sign(X[i]) is coded separately.
A representative value of the transform coefficients in each scalefactor band is chosen. Let us denote this value by X_b.

$$X_b = \max_{i\in b}\{|X[i]|\} \qquad \forall b \quad (13)$$

Let $X_{b,c} = (X_b)^{\frac{3}{4}}$ denote the compressed coefficient and

$$\Delta[b] = 2^{\frac{3}{16}\times(global\_gain-sf[b])} = 2^{\frac{3}{16}\times q[b]} \quad (14)$$

All the coefficients in the b^{th} scalefactor band are quantized with a step size of $\Delta[b]$
In a particular scalefactor band, X_b is quantized using equation(12) as

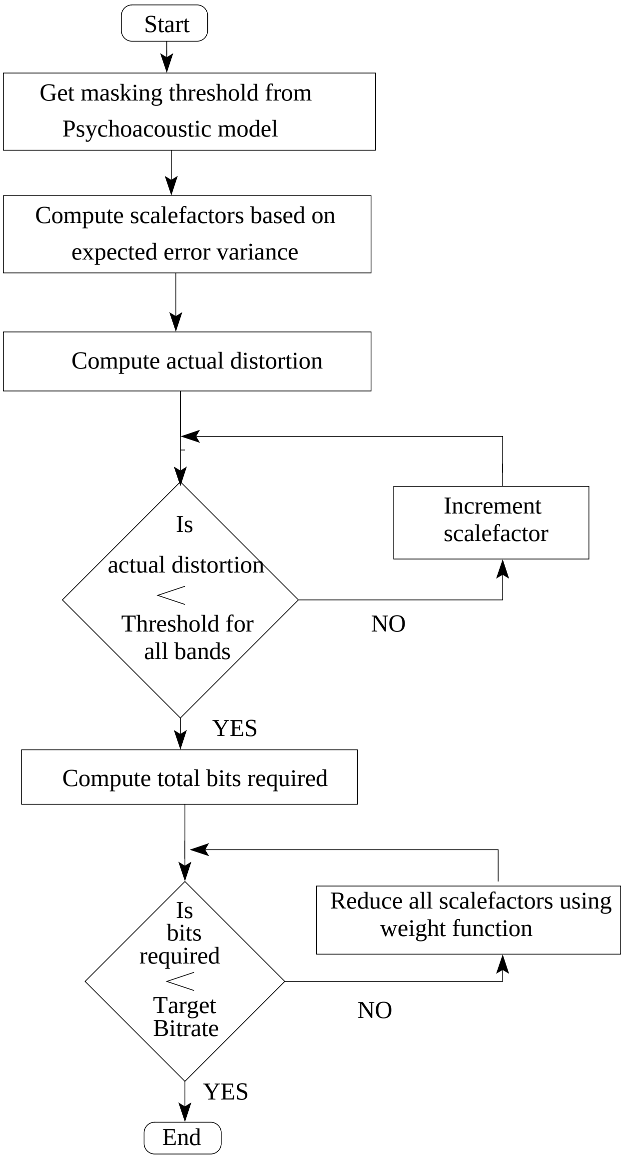

Fig. 2: Limited Iteration Algorithm

$$Xq = int\left(\frac{X_{b,c}}{\Delta[b]} + 0.4054\right) \tag{15}$$

We know that for a uniform quantizer with step size Δ, the quantization error variance, σ^2, is given by $\sigma^2 = \frac{\Delta^2}{12}$. Therefore, error variance corresponding

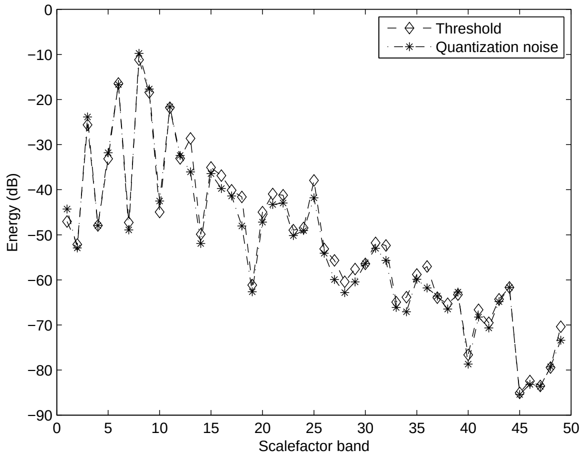

Fig. 3: Quantization noise after initialization

to $X_{b,c}$ is $\frac{\Delta^2[b]}{12}$ and the error variance corresponding to X_b will be

$$\frac{\Delta^2[b]}{12} \times \frac{X_b^2}{X_{b,c}^2} = \frac{\Delta^2[b]}{12} \times (X_b)^{0.5} \tag{16}$$

Now, this error should not exceed the masking threshold of the corresponding scalefactor band. Let 'Th[b]' denote the masking threshold for the b^{th} scalefactor band. Equating this to error variance,

$$\frac{\Delta^2[b]}{12} \times (X_b)^{0.5} = Th[b] \tag{17}$$

$$\Delta^2[b] = 12 \times Th[b] \times (X_b)^{-0.5} \tag{18}$$

From eqn(14) $\quad \Delta^2[b] = 2^{\frac{3}{8} \times q[b]}$
Therefore,

$$q[b] = \frac{8}{3} \times log_2\left[12 \times Th[b] \times (X_b)^{-0.5}\right] \tag{19}$$

$$global\_gain = \max_b \{q[b]\} \qquad b = 1 : 49 \tag{20}$$

$$sf[b] = \text{global\_gain} - q[b].$$

Thus, the *scalefactor* values are found for all the bands.

2. Distortion control Loop

Step (i): The spectral coefficients are quantized using equation(12).

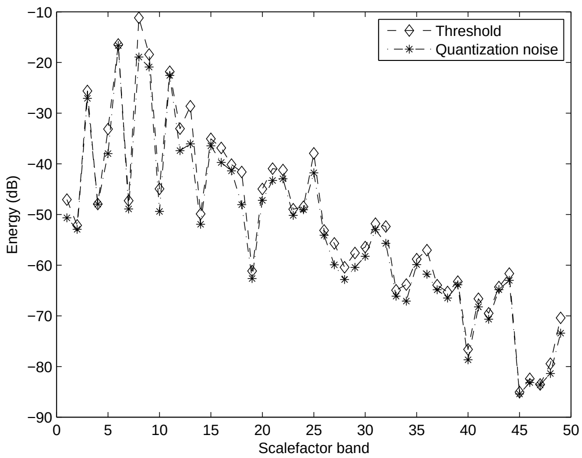

Fig. 4: Quantization noise after the Distortion control loop

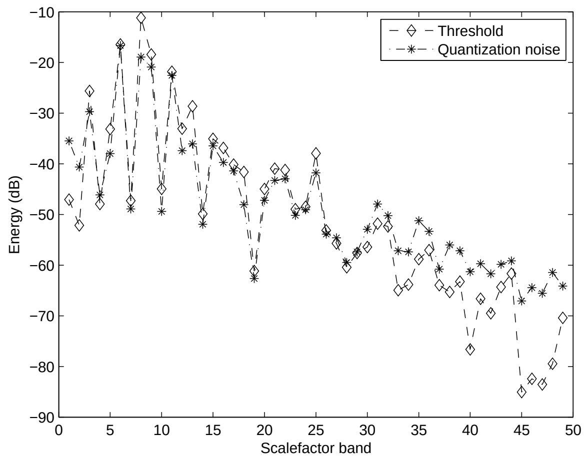

Fig. 5: Quantization noise after the Rate control loop

Step (ii): For each scalefactor band, the actual quantization noise obtained due to quantization is calculated. The spectral quantization noise is given by $e[i] = X[i] - \hat{X}[i]$, where X[i] and $\hat{X}[i]$ denote the original and reconstructed MDCT coefficients respectively.
Distortion in the b^{th} scalefactor band will be

$$D[b] = \frac{1}{L}\sum_{i=k_{b-1}+1}^{k_b} e^2[i] \tag{21}$$

k_b=scalefactor band boundary and L is the number of coefficients in the b^{th} scalefactor band.
The NMR is given by, $NMR[b] = \frac{D[b]}{Th[b]}$
Step (iii): If $NMR[b] > 1$, the quantizer step size for that band is reduced by incrementing the scalefactor.

Step (iv): Steps i-iii are repeated till the NMR in each of the scalefactor band is less than 1.

The actual distortion obtained after the initialization and after the distortion control loop are shown in Figs. 3 and 4.

Now the number of bits required to code the quantized coefficients and the side information is calculated. If the bits used is greater than the target bit-rate, the rate control loop is executed.
3. Rate control Loop

In this step, to reduce the required bit-rate, the quantizer step sizes of those bands are increased which are less perceptible to human ear. To increase the quantizer step sizes, the *scalefactors* are reduced by multiplying with weights which are based on the inverse of the threshold in quiet. The weights used is plotted in figure 1. After experimenting with weights of various slopes, these weights were chosen, as it gives better granularity in controlling the rate and distortion.
In this way, we are reshaping the quantization noise based on the absolute threshold of hearing. The quantizer step sizes are increased based on the weights, till the bit-rate constraint is met. The shaping of quantization noise after the Rate control loop is shown in Fig.5.

As the initial values of the step sizes are close to optimal, the search process requires very few iterations.

4. EXPERIMENTS

The proposed algorithms are compared with the Single pass algorithm [1], nested loop algorithm of AAC and MP3 [13]. The audio tracks given in Table 1, which are taken from SQAM database were considered for the comparison of algorithms. The chosen

1	Track08	Violin
2	Track20	Saxophone
3	Track21	Trumpet
4	Track29	Bass Drum
5	Track60	Piano
6	Track64	Choir

Table 1: Audio tracks used in the tests from SQAM database

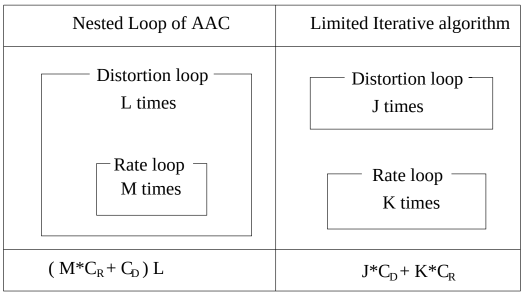

Fig. 6: Comparison of complexities

audio tracks were sampled at 32kHz.

4.1. SPA Algorithm

Single pass algorithm uses an approximate solution to the bit allocation in a single step. The quantizer step sizes are calculated based on the estimated maximum quantization noise. Starting from the first scalefactor band, the spectral coefficients are quantized. These coefficients are Huffman coded and the number of bits required to code these coefficients plus the side information is calculated. Then the next scalefactor band is taken up for coding. If the number of bits used is less than the available bits for the frame, the next scalefactor band is chosen until all the scalefactor bands are finished. Otherwise, the coefficients of the present scalefactor band are withdrawn and the process exits, without quantizing the subsequent bands. Thus, the SPA is a greedy algorithm of allocating bits to the lower frequency components at the cost of ignoring higher frequencies.

4.2. Complexity

We can compute the complexity between nested loop algorithm and decoupled loops of limited iterative algorithm, similar to the approach in [3]. If we represent the distortion computation load as C_D and rate computation as C_R, the number of outer loops as L and typical number of inner loops as M, the total computation load can be expressed as $L\,(MC_R + C_D)$. In the case of decoupled loops, the same can be expressed as $JC_D + KC_R$. where J and K are the number of iterations in the distortion loop and rate control loop respectively. Figure 7 shows a plot of L for successive frames of a typical music piece and we can see that at least $L = 40$ iterations is required, not to loose the optimum performance of the nested loops algorithm. For the same music piece Figure 8 and 9 show the plot of J and K for the successive frames. We can see that, approximately $J = 5$ and $K = 6$. Substituting these numbers and assuming M is small, we can see that there is a saving of at least, factor of 10.

4.3. Objective Quality

The average PEAQ [11] score obtained for the algorithms at 2 bits/sample, 1.5bit/sample, 1 bit/sample and 0.75 bit/sample are shown in Table 2. We can see that the nested loop approach of AAC gives the best reconstruction error at all bit-rates; however it is computationally expensive. The single pass algorithm does perform poorer as expected at higher bit-rates, but much poorer at lower bit-rates. Similarly the Limited Iterative algorithm provides a performance comparable to that of nested loops, but with much less complexity. The weighted NMR algorithm without the iterations (less complexity) does appear similar to that of Single pass algorithm [1], but perceptually it is found to have an edge over SPA, through the MUSHRA test[12].

4.4. MUSHRA Test

To evaluate the performance of the algorithms subjectively, we have conducted MUSHRA test. Six music pieces from the SQAM database, as mentioned in the Table 1 were used in the test. Low pass version of the original signal filtered at 3.5 kHz is used as the hidden anchor for the MUSHRA test. Nine trained listeners participated in the test. Test results for 2 bits/Sample, 1 bit/Sample and 0.75 bit/Sample are shown in Figs. 10, 11 and 12 respectively.

The order of the test signals is as follows:

1.Hidden reference.

2.Hidden anchor.

3.Nested loop algorithm of AAC.

Bitrate (bits/Samp.)	2	1.5	1	0.75
Nested loops	-0.03	-0.25	-0.5	-1.1
Single Pass	-0.3	-0.9	-1.54	-3.6
Weighted NMR	-0.33	-0.7	-1.19	-3.1
Limited Iteration	-0.31	-0.6	-0.8	-1.23

Table 2: Average PEAQ score

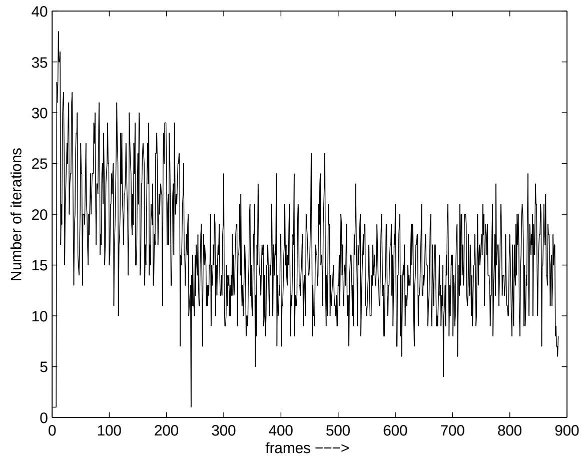

Fig. 7: Iteration index with the best performance in the nested loop algorithm.

4.Single pass algorithm[1]
5.Limited iteration algorithm
6.WNMR based optimum bit allocation.
7.MP3 (LAME encoder)[14]

At 2 bits/Sample, the performance of all the algorithms are almost equal, as sufficient bits are available for coding. At all bitrates, the performance of Limited Iterative algorithm is comparable to that of nested loop algorithm of AAC. Limited Iterative algorithm performs better than MP3 at lower bitrates. Comparing the single iteration algorithms, the performance of our algorithm(WNMR) is better than the Single pass algorithm [1] at lower bitrates. As expected, the performance of single iteration algorithms are poorer at low bitrates.

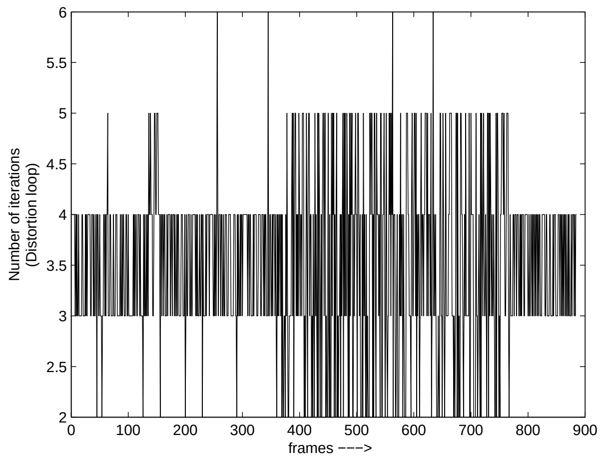

Fig. 8: Number of iterations for convergence in distortion control loop of Limited Iterative algorithm.

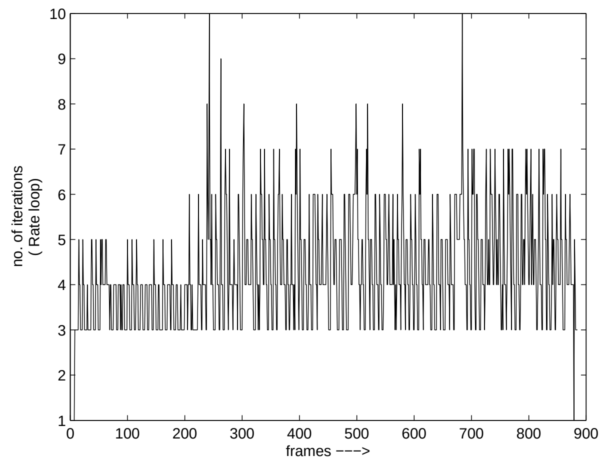

Fig. 9: Number of iterations for convergence in Rate control loop Limited Iterative algorithm.

5. CONCLUSIONS

As can be seen from the performance difference between the one single pass algorithm and the decoupled iterations algorithm, optimization based on actual quantization noise is important for good percep-

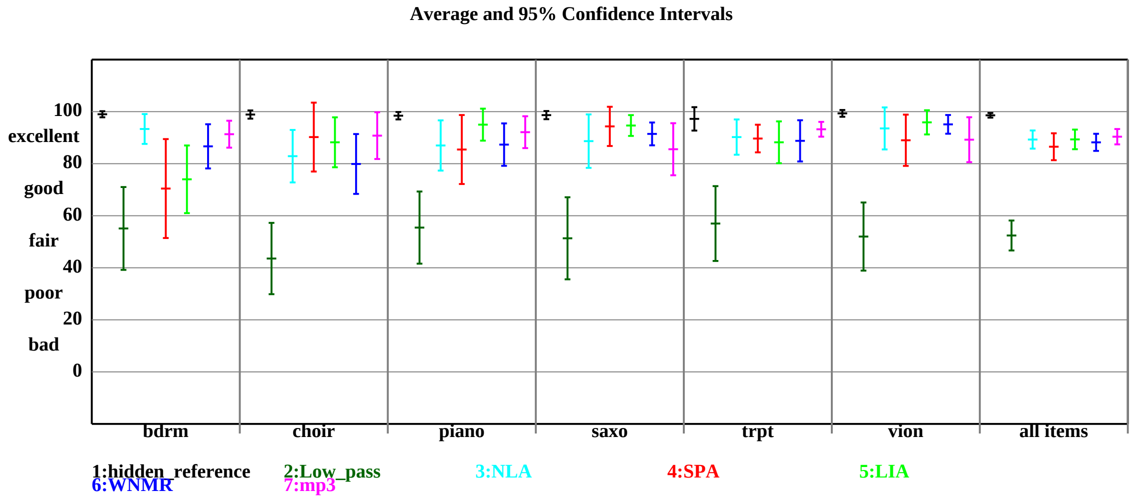

Fig. 10: Listening test results at 2 bits/Sample

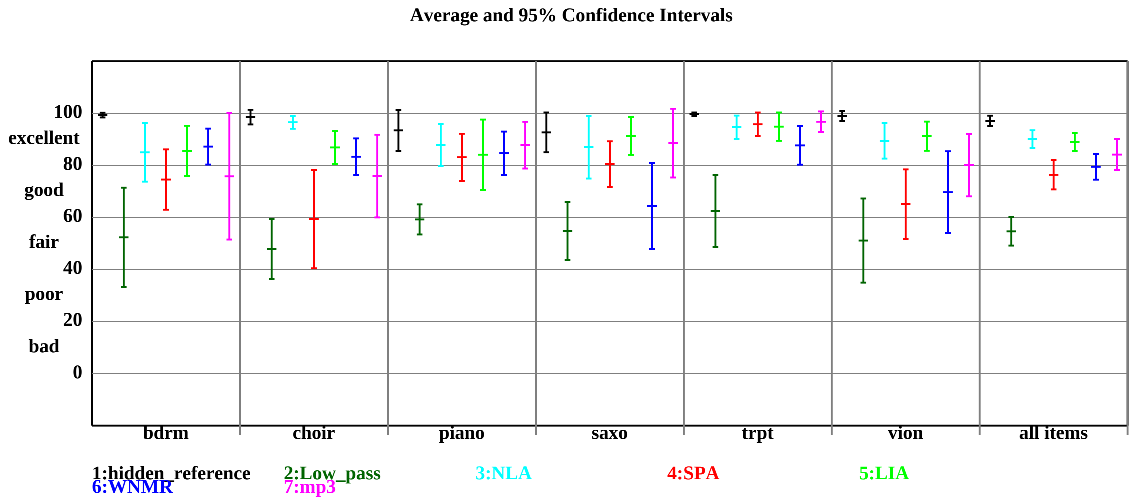

Fig. 11: Listening test results at 1 bits/Sample

tual quality, compared to the measure of expected quantization noise, particularly at lower bit-rates. There is need for further improvement of quality at low bit-rates.

6. ACKNOWLEDGMENT

We would like to thank Ms R.S. Vasupradha for conducting the listening tests.

7. REFERENCES

[1] H. Yamashita, "Audio coding device with fast

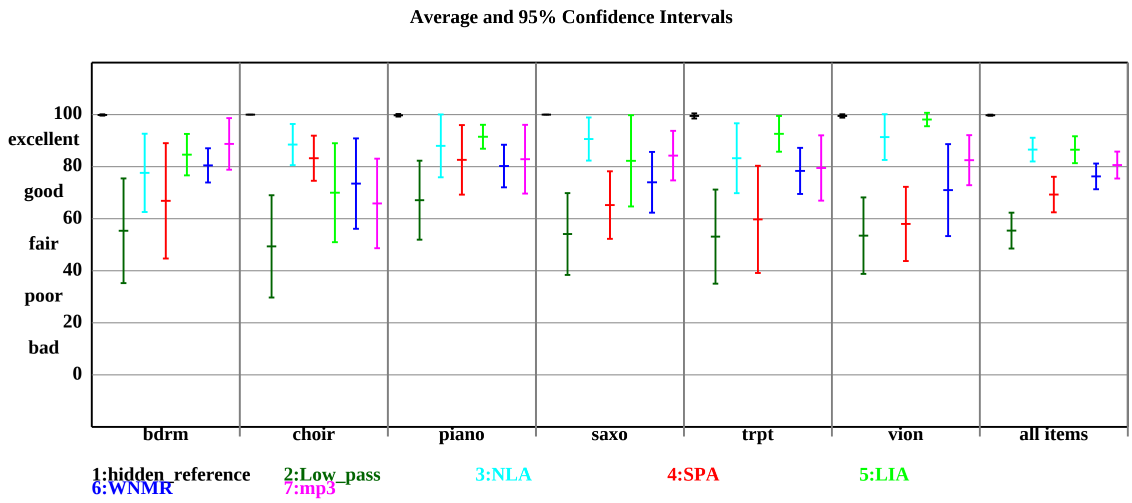

Fig. 12: Listening test results at 0.75 bits/Sample

algorithm for determining quantization step sizes based on psycho-acoustic model," US Patent No. 272223, Apr 2006.

[2] E. Alexandre, A. Pena and M. Sobreira, "Low complexity bit allocation algorithm for MPEG AAC audio coders," *IEEE Signal Proc. Letters,* Vol.12, No.12, pp 824-826, Dec 2005.

[3] K.H. Bang, K.S. Lee, Y.C. Park and D.H. Youn, "Fast bit allocation method for MP3/AAC encoders," *Proc. 118 AES Convention,* Barcelona, May 2005.

[4] P. Konda and V. Prakash, "Optimal bit allocation strategy for perceptual audio coders employing uniform quantization schemes," *Proc. 116 AES Convention,* Berlin, May 2004.

[5] S. C. Vadapalli and V. Prakash, "Efficient bit distribution strategy for stereophonic audio coders," *Proc. 116 AES Convention,* Berlin, May 2004.

[6] C.Y. Lee, Y.C. Fang, H.C. Chuang, C.N. Wang and T. Chiang, "A fast audio bit allocation technique based on a linear R-D model," *IEEE Trans. Cons. Electron.,* Vol. 48, No.3, pp 662-670, Aug 2002.

[7] K.H.C. Eng, D.Y. Huang and S.W. Foo, "A new bit allocation method for low delay audio coding at low bit rates, " *Proc. 112 AES Convention,* Munich, May 2002.

[8] A.D. Duenas, R. Perez, B. Rivas, E. Alexandre and A.S. Pena, "Real time implementation of MPEG-2 and MPEG-4 natural audio coders," *Proc. 110 AES Convention,* Amsterdam, May 2001.

[9] A.D. Duenas, R. Perez, B. Rivas, E. Alexandre, A.S. Pena, "A robust and efficient implementation of MPEG-2/4 AAC natural audio coders," *Proc. 112 AES Convention,* Munich, May 2002.

[10] N.S. Jayant and P. Noll, *Digital coding of waveforms: Principles and applications to speech and video,* Prentice-Hall, 1984.

[11] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality", TSP Lab Technical Report, Dept. Electrical and Computer Engineering, McGill University, May 2002. (http://www-mmsp.ece.mcgill.ca/Documents/Downloads/)

[12] ITU/*ITU-R BS 1534. Method for subjective assement of intermediate quality level of coding systems,* 2001.

[13] ISO/IEC, JTC1/SC29/WQ11 MPEG, "Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - part 3: Audio", IS11172-3 1992 ("MPEG-1").

[14] http://lame.sourceforge.net/