A NMR OPTIMIZED BITRATE TRANSCODER FOR MPEG-2/4 LC-AAC

Te-Hsueh Lai, Chung-Neng Wang and Tihao Chiang

National Chiao-Tung University, HsinChu 30050, Taiwan

ABSTRACT

This paper presents a noise-to-masking-ratio (NMR) optimized MPEG-2/4 AAC Low-Complexity (LC) transcoder, called as Fast Rate-Distortion Optimized Transcoder (FRDOT). The FRDOT searches for the optimal scalefactors under the NMR criterion according to the magnitudes of quantized input coefficients and the target bitrate. The computation of NMR difference is replaced by the derivation of signal-to-noise-ratio (SNR) difference since the audible masking thresholds of the original and transcoded bitstreams are identical. The SNR value is further converted to a noise-to-signal-ratio (NSR) to represent distortion energy of audio signals. With the NSR values, a table lookup technique is used to speed up the search of the optimal scalefactor increment. To further reduce the execution time, a bandwidth limiter is used to remove the iterative rate-distortion optimization of a frame. In addition, a bitrate control module is proposed to make the averaged bitrate of output bitstream close to the target bitrate. The results show that the FRDOT has smaller NMR values than the cascaded transcoder (CT) by 0.5-3.0 dB at different bitrates with 6-7 times speedup on the average.

1 INTRODUCTION

The multimedia delivery over networks becomes popular recently in real-time audio streaming services including music-on-demand (MOD), digital audio broadcasting (DAB), etc. To stream high quality audio, the required bitrates are typically 64 to 128 kilo-bits per second (kbps). To deliver audio bitstreams of high bitrates over long range (from continent to continent) or over wireless network [1] is still a challenge. The key issue is to reliably provide high quality audio services over varying bandwidth and heterogeneous networks including GSM, Wireless LAN, UMTS, etc. Even in a single connected network session, the bandwidth fluctuates during the transmission session. In case of channel bandwidth fluctuation or network congestion, packets may be lost or delayed during transmission, which is not acceptable at real-time applications. To prevent packet loss, the audio services should seamlessly lower the bitrates of delivering bitstreams. The dynamical bitrate adaptation is required to simultaneously serve multiple clients with different capacities of receiving, decoding and playback as in MOD applications. Bitrate adaptation methods [2]-[3] have been used to trade audio quality for transmission throughput in real-time content delivery applications. The bitrates of delivering bitstreams are adapted to fit to channel bandwidth and receivers' capacities. In addition, to save storage space, single bitstream of the specified format for each audio sequence is archived at the server database.

For audio streaming, the archived bitstreams can be generated by scalable coding and non-scalable coding. For scalable audio coding, MPEG-4 bit slice arithmetic coding (BSAC) could be used. With the side information for scalability, the coding efficiency of BSAC is generally much lower than non-scalable MPEG-2/4 AAC (AAC) coding at the same rate. Thus from the rate-distortion performance efficiency viewpoint, a non-scalable AAC coding scheme is adopted and it is followed by a bitrate adaptation transcoder that directly converts the compressed bitstream from high bitrates to low bitrates [4]-[6].

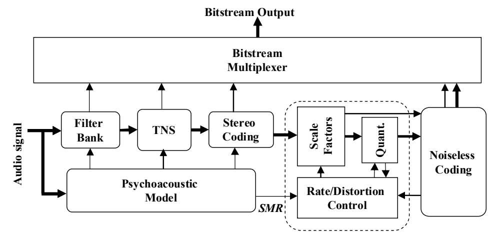


Figure 1. Block diagram of AAC-LC encoder

Based on the bitrate transcoding techniques, the application Shoutcast [7] constructed Internet radio stations where the audio signals are usually compressed at high bitrates prior to low bitrates transmission. In addition, the application AllofMp3 [8] that started with a 384 kbps MP3 archive and recently updated to lossless source files presents an online music store. AllofMp3 also supports the "Preview" and "Online encoding" functionalities. "Preview" enables consumers to listen to the music before buying. Consumers can download a low quality sample or listen immediately in a streaming audio format. "Online encoding" offers download options of bitstream formats and varying qualities at specified bitrates. The audio sequences are compressed as MP3 bitstreams at 192 kbps by default. Thus, the bitrate transcoding techniques can offer audio streaming of optional quality levels over heterogeneous networks to meet the user demands and price. At a given bitrate, the issue is to retain the maximal rate-distortion performance of transcoded bitstreams with the lowest computation power and the least storage space.

To trade the rate-distortion performance for real-time transmission, a novel and fast NMR optimized transcoding algorithm is adopted in Fast Rate-Distortion Optimized Transcoder (FRDOT). The performance was evaluated on a transcoding platform based on MPEG-2/4 AAC with Low Complexity (LC) profile [9]. Since AAC has been widely applied in compression for PC-based and portable devices, compression for terrestrial digital audio broadcast, streaming of compressed media for both Internet and mobile telephone channels [10]. In addition, AAC is also the core of High Efficiency AAC (HE-AAC) that has been applied to both satellite-delivered digital audio broadcast and mobile telephony audio streaming [10]. LC-AAC can retain a high quality with considerably reduced memory and power requirement [11]. Figure 1 shows main modules of an AAC LC Profile encoder. The results showed that the FRDOT can retain good coding performance with less complexity as compared to the cascaded transcoding that provides the best rate-distortion performance with the highest computation cost.

2 BITRATE ADAPTATION TRANSCODERS

To deliver audio signals over heterogeneous networks and to the devices with varying capabilities, the bitrate adaptation coding schemes are required. With single bitstream for each

1-4244-0921-7/07 $25.00 © 2007 IEEE.

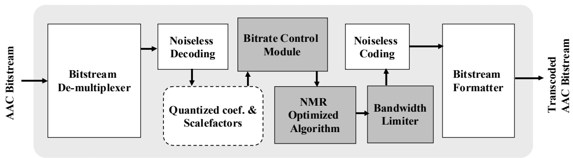


Figure 2. Block diagram of the FRDOT transcoder.

audio sequence, the bitrate scalability can be done by directly convert the compressed bitstream from higher bitrates to lower bitrates in the bitrate scaling transcoders [4]-[6].

The cascaded transcoder that inherits all modules from an AAC single layer coding system can get listening quality with a great amount of computation cycles.

The re-quantization reflecting psychoacoustic models can obtain the best transcoding quality when the uncompressed audio sequences are used as the input of psychoacoustic model [4]. As the input becomes the reconstructed audio signals, the masking thresholds are not accurate to prevent the audible distortion at the transcoder. In addition, the re-quantizing process that iteratively decreases the allocated bits may still take lots of computation time.

SLAT [5] can retain the listening quality with rapid transcoding algorithms. The three combined algorithms still have a few implementation issues. In the beginning, the bandwidth is reduced with the ρ-domain model, but the low bound of bandwidth reduction is difficult to set at different target bitrates. The scalefactor increment is in proportion to the bits to reduce. To meet the target bitrate, it will re-quantize the coefficients to zero when the number of bits to reduce is larger than a specified threshold after the bandwidth reduction. The re-quantization violated the original hypothesis that the minimum re-quantized coefficients should be bounded to the unity. The violation avoids the transcoding from converging to the target bitrate rapidly.

3 FAST RATE-DISTORTION OPTIMIZED TRANSCODER (FRDOT)

At the architecture of FRDOT in Figure 2, to address the high complexity issues of cascaded transcoder and to retain the better audible quality, the time-consuming modules like PAM and the quantization are removed. In addition, a bitrate control module (BCM) is used to meet the target bitrate. To improve the rate-distortion performance, a novel NMR optimized algorithm with a bandwidth limiter is presented. The bandwidth limiter iteratively will discard the non-zero quantized coefficients from the highest scalefactor band to the lowest scalefactor band until the allocated bit budget is met.

3.1 Bitrate Control Module (BCM)

Based on inter frame relationship, the BCM can dynamically estimate the total number of bits to be reduced every frame at lower bitrates. At a given bitrate, the estimation for each scalefactor band is based on the difference of the target bitrate and the original bitrate and the total amount of bits that are reduced at the previous scalefactor bands.

$$Bd_{frame,n} = Bd_{frame,n-1} + \Delta Bit_{n-1}, \tag{1}$$

where the symbols $Bd_{frame,n}$ and $Bd_{frame,n-1}$ mean the amount of the bits to reduce at the n-th and the $(n-1)$-th frames respectively. ΔBit_{n-1} presents the amount of reservoir bits at the $(n-1)$-th frame.

$$\Delta Bit_{n-1} = Bit_{used,n-1} - Bit_{budget}, \tag{2}$$

where the symbol $Bit_{used,n-1}$ means the actual number of used bits at the $(n-1)$-th frame, which is calculated by the noiseless coding after the bit reduction of transcoding algorithms.

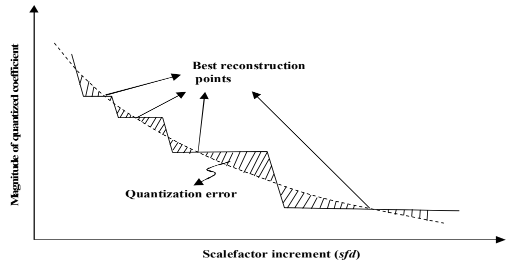


Figure 3. Illustration of relationship between the re-quantized coefficient and the scalefactor increment.

Bit_{budget} means the bit budget of current frame at a target bitrate R_{tar}.

$$Bit_{budget} = R_{tar} \times 1024 / sampling\_rate \tag{3}$$

The number of saving bits in the n-th frame is evaluated by

$$Bd_{frame,n} = (R_{ori} - R_{tar}) \times BFAC, \tag{4}$$

where the symbol R_{ori} indicates the original bitrate. The symbol $BFAC$ is defined as

$$BFAC = \frac{Bit_{ori} - Bit_{tar}}{R_{ori} - R_{tar}}. \tag{5}$$

The observations showed that the magnitudes of $BFAC$ are in a small rage of [9.3..9.8]. The $BFAC$ is updated by (6) to avoid underestimation and overestimation of the saving bits.

$$BFAC_n = \begin{cases} 9.5 + \Delta BFAC, & \text{if } \Delta Bit_{n-1} > \text{threshold} \\ 9.5 - \Delta BFAC, & \text{if } \Delta Bit_{n-1} > \text{threshold} \\ 9.5, & \text{otherwise} \end{cases} \tag{6}$$

Thus, by (2) and (4), (1) is changed to

$$Bd_{frame,n} = (R_{ori} - R_{tar}) \times BFAC_{n-1} + (Bit_{used,n-1} - Bit_{budget}) \tag{7}$$

3.2 NMR Optimized Transcoding Algorithm

For fast rate-distortion optimized content adaptation, the NMR optimized AAC transcoder will simultaneously address the issues including the maximal listening quality in NMR, the low complexity, and the small memory bandwidth.

3.2.1 Rationale

The difference of scalefactors between the transcoded and original bitstreams is called as a scalefactor increment. The final scalefactor is encoded by lossless coding module as the side information. Therefore, to reduce the bits of side information, the range of scalefactor increments shall be limited. In addition, the final scalefactor increment will affect the total amount of bits to encode each MDCT coefficient and the reconstructed magnitudes of audio samples. Thus, rate-distortion performance of rate adaptation transcoding is strongly dependent on the scalefactor increment of each band.

Figure 3 shows the relationship between the scalefactor increment and the quantization error. With different scalefactor increments to quantize specified MDCT coefficients, the dotted curve and the stair-like curve present the levels of quantized coefficient before and after rounding respectively. The dashed area indicates the quantization error of an input coefficient with varying scalefactor increments. As the scalefactor increment is enlarged from the minimum to the maximum, the levels of the quantized coefficients are decreased. In addition, for any input coefficient, applying a small set of scalefactor increments will generate the quantized coefficients of same level. The same quantization level will take an identical number of bits into the transcoded bitstream after AAC lossless coding. At the decoder side, applying different scalefactors to an identical quantization level, the reconstructed values are varied, which indicates the different quantization error and listening quality of the transcoded bitstreams. Thus, the best scalefactor increment can be found at the best reconstruction point of each input coefficient.

3.2.2 NMR-based Rate-Distortion Optimization

In the cascaded transcoder, the NMR values are derived from the psychoacoustic model. The masking thresholds that are derived using compressed audio signals at the archived bitstreams are not proper for perceptual audio coding. Thus, we speed up the transcoding by using embedded information within the input bitstreams. The NMR-based rate-distortion optimization can derive the best scalefactor increment (sfd_{best}) under the NMR criterion at a given bitrate.

The noise-to-masking ratio is to quantify the energy of audible noise. For the same audio source, smaller NMR values indicate higher audible quality of the reconstructed audio signals. NMR can be represented by a signal-to-masking ratio (SMR) minus a signal-to-noise ratio (SNR). The SNR value is measured with the audio source waveforms and the reconstructed audio waveforms. The SMR value is derived by the psychoacoustic model with the audio source signals. Thus, for the same audio source and the same psychoacoustic model, the SMR difference is set as zero and the NMR degradation can be formulated as a function of the SNR difference for bitrate adaptation transcoding.

$$\begin{aligned} NMR_{\deg rade} &= NMR_o - NMR_i = (SMR_o - SNR_o) - (SMR_i - SNR_i) \\ &= (SNR_i - SNR_o) = f(SNR_o), \end{aligned} \tag{8}$$

where the subscripts 'o' and 'i' indicate the output and input bitstreams respectively. As the SNR_o approaches to the SNR_i, the audible qualities of reconstructed signals from the input bitstream and the output bitstream are close. Based on the SNR criterion, the schema of NMR optimized algorithm for each frame are shown in Figure 4.

For complexity reduction, the quantized output coefficients are re-quantized without inverse quantization. In AAC coding system, the SNR values of the output bitstreams can be derived from the quantization formula of AAC standard.

$$q_i = \mathrm{int}\left(x \cdot 2^{-\frac{3}{16}\cdot sf_i} + 0.4054\right), \quad x = |mdct\_line|^{\frac{3}{4}}, \tag{9}$$

where the symbol x is the compressed MDCT coefficient. q_i and sf_i represent the i-th quantized coefficient and the i-th scalefactor of input bitstream respectively. The operator 'int' means to cast the remainder. '*mdct_line*' is the MDCT coefficient. When q_i is inversely quantized, the reconstructed value of the MDCT coefficient x_{Ri} is $q_i \cdot 2^{-\frac{3}{16}\cdot sf_i}$.

When the bitrate of input bitstream is reduced, the quantized coefficient is decreased from q_i to q_o and the scalefactor quantity is increased from sf_i to sf_o. In (10), q_o is obtained by re-quantizing the reconstructed MDCT coefficients x_{Ri}. The re-quantization can simply be done by applying a scalefactor increment to the original quantized coefficients without the inverse quantization.

$$\begin{aligned} q_o &= \mathrm{int}\left(x_{R_i} \cdot 2^{-\frac{3}{16}\cdot sf_o} + 0.4054\right) = \mathrm{int}\left(q_i \cdot 2^{-\frac{3}{16}\cdot(sf_o - sf_i)} + 0.4054\right) \\ &= \mathrm{int}\left(q_i \cdot 2^{-\frac{3}{16}\cdot sfd} + 0.4054\right) \end{aligned} \tag{10}$$

With the scalefactor increment, the SNR is defined by

$$\begin{aligned} SNR &= 10 \times \log\left(\frac{|mdct\_line|_{R_i}}{|mdct\_line|_{R_i} - |mdct\_line|_{R_o}}\right)^2 = 10 \times \log\left(r^2\right), \\ r &= \frac{x_{R_i}^{\frac{4}{3}}}{x_{R_i}^{\frac{4}{3}} - x_{R_o}^{\frac{4}{3}}} = \frac{q_i^{\frac{4}{3}} \cdot 2^{\frac{1}{4}\cdot sf_i}}{q_i^{\frac{4}{3}} \cdot 2^{\frac{1}{4}\cdot sf_i} - q_o^{\frac{4}{3}} \cdot 2^{\frac{1}{4}\cdot sf_o}} = \frac{\left(q_i \cdot 2^{-\frac{3}{16}\cdot sfd}\right)^{\frac{4}{3}}}{\left(q_i \cdot 2^{-\frac{3}{16}\cdot sfd}\right)^{\frac{4}{3}} - q_o^{\frac{4}{3}}} \end{aligned} \tag{11}$$

The suffixes R_i and R_o mean the reconstructed signals of input and output bitstreams respectively.

By substituting the variables at (9) into (10), the SNR values of the output signals can be derived by a function of the quantized input coefficients q_i and the scalefactor increment sfd. Given an input coefficient q_i, we can make an observation on the correlation between SNR and sfd according to (11). The observation shows that the SNR values may decrease as the magnitude of sfd increases.

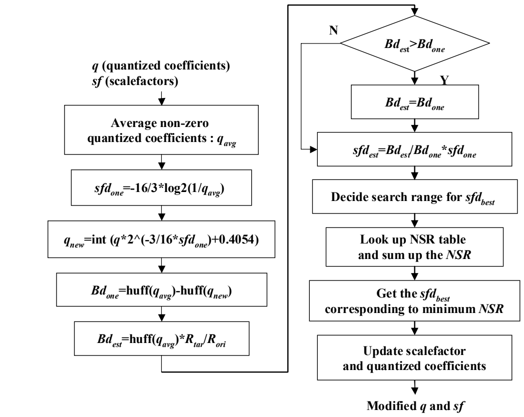


Figure 4. Flow chart of NMR optimized transcoding algorithm

NSR Lookup table **Defined search range=3-10**

sfd	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
q_i=1	66	72	77	80	34	34	34	34	34	34	34	34	34	34	34	34
q_i=2	66	73	70	66	55	62	70	75	79	80	80	80	80	80	80	80
q_i=3	66	65	46	64	73	71	67	58	60	70	75	79	80	80	80	80
q_i=4	66	51	63	66	55	62	70	71	68	61	55	68	74	78	80	80
q_i=5	61	54	63	42	66	64	39	65	73	71	67	56	61	70	75	79

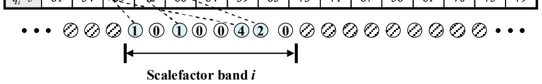


Figure 5. An illustration of the re-quantization process

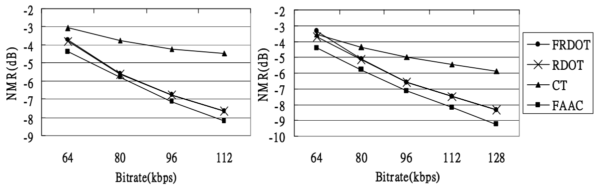
To be consistent with the basic principle of NMR measure, the SNR measure is converted to NSR measure.

$$NSR \equiv round\left(10 \times \log\left(\frac{\left(q_i \cdot 2^{-\frac{3}{16}\cdot sfd}\right)^{\frac{4}{3}} - q_o^{\frac{4}{3}}}{\left(q_i \cdot 2^{-\frac{3}{16}\cdot sfd}\right)^{\frac{4}{3}}}\right)^2 + 80\right), \tag{12}$$

An offset is added to make positive NSR values. With (12), given a quantized coefficient q_i from the input bitstream, the NSR value of re-quantized coefficient q_o at the output bitstream can be computed by applying the scalefactor increment sfd. With a specified range of sfd, the minimum NSR is found at the intersection of the dotted line and the curve in Figure 3.

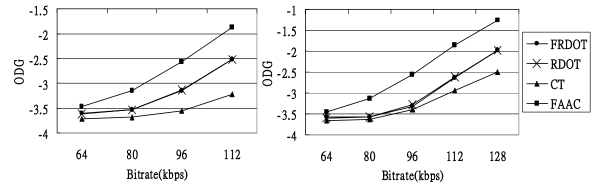
Based on the NSR, we can get the best scalefactor increment for every scalefactor band by exhaustively searching for sfd_{best} from all legal values defined in MPEG-2/4 AAC standard. The search strategy is sped up by the fast NSR-based search algorithm which reduces the overall complexity using the starting point prediction, the reduced search range and the rapid quality comparison for varying sfd. The starting point is predicted based on the bitrates of the original bitstream and the converted bitstream. The search range is reduced based on the distribution of NSR values that are obtained by simulating AAC transcoding at varying bitrates. The re-quantization and quality comparison are done by a table lookup technique.

The sfd_{est} is derived as a proper quantization step-size to meet the allocated bits. We can make an assumption that the neighboring sfd values of sfd_{est} will fit the bit allocation requirement. Based on the distribution of local NSR minimum and the best reconstruction point with an estimated level of the quantized coefficients, the neighbors of sfd_{est} form a range to search for a better sfd in the NSR table constructed by (12). The best sfd can induce the least NSR.



(a) 128 kbps (b) 160 kbps

Figure 6. Averaged NMR values with the input bitstreams.



(a) 128 kbps (b)160 kbps

Figure 7. Averaged ODG values with the input bitstreams.

The sfd_{best} is found within an asymmetric search range. To get sfd_{best} having the locally minimal NSR value at the search range, the NSR values of non-zero quantized coefficients q_i are summed up for each scalefactor band. As illustrated in Figure 5, there are four non-zero quantized coefficients, q_i= 1, 1, 4, 2. The search range is from 3 to 10. As sfd is 5, the NSR summation is 178, which is the minimum within the specified search range. Thus, the sfd_{best} which is 5 is added to the scalefactor of the input bitstream to get the final scalefactor.

The bitrate reduction by the NSR-based search algorithm is started from the highest scalefactor band and terminated when the target bitrate of processing frame is achieved. To meet with the specific bitrate with the smallest NSR value, the table lookup technique can reduce the processing time.

The size reduction of the lookup tables is done based on statistical distribution of quantized input coefficients. Since most of the quantized MDCT coefficients at AAC compressed bitstreams are small even at high bitrates. To investigate the magnitude histogram of quantized coefficients, the faac [12] and the two test sequences from European Broadcasting Union (EBU) are used. Our observations showed that more than 98% of non-zero MDCT coefficients are quantized to magnitudes smaller than 5 at high bitrates of 128 kbps and 160 kbps. Therefore, the index of q_i is only constructed from 1 to 5 to save the memory requirement of the NSR look-up table.

For rapidly converging to the target bitrate, some coefficients equal to one are allowed to be quantized to zero, as long as the averaged value of quantized coefficients in a scalefactor band is larger than one. When sfd_{est} is set than 5 with large coefficients in the scalefactor band, diminished NSR values are summed up to the NSR summation. Thus, the neighboring quantized coefficients with magnitude equal to the unity (q_i=1) can be zeroed out by re-quantization with least degradation of audio quality because there are stronger masking effects caused by the coefficients of large amplitudes.

4 EXPERIMENTS

For fair comparison, all transcoders are built on FAAC and FAAD [12]. The eight stereo sequences of sampling rate 44.1 kHz from EBU are encoded at 128 kbps and 160 kbps as the achieved bitstreams. The four methods including the FAAC encoder 'FAAC', the cascaded transcoder 'CT', NMR-based RDO algorithms 'FRDOT' and 'RDOT' are compared. The 'RDOT' adopted multiple iterations in searching for the best scarefactor increment. The results in TABLE I showed that the averaged bitrate of transcoded bitstreams by the BCM of FRDOT is closer to the target bitrate than the CT and the FAAC. The listening qualities of transcoded bitstreams are measured by NMR and ODG (Objective Difference Grade) [13]. The execution time is measured by Microsoft Visual C++ 6.0 profiling tool.

In Figure 6 and Figure 7, the rate-distortion curves show that the averaged NMR and ODG values of the FRDOT are better than those of the CT. When the bitrate of the input bitstreams increases from 128 kbps to 160 kbps, the quality loss by the FRDOT is less than 0.5 dB in NMR. As shown in TABLE II, the FRDOT can speed up the CT by about 6-7 times. The improvement of the FRDOT over the CT is increased at the high target bitrates.

TABLE I. Difference ratio of the transcoded bitrate over the target bitrates of FRDOT, CT and FAAC.

Target bitrate	FRDOT(%)	CT(%)	FAAC(%)
64 kbps	1.81	1.30	2.95
80 kbps	0.51	-0.72	1.84
96 kbps	0.01	-0.30	-2.01
112 kbps	-1.33	0.39	1.46

TABLE II. Averaged execution time of FRDOT and CT (transcoding from 128kbps to varying target bitrates).

Target bitrate	FRDOT(msec)	CT(msec)	CT/FRDOT
64 kbps	2285.54	13983.92	6.14
80 kbps	2342.84	15276.45	6.53
96 kbps	2298.24	16220.91	7.07
112 kbps	2254.16	16967.67	7.55

5 SUMMARY AND CONCLUSION

This paper has presented a fast bitrate transcoding algorithm called as the FRDOT for real-time audio delivery application. The FRDOT transcoder is to retain better quality at a given bit budget under the NMR criterion by finding the best scalefactor increment. The NMR value is represented by the NSR value to measure the distortion without the source audio signals. Based on the NSR values, a fast search algorithm of the best scalefactor increment at a given bitrate is done by the table lookup technique. In addition, the BCM can provide a model to estimate the bit difference for spectral coefficients between the original and target bitrates. With the BCM, the averaged bitrate of transcoded bitstream is close to the target bitrate.

REFERENCES

[1] J. Zhou and J. Li, "Scalable audio streaming over the Internet with network-aware rate-distortion optimization," in *Proc. IEEE ICME*, vol. 2, Aug. 2001, pp. 567-570.

[2] H. Park, et al., "Multi-layer bit-sliced bit rate scalable audio coding," *AES 103rd Convention*, New York, Aug. 1997.

[3] A. Kassler, et al., "Generic QoS aware media stream transcoding and adaptation, " in *Proc. Packet Video Workshop*, Apr. 2003.

[4] Nakajima. Y, et al., "MPEG audio bit rate scaling on coded data domain," in *Proc. ICASSP*, vol. 6, May 1998, pp.3669-3672.

[5] C. Y. Lee, et al., "Efficient AAC single layer transcoder," *AES 117th Convention*, San Francisco, Oct. 2004.

[6] Mat Hans, et al., "An MPEG audio layered transcoder," *AES 105th Convention*, San Francisco, Aug. 1998 (preprint 4812).

[7] http://www.shoutcast.com

[8] http://www.allofmp3.com

[9] ISO, *Information technology–Coding of Audio-Visual Objects*, 1999. ISO/IEC JTC1/SC29, ISO/IEC IS 14496-3 (Part 3, Audio).

[10] S. Quackenbush, "MPEG technologies: advanced audio coding," *ISO/IEC JTC1/SC29/WG11*, Nice, FR, Oct. 2005.

[11] Y. Takamizawa, et al., "High-quality and processor-efficient implementation of an MPEG-2 AAC encoder," in *Proc. ICASSP*, vol. 2, May 2001, pp. 985-988.

[12] http://www.audiocoding.com

[13] Draft ITU-T Recommendation BS.1387 "Method for Objective Measurements of Perceived Audio Quality, " July 2001,.