MPEG AUDIO BIT RATE SCALING ON CODED DATA DOMAIN

Y. Nakajima, H.Yanagihara, A.Yoneyama, and M.Sugano

KDD Co.Ltd., R&D Labs.
2-1-15, Ohara,
Kamifukuoka, Saitama 356, JAPAN

ABSTRACT

Formerly, once the audio data is compressed, transcoding is used to scale the bit rate, where decoding and re-encoding are taken place. Therefore, data manipulation of coded data has been very complex and time consuming work. In this paper, we describe three algorithms for bit rate scaling on coded MPEG data domain. One is bandwidth limitation method cutting higher frequency components until target data rate is satisfied. The other two use re-quantization process where a quantization step in each subband is modified. One of them reflects psychoacoustic model from bit allocation information obtained in the bitstream in order to improve bit rate scaling efficiency. The simulation results show that re-quantization process provides very high conversion efficiency and nearly equal sound quality to direct coding one can be obtained by reflecting psychoacoustic model. It is also shown that very fast scaling(factor of six) have been achieved when compared with transcoding method.

1. INTRODUCTION

MPEG is used in a wide variety of fields such as communication, broadcasting, and storage applications. For example, MPEG1 is used in Video CD for such an application as Karaoke and CD-ROM for PC applications[1,2]. As for MPEG2, wider application fields are expected, where some of them have already been practiced in digital satellite/terrestrial broadcasting and DVD(Digital Video Disc)[3,4]. Recently audio coding algorithms for software implementation and lower bit rate have been studied actively[5-12].

Since required audio-video quality totally depends on application, available network bandwidth, and terminal capability, the coding algorithm and coding bit rate are selected according to these conditions and requirements. For example, in broadcasting applications using MPEG2, studio quality video is coded at more than 10 Mbit/s in order to maintain high quality audio-video throughout the transmission in between stations, whereas distribution-quality video is coded at 4-6Mbit/s in order to achieve cost effective multi-channel transmission using 6MHz bandwidth. In intranet applications, high quality audio-video corporate information is stored as master data and lower bit rate AV contents are distributed through the intranet. This is mainly due to the reason that the available bandwidth is limited and lower quality than that of broadcast TV may be acceptable as long as the contents are comprehensive. In order to correspond to these heterogeneous network environments, several methods can be applied. One is the use of transcoding where compressed audio-visual information at bit rate $R1$ is once decoded to obtain pixel level video and PCM level audio data at $R0$ ($>>R1$), and then re-encoding at target bit rate $R2$ ($<R1$) is performed(Figure 1). Another method is the use of MPEG scalability functions based on hierarchical coding architecture(Figure 2). For example, MPEG2 SNR scalability provides two-level coding bit rate as well as picture quality. In this case, decoding of base layer gives standard quality whereas the combination of enhancement layer in decoding provides higher quality.

Although these methods can provide multi-bit rate streams, each method has its own drawbacks. Transcoding usually requires a storage media with large capacity in order to store the reconstructed audio-visual information in digital form. Otherwise, a high quality analogue VCR is required for the storage. In addition, MPEG encoder is necessary to re-encode the reconstructed audio-video. The quality of re-encoded audio-video also becomes a crucial issue. Several research works have been reported in terms of coding conditions for transcoding over multi-generations[13,14]. As for scalability, since the number of layers is predetermined in the encoded bitstream and its number is at most three, application using this function would be limited to such objective as graceful degradation for transmission error[3]. In addition, the scalability functions are provided only for MPEG2 video data. Therefore, neither MPEG1 audio-video nor MPEG2 audio can use such capabilities. Assume that 4Mbit/s MPEG2 audio-video consists of 3.6Mbit/s video and 256kbit/s audio. Even if MPEG2 video can provide base layer at 1/10th(=360kbit/s) of overall video stream using the scalability function, total bit rate becomes more than 600kbit/s since no scalability function is provided for audio.

In recent literature, we have proposed an algorithm for bit rate scalability of MPEG video using re-quantization process[15]. We have shown that bit rate conversion on quantization stage can provide good conversion efficiency when compared with transcoding. In the conversion, adaptive re-quantization is performed taking into account of local and global activities of picture.

In this paper, we propose novel algorithms for bit rate scaling of MPEG audio on coded data domain. The proposed techniques for bit rate scalability exploit bit rate reduction on subband domain. Here, we propose three types of coding rate conversion using MPEG1 layer II audio. Firstly, we propose bandwidth limitation method where bandwidth of subband is limited until target bit rate is met. Secondly, re-quantization method is introduced where re-quantization is performed for every subband using a rate control. Thirdly, re-quantization method reflecting psychoacoustic model is proposed for higher conversion efficiency. Simulation has been performed for several test sequences coded by MPEG1 Layer II.

2. BIT RATE SCALING

2.1 Bandwidth limitation

Figure 3 shows basic diagram of MPEG1 audio layer I/II decoder. As can be seen from the figure, neither psychoacoustic model nor bit allocation process can be known in the decoder. Therefore, it is necessary to exploit these characteristics on coded data domain in order to perform the efficient bit rate scaling. The simplest way to achieve the scaling is bandwidth limitation. Since audio information of lower frequency band is more important than that of higher one for human ear, audio signal compressed at the lower bit rate reconstructs a lowpass version of the audio signal[2,3]. Figure4(a) illustrates the bandwidth limitation method. The horizontal axis shows subband whereas the vertical axis shows bit allocation. Firstly, the target bit *Tbf* for each frame is calculated according to the target bit rate. Then zero bits are assigned to all the samples in the highest subband with nonzero bit allocation(Figure4(a)-(ii)). The total bit in the frame *Cbf* is calculated including adjustment of the bit count of side information such as scalefactor. If the amount of bit count is still larger than the target *Tbf* then this low pass filtering procedure is iterated for the next highest subband until *Cbf* becomes smaller than *Tbf*(Figure4(a)-(iv)). The new bit-rate scaled bit stream is created removing corresponding subband information. This method is very simple in implementation point of view since no subband synthesis/analysis process is required nor inverse/re-quantization process is required. However, when the bit rate scaling factor is very large, for example, scaling from 128kbit/s per channel to 32kbit/s, very limited bandwidth can be remained since bit count of a few subbands may cause a large amount of bit due to fine quantization. This will lead to degradation of conversion efficiency and hence lower audio quality than that of the audio coded from PCM at the same bit rate.

2.2 Re-quantization

By incorporating re-quantization process, more efficient scaling can be achieved than the previous method without drastic sacrifice of audio bandwidth. Figure 4(b) shows the bit rate scaling flow by re-quantization process. In this method, instead of zero bit allocation from higher subbands, coarser re-quantization is performed from the higher subbands. For the highest subband with nonzero bit allocation, one bit is subtracted for each sample(Figure 4(b)-(ii)). Then bit count in the frame *Cbf* is calculated reflecting bit count for bit allocation data. If *Cbf* is larger than *Tbf* then further re-qunatization is performed at the next lower subband with nonzero bit allocation. This process is repeated until *Cbf* becomes smaller than *Tbf*(Figure 4(b)-(iv)). Note that when *Cbf* is still larger than *Tbf* at the lowest subband, another iteration is performed from the highest subband. When the optimum bit allocation is found where *Cbf* is just below the *Tbf*, all the samples in subbands whose bit allocations are modified are inverse quantized and re-quantized according to the new bit allocations. Then the bit-rate scaled bitstream is created using the re-quantized subband samples and new bit allocation information. Figure 5 shows the block diagram of bit rate scaling using re-quantization process. In this process much higher bit rate scaling efficiency than the bandwidth limitation method can be expected since balanced bit count reduction from each subband is achieved.

2.3 Re-quantization reflecting psychoacoustic model

One of the key technologies that plays an important role in very high compression coding of MPEG audio is the use of perceptual coding[2,3]. In the course of MPEG encoding, it removes components that are perceptually irrelevant to the ear. In the encoder, after subband analysis, the psychoacoustic model calculates the masking threshold and SMR(Signal to Mask Ratio) for each subband. Then in bit-allocation process, bits are allocated to the subbands according to the masking threshold provided by the psychoacoustic model. Each subband sample is quantized following the allocated bits and transmitted to the multiplexer. Since the psychoacoustic model is required only in the encoder and audio data can be reconstructed without the knowledge of this model, the decoder is made less complex. However for the optimum bit rate scaling point of view, the information directly obtained in the decoder is not sufficient and psychoacoustic model needs to be reflected in the bit rate scaling process.

We have focused on the iterative encoding process controlled by dynamic bit-allocation rule and exploited the algorithm of re-quantization reflecting psychoacoustic model. In the encoder, bit allocation process performs minimization of the NMR(Noise to Mask Ratio) in each subband. In each iteration step quantizer step is increased to produce the smaller value of NMR(= SMR-SNR). This iteration is repeated as long as bit-allocation is available. After the iteration, NMR for each subband becomes almost the same[5]. Therefore, by re-quantizing subband samples toward equal increase of NMR for each subband, it is possible to achieve bit rate scaling reflecting psychoacoustic model. Although no information on absolute NMR nor SMR value can be obtained in the decoder, relative values can be used for the calculation.

Figure 4(c) shows bit rate scaling flow by re-quantization process reflecting psychoacoustic model. Firstly all the NMR values in the subbands are set to zero. Then quantization step of the subband that has minimum increase of NMR by the decrease of bit allocation is decreased by one step and the total bit in the frame *Cbf* is calculated. Then this process is repeated until *Cbf* becomes smaller than the target bit *Tbf*(Figure 4(c)-(iv)). Since this process nearly emulates MPEG audio encoding procedure, highest conversion efficiency among these three algorithms can be expected.

3. SIMULATION RESULTS

We have used several music sequences(jazz, classic, and rock music) to evaluate the bit rate scaling performance. These sequences are at first coded by MPEG1 Layer II at 112kbit/s per channel. Then the encoded streams are converted to the bit-rate scaled bitstreams using three algorithms mentioned above. Transcoding is also performed where MPEG coded 112kbit/s stream is once decoded and re-encoded at the reduced bit rates. In addition, these music sequences are directly coded at the reduced rates from PCM. Some of the simulation results are shown in Figure 6 using rock music sequence. Figure 6(a) shows the reconstructed waveform from 64kbit/s MPEG audio directly coded from PCM. Figure 6(b) shows the reconstructed waveform from the same bit rate MPEG audio that is bit-rate scaled from 112kbit/s using bandwidth limitation method. When compared with Figure 6(a), it can be seen that only low frequency components are retained and high frequency components are removed(around 100th and 200th sample) in Figure 6(b). From the subjective point of view, the sound by the bandwidth limitation method can be clearly distinguished from the one directly coded from PCM. We have also found that the sound by the bandwidth limitation method converted to 56kbit/s has about the same subjective quality as the one directly coded at 32kbit/s. Figure 6(c) is the case of re-quantization method reflecting psychoacoustic model. It can be seen from the figure that almost the same waveform as Figure 6(a) can be obtained. It is also found that subjectively almost no difference are perceived in all of the test sequences. As for the re-quantization method, although very small difference are found in the waveform comparison with the case of direct coding, subjectively slight degradation in high frequency component are found in the re-quantization method. As for the transcoding when compared with direct coding, there is no noticeable difference in both waveform and subjective test. This is mainly due to the reason that transcoding is performed in only one generation.

Table 1 shows computation time for the simulated methods using one of the test sequences with 90sec. The simulation is performed using about 140MIPS workstation. As can be seen from the table, very fast processing of bit rate scaling can be realized in the proposed three algorithms. It is also found that more than six times faster processing is possible in the method of re-quantization reflecting psychoacoustic model when compared with transcoding.

4. CONCLUSION

In this paper we have presented audio bit rate scaling algorithms on coded data domain from MPEG audio bitstream. Although bandwidth limitation can be simply implemented, degradation of sound quality is large when compared with directly coded audio at the same reduced bit rate. By incorporating re-quantization process, significant gain has been obtained in bit rate scaling efficiency. Although the reflection of psychoacoustic model to the re-quantization provides slight improvement at low bit rate region, it is found that almost the same quality as the one coded directly from PCM can be obtained. In addition, processing time has been greatly reduced to 1/6th of the transcoding method.

5. ACKNOWLEDGEMENT

Authors would like to express thanks to Dr.H.Murakami and Dr.M.Wada for their continuous support and also wish to thank to Mr.T.Aoyagi and Mr.H.Takayama for the computer simulation.

6. REFERENCES

[1] ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5Mbit/s," 1993.

[2] D.Pan, "A Tutorial on MPEG/Audio Compression," *IEEE Multimedia*, pp.60-74,Vol.2,No.2,1995.

[3] B.G.Haskell, A.Puri and A.N.Netravali, "Digital Video : An Introduction to MPEG-2," Chapman & Hall, 1997.

[4] P.Noll, "MPEG Digital Audio Coding," *IEEE Signal Processing Mag.*,pp.59-81, Sept.1997.

[5] M.Kumar and M.Zubair, "A High Performance Software Implementation of MPEG Audio Encoder," *Proc. IEEE ICASSP*, pp.1049-1052, 1996.

[6] L.Mainard and M.Lever, " A Bi-dimensional Coding Scheme Applied to Audio Bitrate Reduction," *Proc.IEEE ICASSP*, pp.1017-1020, 1996.

[7] S.Boland and M.Deriche, "Audio Coding Using the Wavelet Packet Transform and a Combined Scalar-Vector Quantization," *Proc. IEEE ICASSP*, pp.1041-1043, 1996.

[8] K.N.Hamdy, M.Ali and A.H.Tewfik, "Low Bit Rate High Quality Audio Coding with Combined Harmonic and Wavelet Representations," *Proc. IEEE ICASSP*, pp.1045-1048, 1996.

[9] D.Sinha and J.D.Johnston, "Audio Compression at Low Bit Rates Using a Signal Adaptive Switched Filterbank," *Proc. IEEE ICASSP*, pp.1053-1056, 1996.

[10] W.K.Dobson, J.J.Yang, K.J.Smart, and F.K.Guo, " High Quality Low Complexity Scalable Wavelet Audio Coding," *Proc. IEEE ICASSP*, pp.327-330, 1997.

[11] R.Lefebvre and C.Laflamme, "Spectral Amplitude Warping (SAW) for Noise Spectrum Shaping in Audio Coding," *Proc. IEEE ICASSP*, pp.335-338, 1997.

[12] S.Bolan and M.Driche, "New Results in Low Bitrate Audio Coding Using a Combined Harmonic-Wavelet Representation," *Proc.IEEE ICASSP*, pp.351-353, 1997.

[13] M Pereira, and A.Lippman, "Re-codable video," *Proc. IEEE ICIP* , vol.II, pp.952-956, 1994.

[14] A.T.Erdem and M.I.Sezan, "Multi-generation Characteristics of the MPEG Video Compression Standards," *Proc. IEEE ICIP*, Vol.II, pp.933-937, 1994.

[15] Y.Nakajima, H.Hori, and T.Kanoh, "Rate Conversion of MPEG Coded Video by Re-quantization Process," *Proc. IEEE ICIP*, Vol.III, pp.408-411, 1995.

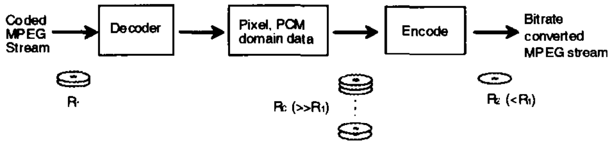


Figure 1 Transcoding

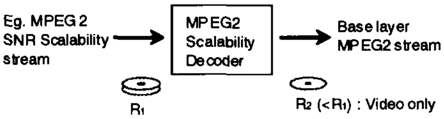


Figure 2 MPEG2 scalability

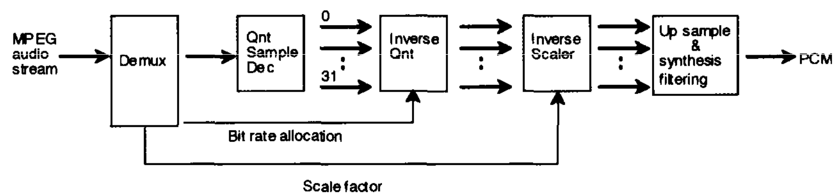


Figure 3 MPEG-1 audio layer I/II decoder

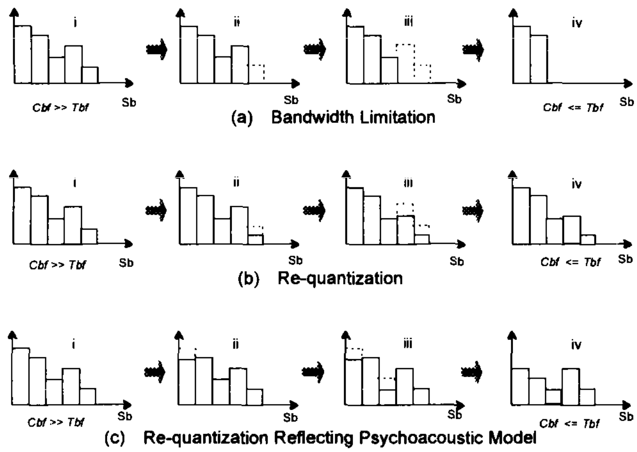
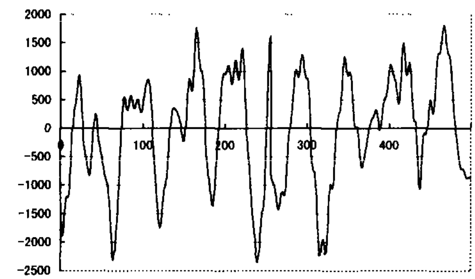


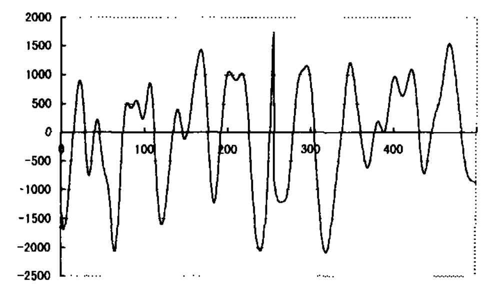
Figure 4 Bit rate scaling on coded data domain

Table 1 Execution time

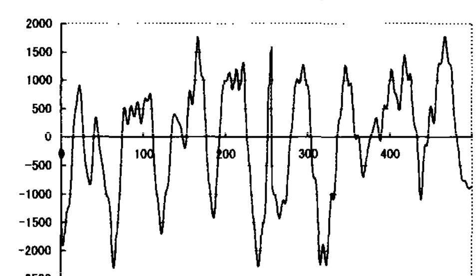
Method	Direct	BW limit	Re-Qnt	Re-Qnt + Psych.	Trans-coding
Time(sec)	91	11	13	16	105



(a) PCM to 64kbit/s coded audio



(b) 112kbit/s to 64kbit/s conversion by bandwidth limitation



(c) 112kbit/s to 64kbit/s conversion by re-quantization reflecting psychoacoustic model

Figure 6 Simulation results

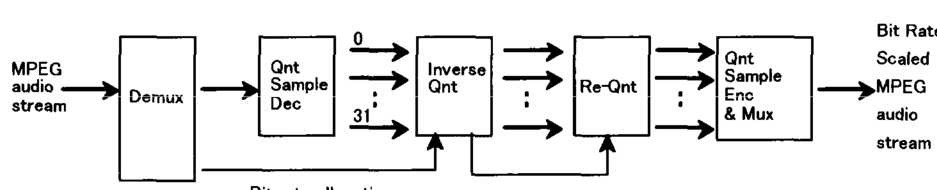


Figure 5 Block diagram of bit rate scaling using re-quantization