Conversion of MP3 to AAC in the Compressed Domain

Koichi Takagi Satoshi Miyaji Shigeyuki Sakazawa Yasuhiro Takishima
KDDI R&D Labs. Inc.
Fujimino-shi, Saitama 356-8502 Japan

***Abstract*— In this paper, we propose an efficient algorithm for conversion of an MP3 stream into AAC. Generally, this kind of conversion, transcoding, requires full decoding and re-encoding. However, the re-encoding based transcoding process may result in degradation of quality and take longer than encoding from a PCM signal. This paper proposes a method to inherit the frame structure and quantization scale from MP3 to AAC. This enables a reduction in the iteration process, which requires the most time in the AAC encoding process, without incurring degradation in quality. Experimental results show that the proposed method can perform bitstream domain transcoding at high speed while maintaining a high level of audio quality.**

Keywords—MP3; AAC; Transcoding; Audio conversion; Scale factor; Subband

Topic area—Compressed domain processing

I. INTRODUCTION

Since an audio decoder can easily be implemented even in a small mobile handset, the digital audio market has been deployed all over the world. As the market grows and technology advances, a variety of audio compression schemes have been developed. Regarding the international standard, MP3 (MPEG-1/2 Audio Layer 3) [1, 2], which was standardized by ISO MPEG, is widely used. MP3 is one of the most popular audio compression formats, and almost all portable playback devices are capable of MP3 decoding.

On the other hand, advanced audio coding (AAC) [3, 4], which is not compatible with MP3, is being used and has penetrated into online music distribution, broadcasting services, etc. AAC can encode a CD quality audio signal of 48–64 kbit/s per channel, and the encoding efficiency is 30% higher than MP3. A wide range of AAC products and AAC related services are expected to appear on the market. As a result, demand for converting existing MP3 files into AAC files will increase.

The easiest way to convert MP3 to AAC is uncompressed domain transcoding. This means that an MP3 stream is fully decoded to PCM and then encoded into AAC. However, the uncompressed domain transcoding method is time consuming and quality will degrade. Transcoding applies two different kinds of encoders to audio data, degrading the quality of the data. Reference [6] says that when an MPEG video stream is transcoded, the quality of the MPEG video stream is degraded. Thus, there are many reports on how to minimize the degradation in quality caused by MPEG video transcoding. The same thing is true of audio transcoding, even in transcoding for bit rate conversion where the encoding method is not changed.

Therefore, in this paper, we propose a novel algorithm for conversion of MP3 to AAC in the compressed data domain with the least quality degradation. We analyzed the encoding parameters of MP3 and AAC, which consists of side information, scale factors, and MDCT coefficients, and then tried to figure out the differences and similarities. Consequently, we realized faster conversion of MP3 to AAC compared with the simple transcoding technique.

This paper is organized as follows. Section II provides an overview of the related techniques and outlines the proposed method. In Section III, performance of the proposed algorithm is compared with the existing transcoding technique. The last section concludes this paper.

II. RELATED WORKS

In this section, related works regarding audio data conversion are described, followed by a technical overview of MP3 and AAC.

A. *Related works*

A variety of techniques for digital audio data conversion have been studied to date from several aspects. Regarding uncompressed audio data, sampling frequency conversion techniques have long been studied extensively, such as by [5]. Most of the techniques are based on Shannon's sampling theorem. Furthermore, there have been many reports that address signal conversion of digital compressed video and digital compressed audio [6-8]. In these papers, for example, bit rate conversion, frame rate conversion, and coding format conversion were conducted among the encoding schemes of MPEG-2, MPEG-4, MPEG-4 and AVC/H.264. In addition, [8] focuses on the MP3 bit rate conversion technique and proposes two methods based on elimination of higher frequency components and re-quantization. However, this method is only used with the audio conversion application where the encoding scheme does not change. As mentioned above, the direct conversion method involving two or more audio codecs has never been studied.

B. *MP3 and AAC*

This section provides a technical overview of MP3 and AAC.

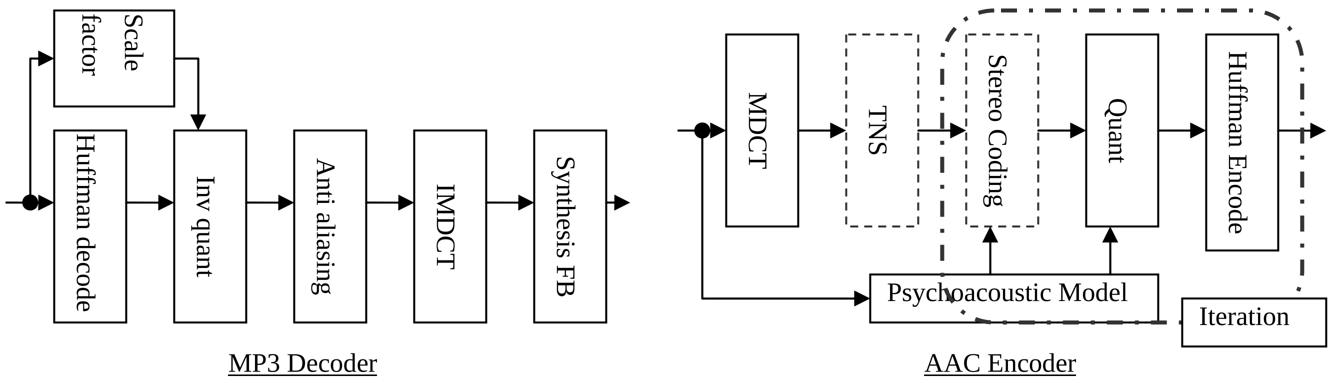

Figure 1. Structures of MP3 decoder and AAC encoder.

MP3, the MPEG-1/2 Layer III standard, introduced many features. The decoder structure is shown in the left side of figure 1. In order to achieve a higher frequency resolution closer to the critical band partitions, the 32 subband signals are further subdivided in the frequency domain by applying a 6-point (short) or 18-point (long) modified DCT (MDCT) with 50% overlap in each subband. The 6-point MDCT is applied when expected pre-echoes occur.

AAC (Advanced Audio Coding) was originally designed to obtain a higher quality compression coding than MP3. Most of the technologies used in AAC were carried over from MP3. The encoder structure is shown in the right side of figure 1. One of the major reasons for the improvement in quality from AAC is the newly developed MDCT filter bank. The AAC filter bank adaptively varies in length between 128 and 1024 frequency bins (length 256 and 2048 point windows, with a 50% overlap). For an audio signal, a high-frequency resolution filter bank is often required. MP3 has 576 bins even in the highest resolution mode, whereas the longer MDCT transform of 1024 frequency bins are allowed in AAC. However, high-frequency resolution is not desired during transients of the audio signal. Both MP3 and AAC have the capability to change the window length adaptively to a shorter length during transients. In this case, MP3 has a frequency resolution of 192 bins, compared to 128 bins for AAC. Because AAC does not use a cascaded Quadrature Mirror Filter/MDCT filter bank employed in MP3, no cross-talk aliasing reduction is required.

As mentioned above, one of the major differences between MP3 and AAC is the transform method. Furthermore, transform sizes are also different. For all of these reasons, it is very challenging to convert MP3 to AAC in the compressed domain.

For encoding MP3 and AAC, the coded bits are generally assigned using the same psychoacoustic model. Moreover, the time required for the bit allocation processes occupies a significant portion of the encoding processes because this process uses iteration. Conversely, reducing the processing time for this operation can improve the overall speed of audio encoding. Therefore, it is strongly desirable that all AAC variables determined on the iteration process be calculated from the MP3 parameters.

III. Proposed Method

In this section, we will describe the similarities in the quantization scales of both MP3 and AAC. And then, the details of the proposed conversion method will be explained, which provides audio transcoding from MP3 into AAC at high speed and with high audio quality.

A. *Preliminary experiment*

One of the greatest differences between MP3 and AAC is the structure of the filter bank as described previously, whereas the basic block diagram of the decoder is similar. Therefore, we focused on the re-quantization process placed immediately before the filter bank in the decoder. The re-quantization coefficient of MP3 is calculated as follows by the use of the scale factor:

$$xr[i] = is[i]^{\frac{4}{3}} \times 2^{\frac{1}{4}A} \tag{1}$$

$$A = \text{global\_gain}[gr] - 64 - 8\,\text{subblock\_gain}[w][gr]$$
$$-2(1+\text{scalefac\_scale}[gr])(\text{scalefac}[cb][w][gr] + \text{preflag}[gr]\,\text{pretab}[cb])$$

where $xr[i]$ and $is[i]$ are the quantized MDCT coefficient and MDCT coefficient, respectively. i, gr, w, and cb denote the MDCT coefficient index, granule index, window index, and codebook index, respectively. The re-quantized coefficient of AAC also uses the scale factor and is calculated as follows:

$$xr[i] = is[i]^{4/3} \times 2^{\frac{1}{4}(\text{scale\_factor}[g][sfb]-100)} \tag{2}$$

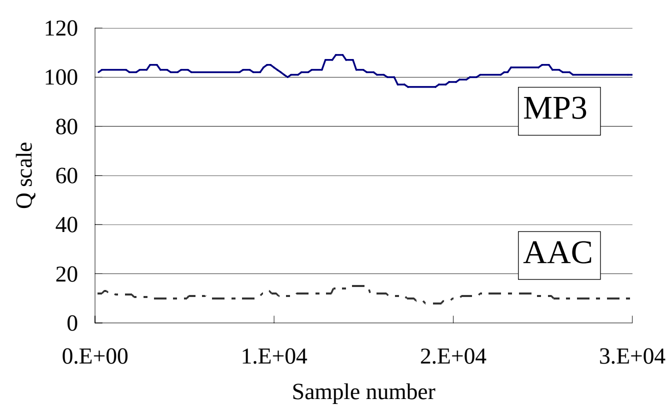

Figure2. Quantization scale for MP3 and AAC

where *g* and *sfb* mean the window group and scale factor band, respectively. The distribution of these quantized scales, which denotes the exponent parts of 2 in eq. (1) and (2), are measured and shown in figure 2. There is a correlation between the quantization scale of MP3 and AAC, although their absolute values are different.

B. *Proposed method*

Based on the results of the preliminary experiments discussed previously, we propose an MP3 to AAC transcoding method where the quantization scale is shared by both encoding methods. Details of the proposed algorithm are described as follows:

1. (Initialization) N-frame MP3 data are stored in the buffer. The size of the transformation unit for MP3 is 1152 [samples] and for AAC is 1024 [samples]. Because the least common multiple is 9216 [samples], let N = 8 (=9216/1152). This resolves the problem of the difference between the MP3 and AAC frame sizes.

2. MP3 data in the buffer are extracted and decoded. The dequantization scales are preserved as Q[*j*], where *j* indicates sample number. Frame structures (short or long) are also maintained as frame[*k*], where *k* means frame number.

3. AAC MDCT is applied. When the frame structure is a short block, where 192 bins have been selected with MP3 at the corresponding time, the frame is determined as transient and the short window is selected for the corresponding frame. This is illustrated in figure 3. AAC frames which includes the most part of each short block S in MP3 frames adopt short window S'.

4. The number of coding bits is assigned according to Q[*i*], which is derived from the MP3 data as a quantization scale. Furthermore, minor adjustments are applied based on the psychoacoustic model.

5. Apply Huffman encoding and then output the AAC data.

6. If the data in the buffer are sufficient for the process, the next N-frame MP3 data are provided to the buffer, and then procedures 2 through 6 are applied. The conversion process is complete when there are no more MP3 data to convert.

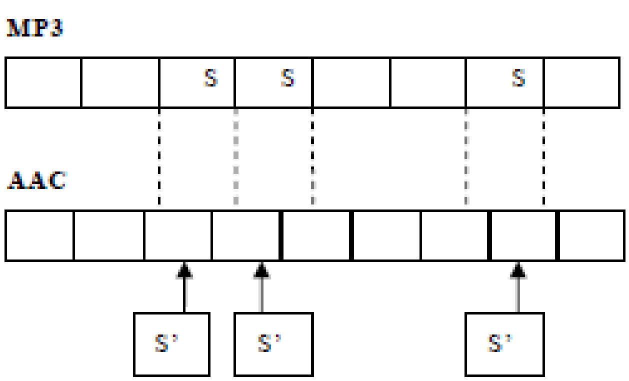

Figure 3. Frame structure inherritance from MP3 to AAC

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we performed a quality evaluation of the AAC data converted from MP3 using the proposed method (a) comparing the bitstream generated by the uncompressed domain transcoding method (b) and the data directly encoded from the uncompressed PCM for reference purposes (c). The computational complexity of each method was also evaluated. Figure 4 illustrates how data (a), (b) and (c) were prepared.

The experiments were executed on a PC with Intel Mobile Pentium 1.5 GHz, 512 MB RAM, and Windows XP. Two audio music data segments of 300 seconds (pop and rock) were used for the test. MP3 was encoded and decoded with the LAME codec [10], and AAC was encoded with the ISO AAC encoder [11]. Encoding bit rate combinations for MP3 and AAC were (MP3, AAC) = <A>(128 kbps, 96 kbps), <B>(128 kbps, 64 kbps), and <C>(96 kbps, 64 kbps). All bitstreams were 44.1 kHz, mono.

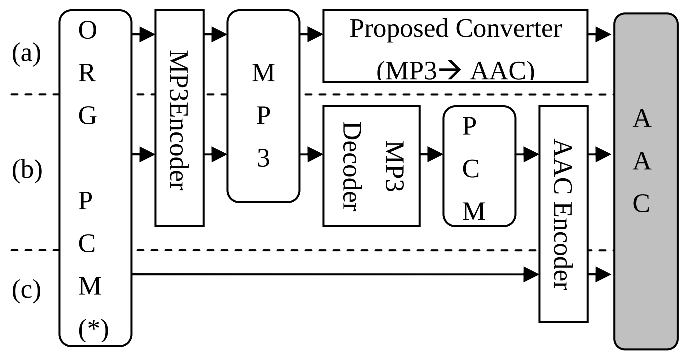

Figure 4. Used AAC streams for evaluation.

A. *Audio quality*

To evaluate the audio perceptual quality of each method, a subjective assessment test [12] was performed with 10 experimental subjects. First, the subjects listened to the original PCM data (corresponding to (*) in Figure 4). Then, they listened to two data (original data/evaluated data) segments, where they did not know which were the evaluated ones, and then rated the perceived difference between the two data segments. The results of the comparison are shown in table I. In this table, the range of values is 1 to 5. Higher values indicate less degradation, where the encoded bitstream is perceptually close to the original PCM data. For example, "4.00" means "perceptible, but not annoying."

TABLE I. SUBJECTIVE ASSESSMENT RESULTS

Condition	(a)Proposed	(b)Transcoding	(c)Org→AAC
<A>	4.10	3.75	4.25
<B>	3.80	3.55	4.00
<C>	3.75	3.25	

As mentioned above, when the encoding process is applied twice, audio quality is lower than when encoding is applied only once. Therefore, case (c) achieved the best results in this experiment. However, the quality of proposed method (a) is almost equivalent to case (c) and obtains excellent improvement compared with (b).

To examine the quality difference in detail, waveforms of (a), (b), and (c) in case <B> are shown in figure 5. The difference between (a) and (c) is smaller than between (b) and (c). This can be proven by the results of PSNR measured for all the audio data, which are (a) 30 dB, (b) 21 dB, and (c) 35 dB in case <B>.

B. Conversion Speed

The conversion speed from MP3 to AAC is compared between the proposed method (a) and the uncompressed domain transcoding method (b), and the results are shown in table II.

This table shows that the proposed method (a) can significantly reduce the time required for conversion compared with transcoding method (b) employing full-decoding and re-encoding.

TABLE II. CONVERSION SPEED

Condition	(a)Proposed	(b)Transcoding
<A>	42[sec]	74[sec]
<B>	40[sec]	72[sec]
<C>	40[sec]	70[sec]

V. CONCLUSION AND FUTURE WORK

This paper reports on an investigation of an efficient audio transcoding method. For MP3 and AAC, we focused on the difference in frame sizes and the similarity in the characteristics of quantization scales and designed a fast conversion method without degrading audio quality. Experimental results demonstrated that the proposed method was better than the full-decoding and re-encoding method. Conversion of MP3 to other formats requires further study.

ACKNOWLEDGMENT

The authors would like to thank Dr. Shigeyuki Akiba, Dr. Shuichi Matsumoto, and Dr. Yasuyuki Nakajima for their continuous support and encouragement.

REFERENCES

[1] ISO/IEC 11172-3:1993, "Coding of moving pictures and associated auido and digital storage media at up to about 1.5Mbit/s, Part3: Audio," 1993.

[2] ISO/IEC 13818-3:1998, "Generic coding of moving pictures and associated audio, Part 3: Audio," 1998.

[3] ISO/IEC 13818-7:2006, "Generic coding of moving pictures and associated audio, Part 7: Advanced Audio Coding," 2006.

[4] ISO/IEC 14496-3:2005, "Information technology — Coding of audio-visual objects — Part 3: Audio," 2005.

[5] S. Park, G. Hillman, R. Robles, "A novel structure for real-time digital sample-rate converters with finite precision error analysis," *Proc. IEEE ICASSP91*, Vol.5, pp. 3613 -3616, 1991.

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

[7] J Zhang, A Perkis, N Georganas, "H. 264/AVC and Transcoding for Multimedia Adaptation," *Proc. the 6th COST*, 2004.

[8] Y. Nakajima, H. Yanagihara, A. Yoneyama, M. Sugano, "MPEG Audio Bit Rate Scaling on Coded Data Domain," *Proc. IEEE ICASSP98*, Vol.6, pp.3669-3672, 1998.

[9] T. H. Tsai, C. Yen, "A high quality re-quantization / quantization method for MP3 and MPEG-4 AAC audio coding," *Proc. IEEE ISCAS*, Vol.3, pp. 851-854, 2002.

[10] LAME Project, "http://lame.sourceforge.net/"

[11] ISO/IEC 14496-5: 2001, "Information technology — Coding of audio-visual objects — Part 5: Reference software," 2001.

[12] ITU-R BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

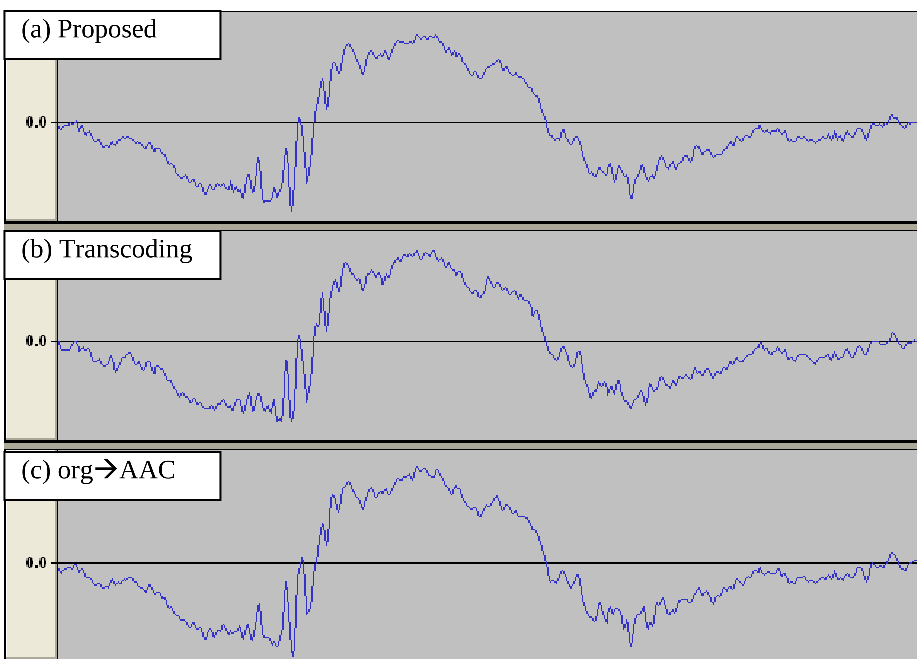

Figure.5 Waveforms (horizontal axis: time, vertical axis: amplitude)