A SCALABLE LOSSY TO LOSSLESS AUDIO CODER FOR MPEG-4 LOSSLESS AUDIO CODING

[1]*Rongshan Yu,* [1]*Xiao Lin,* [1]*Susanto Rahardja and* [2]*C. C. Ko*
[1]Institute for Infocomm Research (I^2R), A*STAR,
21 Heng Mui Keng Terrace, Singapore 119613, email: *{rsyu, linxiao, rsusanto}@i2r.a-star.edu.sg*
[2]Department of Electrical and Computer Engineering,
National University of Singapore, Singapore 119260, email: elekocc@nus.edu.sg

ABSTRACT

In this paper, we present Advanced Audio Zip (AAZ), a scalable lossless audio coding technology that was recently selected as the Reference Model for MPEG audio scalable lossless coding (SLS) work. AAZ provides excellent compression performance while delivering fine grain bit-rate scalability from lossy to lossless coding. Moreover, AAZ provides backward compatible to the MPEG Advanced Audio Coding (AAC) system by embedding AAC compliant bit-stream into the lossless bit-stream. As a result, AAZ serves as an universal coding solution with functionalities that were previously offered by several distinct audio coding technologies such as lossless audio coding, perceptual audio coding, or scalable audio coding; and maximizes the interchangeability for digital audio contents migrating among these application domains.

1. INTRODUCTION

During the last two decades, audio in digital formats has been widely used in numerous applications, and it has essentially replaced its analog counterparts with its unprecedented quality and flexibility. However, its high data rates, if uncompressed, could be a heavy burden to many applications with constrained bandwidth and/or storage resources such as network audio streaming or portable audio players. For example, CD-quality (44.1 kHz sampled, 16 bit quantization) digital audio signal can easily consume bandwidth up to 705.6 kbps per audio channel or 1.41 Mbps for a stereo pair. In response to such a need, considerable research has been devoted to the development of audio compression technology and there have been some fruitful results. Nowadays, many audio compression algorithms, such as the MPEG Advanced Audio Coding (AAC) technology [1], can deliver CD-quality stereo audio at bit-rates from 96~192 kbps, or at 10~20 times compression compared with the uncompressed audio. Many of these high compression audio coding algorithms can be classified into the category of perceptual audio coding [2], where the masking properties of the human auditory system are employed to minimize the perceptibility of signal distortion due to the compression.

Clearly, the perceptual audio coding technology described above is not suitable for applications with lossless quality requirement like audio archiving, studio, and high-end consumer electronic applications. For these applications, lossless audio compression technology [3] may be the only solution as it preserves every bit in the original audio after compression. Unlike perceptual audio coding, which usually performs signal compression in the time/frequency domain since the masking property can be best represented as a time/frequency phenomenon, many lossless audio coding algorithms use time domain predictive coding due to its simplicity, and can only achieve very limited compression ratio of 1.5 ~ 3.0.

It is expected that, with advances in computer, broadband networking and storage technologies, there will be more and more digital audio applications that could provide high quality audio services with high sampling rate, high amplitude resolution (e.g., 96 kHz, 24 bit/sample) audio and lossless quality. Meanwhile, there will still be many applications that require high-compression digital audio. Therefore, it can be foreseen that both lossy and lossless audio compression technologies will continue to prevail in their particular application domains in the predictable future. As such, a solution that provides interchangeability across these two application domains would greatly simplify the problem of migrating audio contents in these domains, and facilitate the transition from lossy to lossless digital audio service. For these reasons, the international standard body ISO/IEC JTC1/SC29/ WG11 (MPEG) has recently issued a Call for Proposal (CfP) on lossless audio compression [4] to invite contributions for such a scalable solution, and the AAZ technology presented in this paper was evaluated and selected as Reference Model for this work [5].

In order to achieve the desirable scalability, AAZ adopts the integer Modified Discrete Cosine Transform (IntMDCT) that was previously described in [6] for lossless audio coding. This is coupled with an embedded entropy coder, namely Bit-Plane Golomb Code (BPGC) [7], to generate Lossless Enhancement (LLE) bit-stream on top of an MPEG AAC core bit-stream which then provides the scalable lossy to lossless coding. In order to efficiently embed the core MPEG AAC compliant bit-stream, an error mapping process is employed so that only the residual IntMDCT spectrum, obtained by removing the information that has already coded in the core bit-stream from the original one, is coded in the LLE bit-stream. Moreover, for best perceptual quality at intermediate rates when the LLE bit-stream is truncated, the order of BPGC coding is selected in such a way that the spectral shape of the quantization noise of the core AAC bit-stream, which has been perceptually optimized

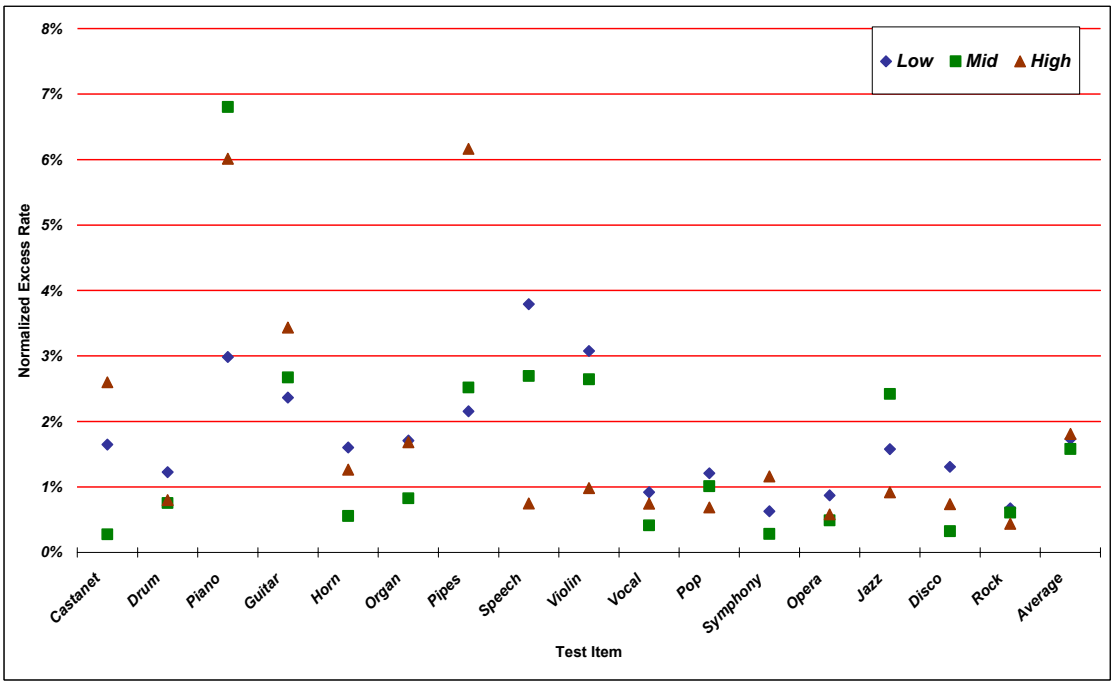

Fig 1. The relative redundancy of BPGC coding of IntMDCT transformed audio due to the divergence of its actual distributions from the Laplacian Model.

by the AAC encoder, is always preserved in the coding process.

2. STATISTICAL STUDY OF INTMDCT TRANSFROMED AUDIO

It is well known that digital samples from many real-world sources, such as audio [8], speech [9], and video/image signals [10][11], exhibit peaky distributions that are approximately Laplacian (two-side Geometric) and their amplitude can be closely described with the Geometric probability distribution function (pdf): $\Pr(i) \stackrel{\Delta}{=} P_\theta(i) = \theta^i(1-\theta)$, where i are positive integers, and $\theta \in (0,1)$ is the distribution parameter.

In AAZ, the BPGC, an embedded coder that provides near optimal coding/truncating performance for sources with Laplacian distribution [7], is selected for scalable coding of the IntMDCT coefficients by assuming that they are closely Laplacian distributed. To further justify this selection, the redundancy due to this Laplacian assumption is investigated by estimating the excess rate resulted from the divergence of the actual distribution of IntMDCT transformed audio from the Laplacian pdf.

To elaborate, we denote the Laplacian distribution and the sample distributions measure from the real IntMDCT spectral data by P_θ and Q, respectively. It is well-known that the minimum redundancy R for a code ϕ that is designed for P_θ but is used to encode samples with distribution Q is bounded by the divergence between theses sources:

$$R \geq D(Q \| P_\theta) = E_Q\left[\log_2 \frac{Q(i)}{P_\theta(i)}\right]. \quad (1)$$

Consequently, the relative redundancy r, obtained by normalized R with the empirical entropy of that source, $H(Q) = E_Q[-\log_2 Q(i)]$, cannot be less than

$$r \geq \frac{D(Q \| P_\theta)}{H(Q)}. \quad (2)$$

In our experiment, we evaluate the relative redundancy r for 16 CD quality audio sequences that covers a wide range of music sources such as instrumental, pop, rock, classical, speech. The duration of each audio sequence is 10,240 samples and the window length of IntMDCT is 2,048. For each audio sequence, the relative redundancy is computed, for three different frequency regions, namely, Low Band (0 – 2.3 kHz), Mid Band (8.3 – 11.0 kHz), and High Band (16.5 – 19.3 kHz). The results are given in Fig. 1, which strongly justify the entropy coder selection as it is evidenced that in most cases the relative redundancy of this Laplacian assumption is in fact trivial (<4%), and the average redundancy for all the testing audio files is only around 2%. In addition, these results further imply that additional improvement in coding performance by replacing the BPGC with more complicated encoding mechanism may be very limited in the AAZ framework presented in this paper.

3. STRUCTURE OF AAZ

In this section, a brief description of the structure of AAZ is presented. Interested readers can refer to [12] for a more comprehensive description of its implementation details.

The system diagram of the AAZ codec is given in Fig 2, which comprises of two distinguished layers in both the encoder and decoder, namely, a core layer and lossless enhancement (LLE) layer. In particular, the core layer is simply an MPEG AAC audio codec [3]. Both layers are found in the AAZ encoder and decoder.

In the AAZ encoder, the input audio are losslessly transformed frame by frame to generate the IntMDCT coefficients $c(k)$, where $k = 1,...,1024$, which are fed to the core layer AAC encoder to generate the core layer AAC bit-stream. In this AAC encoder, $c(k)$ are first grouped into scalefactor bands as defined in [1], which are then quantized with an AAC quantizer, usually with different quantization steps in different scalefactor bands to shape the quantization noise so that they can be best masked. The resultant quantized spectrum is Huffman coded and multiplexed with some side information, e.g., quantization step size in each scalefactor band, to produce the core layer bit-stream.

In order to efficiently employ the information of the spectral data that has been carried in the core layer bit-stream, the following error-mapping procedure is employed in the LLE layer to generate the residual spectrum $e(k)$:

$$e(k) = \begin{cases} c(k) & , \quad i(k) = 0 \\ |thr(k) - c(k)| & , \quad i(k) \neq 0 \end{cases}, \tag{3}$$

Here $thr(k)$ is the rounded quantization threshold (closer-to-zero) for $c(k)$ in the core layer AAC quantizer, and *i(k)* is the quantized IntMDCT spectral data vector produced by the AAC quantizer. Clearly, for $c(k)$ that has already been significant in the core layer, (i.e., $i(k) \neq 0$), $e(k)$ is always non-negative and its sign does not need to be coded. In addition, if the amplitude of $c(k)$ is Geometric distributed owing to the "memoryless" properties of a Geometric pdf, the distribution skew of $c(k)$ will be roughly preserved (with reduced range) in the distribution of $e(k)$, which can thus be very efficiently coded by the BPGC.

The IntMDCT residual spectrum output *e(k)* is then bit-plane coded to generate the scalable LLE layer bit-stream. In order to preserve the spectral shape of the noise spectrum of the core layer encoder at intermediate bit-rates, as the first step, the Most Significant Bit (MSB) for spectral data from all scalefactor bands is coded. After that, the coding process is progressed to the 2nd MSB, 3rd MSB and so on until it reach the Least Significant Bit (LSB) for all scale factor bands. As a result, the level of the noise spectrum will be decreased

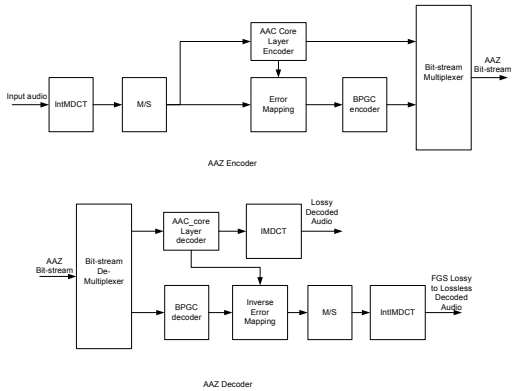

Fig 2. Structure for AAZ encoder and decoder

progressively during this coding process and its spectral shape will be roughly preserved, which is desirable to achieve a good coding margin vs. rate performance in the intermediate rates from lossy to lossless.

For each scalefactor band, the BPGC coding parameter *L* (lazy layer) is calculated according to [7]. After that, the bit-plane symbols are arithmetic encoded using the BPGC probability assignment rules [7] given below:

$$Q^L(j) = \begin{cases} \dfrac{1}{1+2^{2^{j-L}}} & j \geq L \\ \dfrac{1}{2} & j < L \end{cases}. \tag{4}$$

To further reduce the complexity, those bit-plane symbols resided lower than the lazy plane *L* are unprocessed and directly included into the compressed bit-stream without performing any arithmetic coding process since they are coded with equally distributed probability.

As a final step for the encoder, the output of LLE bit-stream is multiplexed with the core AAC bit-stream to produce the final lossless bit-stream.

4. PERFORMANCE

The compression performance of AAZ is evaluated using audio sequences that comprises of vocal, instrumental and symphony music clips that are sampled at 48 kHz, 96 kHz with 16, 20, and 24 bits/sample. As mentioned, the core-layer codec is an MPEG-4 AAC codec. The bit-rates for the core encoder are 64 kbps/channel for 48 kHz and 80 kbps/channel for 96 kHz sampling rates. For comparison, the latest version of Monkey's audio (ver. 3.97), a state of the art lossless audio codec, is used as the benchmark codec. The comparison results are given in Table 1, which clearly shows that AAZ only performs 5% worse compared to Monkey's audio despite its abundant functionalities.

In order to evaluate the lossy performance of AAZ, we further compared the Noise to Masking Ratio (NMR) [13] performance of the reconstructed audio signal when the LLE layer bit-stream is truncated to lower bit-rates.

Items	Monkey	AAZ
avemaria.wav	2.68	2.53
clarinet.wav	2.19	2.10
cymbal.wav	3.36	2.87
etude.wav	2.48	2.37
flute.wav	2.60	2.47
haffner.wav	1.87	1.82
violin.wav	2.15	2.07
Overall	2.40	2.27

Items	Monkey	AAZ
avemaria.wav	2.00	1.98
clarinet.wav	2.32	2.24
cymbal.wav	2.21	2.18
etude.wav	1.94	1.92
flute.wav	2.36	2.26
haffner.wav	2.03	1.98
violin.wav	2.18	2.11
Overall	2.14	2.09

Table 1. Comparisons of compression performance for AAZ and Moneky's Audio. (Top: 48kHz/16Bit audio; Bottom: 96kHz/24Bit audio)

Instead of comparing the average NMR performance, we measure the maximum NMR (MaxNMR) from all scalarfactor banks for each frame, which reflects the most perceptible distortion for a given time instance. Fig. 3 gives an example of the result, which shows the MaxNMR as a function of time for the first 100 audio frames of a piece of symphony. It is clear that the LLE layer improves the perceptual quality of the core AAC layer, with higher bit-rates in the LLE layer bit-stream consistently resulting in smaller MaxNMR for all the audio frames.

5. CONCLUSION

In this paper the structure of Advanced Audio Zip (AAZ), a scalable lossy to lossless audio codec is described. The AAZ has been recently selected as the Reference Model for the work of MPEG-4 audio scalable lossless (SLS) coding. The AAZ provides full range support of lossless audio compression, bit-rate scalability, and backward compatibility; and simulation results have shown that those functionalities are achieved with only marginal increase in the overall compression ratio performance compared with the state-of-the-art lossless only audio codec.

REFERENCES

[1] Information Technology – Coding of Audiovisual Objects, Part 3. Audio, Subpart 4 Time/Frequency Coding, ISO/JTC1/SC29/WG11, 1998.

[2] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio", *Proceeding of the IEEE*, April 2000

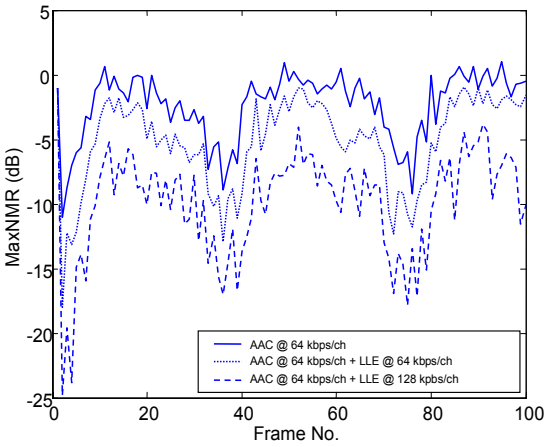

Figure 3. MaxNMR vs time performance for AAZ

[3] M. Hans and R. W. Schafer, "Lossless compression of digital audio", *IEEE Signal Processing Magazine*, pp. 21-32, July. 2001

[4] ISO/IEC JTC1/SC29/WG11 N5040, "Call for Proposals on MPEG-4 Lossless Audio Coding," Awaji Island, Japan, 2002.

[5] ISO/IEC JTC1/SC29/WG11 N5720, "Workplan for Audio Scalable Lossless Coding (SLS)," Trondheim, Norway, 2003.

[6] R. Geiger and *et al*, "Audio Coding based on Integer Transforms," 111th AES Convention, Sep. 2001.

[7] R. Yu, C.C. Ko, S. Rahardja and X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions," Proceeding of ICASSP 2003.

[8] P. Philippe, F. M. de Saint-Martin, M. Lever, "Wavelet Packet Filterbanks for Low Time Delay Audio Coding," IEEE Tran. Speech and Audio Procesing, vol. 7, no. 3, pp. 310 – 322, May, 1999.

[9] M. D. Paez and T. H. Glisson, "Minimum mean-squared-error quantization in speech PCM and DPCM systems," *IEEE Trans. Comm.*, vol. 20, pp. 225 – 230, 1972.

[10] E. Y. Lam and J. W. Goodman, "A Mathematical Analysis of the DCT Coefficient Distribution for Images," IEEE Trans. Image Processing, vol. 9, no. 10, pp 1661 – 1666, Oct., 2000

[11] J. Luo, C. W. Chen, K.J. Parker, T.S. Huang, "Adaptive quantization with spatial constraints in subband video compression using wavelets," Proceeding of International Conference on Image Processing, vol. 1, 1995.

[12] R. Yu, X. Lin, S. Rahardja and H. Huang, "Technical Description of I2R's Proposal for MPEG-4 Audio Scalable Lossless Coding (SLS): Advanced Audio Zip (AAZ)," ISO/IEC JTC1/SC29/WG11, MPEG2003/M10035, October 2003, Brisbane, Australia

[13] B. Beaton and *et al*, "Objective perceptual measurement of audio quality," in Collected papers on digital audio bit-rate reduction, pp. 126 – 152, AES, 1996.