IMPROVING CODING EFFICIENCY FOR MPEG-4 AUDIO SCALABLE LOSSLESS CODING

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

ABSTRACT

The recently introduced MPEG standard for lossless audio coding (MPEG-4 Audio Scalable to Lossless (SLS) coding technology) provides a universal audio format that integrates the functionalities of lossy audio coding, lossless audio coding and fine granular scalable audio coding in a single framework. In this paper, we propose two coding methods that improve the coding efficiency of the SLS, namely, the context-based arithmetic code (CBAC) method and the low energy mode code method. These two coding methods work harmonically with the current SLS framework and preserve all its desirable features such as fine granular scalability, while successfully improving its lossless compression ratio performance.

1. INTRODUCTION

During the last two decades, considerable efforts have been devoted to the development of audio compression technologies, whose primary objective is the transmission of high-quality digital audio over communication channels with limited bandwidth. These efforts have led to fruitful results. Nowadays, most audio compression algorithms, such as the MPEG-1 Layer III (mp3) [1] or MPEG-4 AAC [2], can deliver "perceptually transparent" CD quality (48kHz, 16bit/sample) audio at bit-rates from 48 ~ 64 kbps per channel, or 7 ~ 14 times compression compared with the uncompressed audio.

Recently, with advances in broadband access networking and storage technologies, an increasing number of digital audio applications have emerged to provide high quality audio services with high sampling rate, high amplitude resolution (e.g., 96 kHz, 24 bit/sample) audio at lossless quality. Meanwhile, there will still be many applications that require highly compressed digital audio. A solution that provides interchangeability across these two application domains would greatly simplify the problem of migrating audio content between these domains, and facilitate the transition from lossy to lossless digital audio service. In response to this need, the international standard body ISO/IEC JTC1/SC29/WG11, also known as the Moving Picture Experts Group (MPEG), has recently issued a Call for Proposal (CfP) [3] on lossless audio coding. Contributions for a solution for scalable to lossless (SLS) audio compression are invited. Subsequently, the Advanced Audio Zip [4] has been adopted as the Reference Model (RM) [5] for this work at the 65th MPEG in July 2003.

AAZ adopts the transform coding approach where the Integer MDCT (IntMDCT) [6], an integer implementation for the Modified Discrete Cosine Transform (MDCT) [7] is used as its filterbank for lossless coding. In AAZ, the IntMDCT spectral data of the input audio are first coded with an MPEG-4 AAC encoder to generate an embedded AAC bit-stream, while the residual spectrum between the IntMDCT spectral data and the their quantized value by the embedded AAC encoder is subsequently coded with Bit-Plane Golomb Code (BPGC) [8] to produce the fine granular scalable lossy to lossless portion of the final lossless bit-stream.

In this paper, we propose two coding methods that further refine the BPGC coding process to improve its coding efficiency for lossless compression of IntMDCT residual. The first one is Context-Based Arithmetic Code (CBAC), which uses context modeling technology to capture the statistical dependencies of the probability distribution of the IntMDCT residual signal on its frequency location, amplitude of its adjacent spectral lines, and the core layer AAC quantizer. The second one is the so-call low energy mode coding method, which is used to replace the BPGC process for IntMDCT spectral data from time/frequency (T/F) regions with extremely low energy level. Since these spectral data are dominated by the rounding errors of the IntMDCT algorithm, they have a probability distribution that is far away from the Laplacian distribution for which BPGC/CBAC would lead to suboptimal compression results. The structure of the MPEG-4 SLS codec integrated with the proposed CBAC and low energy mode coding methods is illustrated in Fig. 1.

The CBAC and the low energy mode coding methods have been previously proposed to MPEG as Core Experiments (CE) [11][12] to improve the performance of MPEG-4 SLS. It is found that they successfully improve the compression ratio performance of SLS, while preserving all its other features such as fine granular scalability. As a result, these technologies have been adopted by MPEG, and incorporated into the RM for this work.

2. BIT-PLANE CODING WITH CBAC

The BPGC coding process used in MPEG-4 SLS is basically a bit-plane coding scheme where the bit-plane symbols are arithmetic coded with a structural frequency assignment rule. Consider an input data vector $\mathbf{e} = \{e[0],...,e[N-1]\}$, for which N is the dimension of $\mathbf{e}$. Each element $e[k]$ in $\mathbf{e}$ is first represented in a binary format as

$$e[k] = (2s[k]-1)\cdot\sum_{j=0}^{M-1} b[k,j]\cdot 2^j, k = 0,...,N-1\text{,}$$

which comprises of a sign symbol

$$s[k] \overset{\Delta}{=} \begin{cases} 1 & ,e[k]\geq 0 \\ 0 & ,e[k]<0 \end{cases},\ k = 0,...,N-1$$

and bit-plane symbols $b[k,j]\in\{0,1\}$, $i = 1,...,k$. Here, M is the Most Significant Bit (MSB) for $\mathbf{e}$ that satisfies $2^{M-1}\leq \max\{|e[k]|\} < 2^M, k = 0,...,N-1$. The bit-planes symbols are then scanned and coded from the MSB to the Least Significant Bit (LSB) over all the elements in $\mathbf{e}$, and coded by using arithmetic code with a structural frequency assignment $Q(j)$ given by

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

where the Lazy plane parameter L can be selected using the adaptation rule

$$L = \min\{L'\in\mathbb{Z} \,|\, 2^{L'+1}N\geq A\}. \tag{2}$$

Here A is the absolute sum of the data vector $\mathbf{e}$.

Although the above BPGC coding process delivers excellent compression performance for data that are issued from independent and identically distributed (iid) source with Laplacian distribution [8], it lacks the capability to explore the statistical dependencies that may exist in certain sources to achieve better compression performance. These correlations can be very effectively captured by incorporating context-modeling technology into the BPGC coding process, where the frequency assignment for arithmetic coding of bit-plane symbols is not only depended on the distance of the current bit-plane to the Lazy plane parameter as in the frequency assignment rule (1), but also on other possible elements that may effect the probability distribution of these bit-plane symbols.

In the context of lossless coding of IntMDCT spectral data for audio, elements that possibly effect the distribution of the bit-plane symbols include the frequency locations of the IntMDCT spectral data, amplitude of the adjacent spectral lines, and status of the AAC core quantizer. In CBAC, these correlations are effectively captured by using several types of contexts. The guide in selecting these contexts is to try to find those contexts that are "most" correlated to the distribution of the bit-plane symbols. In addition, cares have to be taken to avoid overpopulating the number of the contexts, which may eventually deteriorate the coding efficiency performance as a result of modeling cost as well as introducing unnecessary implementation burdens.

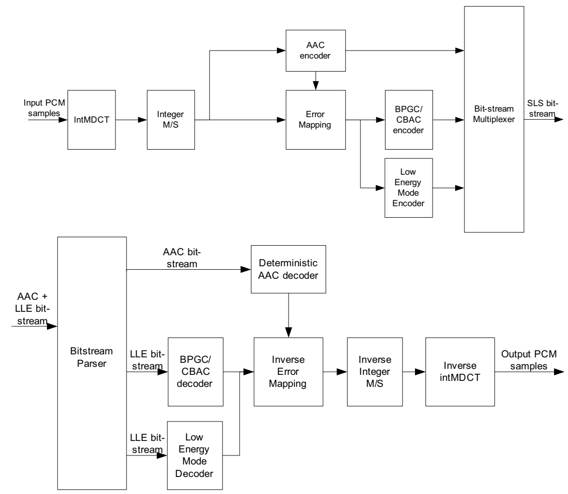

Fig. 1 Block diagram of the MPEG-4 SLS encoder and decoder integrated with CBAC and low energy mode coding

In CBAC, three types of contexts, namely, the *frequency band (FB)* context, the *distance to lazy (D2L)* context, and the *significant state (SS)* context are used. The detailed context assignments are summarized as follows:

- Context 1: *frequency band (FB)* – It is found in our experiments that the probability distribution of bit-plane symbols of IntMDCT varies for different frequency bands. Therefore, in CBAC the IntMDCT spectral data is classified, empirically, into three different *FB* contexts, namely, Low Band (0 ~ 4 kHz, *FB* = 0), Mid Band (4 kHz ~ 11 kHz, *FB* = 1) and High Band (above 11 kHz, *FB* = 2).
- Context 2: *distance to lazy (D2L)* – The D2L context is defined as the distance of the current bit-plane to the BPGC Lazy plane parameter *L*. The introduction of this context follows the same rationale behind the BPGC frequency assignment rule (1), which is based on the fact the skew of the probability distribution of the bit-plane symbols from a source with Laplacian or near-Laplacian distribution tends to decrease as the number of *D2L* decrease [8]. To reduce the total number of the *D2L* context, all the bit-planes with $D2L < -2$ are grouped into one context where all the bit-plane symbols are coded with probability 0.5 since the probability skew in these contexts is so small that the coding gain is negligible if they are arithmetic coded.

- Context 3: *significant state (SS)* – The *SS* context tries to group the factors that may correlate with the distribution of the amplitude of the IntMDCT residual in one place. These include the amplitude of the adjacent IntMDCT spectral lines and the quantization interval of the AAC core quantizer if it has previously quantized in the core encoder. The detailed configuration of the *SS* context can be found at [12].

3. LOW ENERGY MODE CODING

The BPGC/CBAC coding process described above provides excellent compression performance only for source with Laplacian or near-Laplacian distribution. However, in many audio signals, it is found that there exist some "silence" T/F regions, such as the high frequencies regions for some single instrumental music, or silence portion of the music, from which the IntMDCT spectral data are in fact dominated by the rounding errors accumulated from the rounding operation in the IntMDCT algorithm. The distribution of these rounding errors is far away from the Laplacian distribution, and hence the use of BPGC/CBAC results only in suboptimal compression performance. In order to improve the coding efficiency, we replace the BPGC/CBAC coding process with a different coding method, namely, the low energy mode coding for IntMDCT spectral data from those regions.

The low energy mode coding is performed on scale factor bands (sfb) for which the BPGC parameter L is smaller or equal to 0. Here L in effect signals the energy level of the IntMDCT spectrum. At low energy mode coding, the amplitude of the residual spectral data $e[k]$ is first converted into unitary binary string $\mathbf{b} = \{b[0], b[1], \ldots, b[pos], \ldots\}$ as illustrated in table 1:

Amplitude of $e[k]$	Binary string $\{b[pos]\}$
0	0
1	1 0
2	1 1 0
…	…
$2^M - 2$	1 1 … … … 1 0
$2^M - 1$	1 1 … … … 1 1
pos	0 1 2 3 …

Table 1. Binarization of IntMDCT error spectrum at low energy mode. Here M is the maximum bit-plane.

It can be seen that the probability distributions of these symbols are jointly determined by their position *pos*, and the distribution of $e[k]$:

$$\Pr\{b[pos] = 1\} = \Pr\{e[k] > pos \mid e[k] \geq pos\}$$
$$0 \leq pos < 2^M .$$

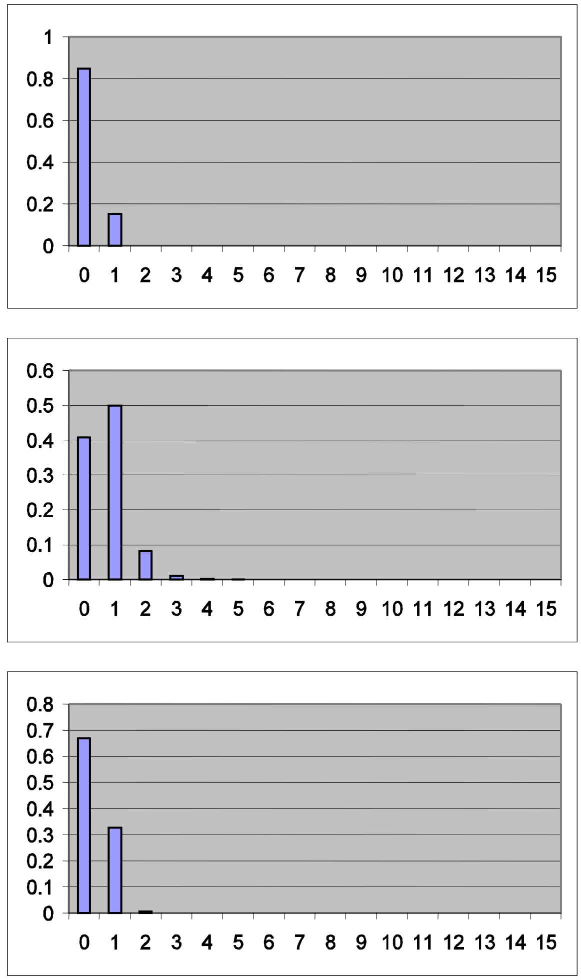

Fig. 2 Histogram of the IntMDCT residual signal for L=0 (Top), L=-1 (Middle), and L=-2 (Bottom) from a piece of audio training set.

In addition, the distribution of $e[k]$ is directly related to the Lazy plane parameter L (see Fig. 2 for an example). Therefore, in low energy mode coding $b[pos]$ is arithmetic coded conditioned on its position *pos* and the BPGC parameter L with a trained frequency table. The sign bit for non-zero $e[k]$ is also coded with probability 0.5 after its amplitude is coded.

4. PERFORMANCE

We compare the compression ratio performances of the MPEG-4 SLS codec and its improved version that integrates the CBAC and low energy mode coding methods. The audio sequences used in our comparison are from the MPEG lossless coding task group [3], which include two popular combinations of sampling rate and quantization word lengths, i.e. 48 kHz/16 bit and 96 kHz/24 bit. Detailed results are listed in Table. 2.

Evidently, these results suggest that the proposed technologies successfully improve the performance of the MPEG-4 SLS encoder, which have improved the compression ratio performance of MPEG-4 SLS encoder by 2.61% and 1.52% for 48/16 testing set and 96/24 testing set respectively.

5. CONCLUSION

As the latest member in the MPEG audio coding family, the MPEG-4 Audio Scalable Lossless (SLS) coding provides fine granular scalable lossy to lossless coding, a functionality that complements the existing MPEG audio coding tools. In this paper we introduce two coding methods to further improve the coding efficiency of MPEG-4 SLS. It is found that the proposed technologies can be readily integrated within the framework of the MPEG-4 SLS, leading to significantly improvement in terms of lossless compression ratio performance as shown in simulation results. The proposed technologies have been adopted by MPEG as successful Core Experiments for MPEG-4 SLS, and they have been incorporated with the RM for this work.

Item	**MPEG-4 SLS**	**Improved MPEG-4 SLS**	**Improve-ment**
avemaria	2.44	2.53	3.66%
blackandtan	1.73	1.75	1.49%
broadway	1.93	1.97	1.72%
cherokee	1.79	1.82	1.72%
clarinet	2.04	2.08	2.24%
cymbal	2.75	2.99	8.66%
dcymbals	1.59	1.60	1.22%
etude	2.30	2.37	2.92%
flute	2.38	2.46	3.04%
fouronsix	2.03	2.08	2.56%
haffner	1.77	1.80	1.93%
mfv	3.06	3.25	6.20%
unfo	1.86	1.90	2.16%
violin	2.01	2.05	2.37%
waltz	1.81	1.84	1.77%
Overall	2.03	2.09	2.61%

Item	**MPEG-4 SLS**	**Improved MPEG-4 SLS**	**Improve-ment**
avemaria	1.95	1.96	0.54%
blackandtan	2.06	2.10	2.03%
broadway	1.71	1.72	0.68%
cherokee	2.09	2.13	2.13%
clarinet	2.20	2.25	2.27%
cymbal	2.13	2.14	0.61%
dcymbals	1.63	1.64	0.61%
etude	1.89	1.90	0.52%
flute	2.23	2.28	2.22%
fouronsix	2.26	2.32	2.67%
haffner	1.95	1.99	1.98%
mfv	1.97	1.98	0.58%
unfo	2.14	2.19	2.55%
violin	2.08	2.13	2.10%
waltz	2.11	2.15	2.26%
Overall	2.01	2.04	1.52%

Table 2. Comparison of compression ratio performance (Top: 48kHz/16bit; Bottom: 96kHz/24bit)

REFERENCE

[1] ISO/IEC JTC1/SC29/WG11, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part3: Audio," IS 11172-3, 1992.
[2] Information Technology – Coding of Audiovisual Objects, Part 3. Audio, Subpart 4 Time/Frequency Coding, ISO/JTC1/SC29/WG11, 1998.
[3] ISO/IEC JTC1/SC29/WG11 N5040, "Call for Proposals on MPEG-4 Lossless Audio Coding," Awaji Island, Japan, 2002.
[4] R. Yu, X. Lin, S. Rahardja and C. C. Ko, "A scalable lossy to lossless audio coder for MPEG-4 audio scalable lossless coding," Proceeding of ICASSP 2004.
[5] ISO/IEC JTC1/SC29/WG11 N5720, "Workplan for Audio Scalable Lossless Coding (SLS)," Trondheim, Norway, 2003.
[6] R. Geiger, T. Sporer, J. Koller and K. Brandenburg, "Audio Coding based on Integer Transforms," 111th AES Convention, Sep. 2001.
[7] J. Princen, A. Johnson, A. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation", *Proc. of the ICASSP 1987*, pp. 2161 – 2164, 1987.
[8] R. Yu, C.C. Ko, S. Rahardja and X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions," Proc. of the ICASSP 2003, pp. 277 – 280, 2003.
[9] 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
[10] R. Yu, C.C. Ko, S. Rahardja and X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions," Proceeding of ICASSP 2003.
[11] R. Yu, X. Lin, S. Rahardja, and H. Haibin, "Proposed Core Experiment for improving coding efficiency in MPEG-4 audio scalable coding (SLS)", ISO/IEC JTC1/SC29/WG11, M10136, Oct. 2003, Brisbane, Australia.
[12] R. Yu, X. Lin, S. Rahardja, and H. Haibin, "Proposed Core Experiment for improving coding efficiency in MPEG-4 audio scalable coding (SLS)", ISO/IEC JTC1/SC29/WG11, M10683, March. 2004, Munich, Germany.