MPEG-4 SCALABLE LOSSLESS AUDIO TRANSPARENT BITRATE AND ITS APPLICATION

Te Li, Susanto Rahardja

Institute for Infocomm Research, A*STAR, Singapore

ABSTRACT

In this paper, the relevance between the bit-plane levels and the perceptual quality of the MPEG-4 scalable lossless audio is explored. It is observed that only the top 3 bit-planes in SLS with AAC core of 64kbps are closely related to the transparent quality. The result is used in the application of the data hiding process. The proposed hiding method has low complexity as no side information is required at the data extraction process. A data hiding capacity of 98kbps is achieved for SLS lossless bitstream.

***Index Terms*—** Audio coding

1. INTRODUCTION

Published in June of 2006, MPEG-4 *scalable lossless* (SLS) [6] audio coding is a unified solution for demands in high compression perceptual audio and high quality lossless audio. It provides a fine-grain scalable extension to the MPEG-4 *advanced audio coding* (AAC) perceptual audio coder up to fully lossless reconstruction. Like most of the perceptual audio coders, SLS is able to provide the transparent-quality audio at a lossy bitrate (transparent bitrate). However, this transparent bitrate is not studied so far.

As an application, the bits beyond the transparent bitrate up to lossless can be exploited to store other useful data such as lyrics, music notes, CD cover art, surround audio [1] and video information. Many approaches on embedding data in audio signal are proposed [2, 3, 4]. Among them, the latest proposed data hiding scheme [4] using integer modified discrete cosine transform (IntMDCT) can be directly applied to SLS bitstream. The test results show that a large amount of data (around 41kbps at frame length of 1024) can be embedded in the IntMDCT stream without degrading the audio quality. However, there are two limitations of the approach. Firstly, a perceptual model has to be applied in the data extraction process, which increases the complexity of the codec. Secondly, there is an assumption that the perceptual significance in the decoded audio should be exactly the same as that for the original audio. This may not be always true if the amount of hidden data is large.

It is thus the purpose of this paper to: 1. study the relevance between the transparent quality and the bit-plane levels of the SLS bitstream. 2. apply this information as the guidance of the data hiding. Experimental results show that for SLS with AAC core of 64kbps (the most commonly used configuration), only the top 3 bit-planes are closely relevant to the transparent quality. This information is used as the only criteria of the data hiding application. No perceptual information is required for the data extraction process, and a minimum of 98kbps data can be embedded in the SLS bitstream with transparent quality. The rest of this paper is organized as follows. Section 2 gives a review on the basic structure of SLS. The study on transparent bitrate and the proposed data hiding algorithm are elaborated in Section 3. The experiment results are presented in Section 4 and in Section 5 conclusions are given accordingly.

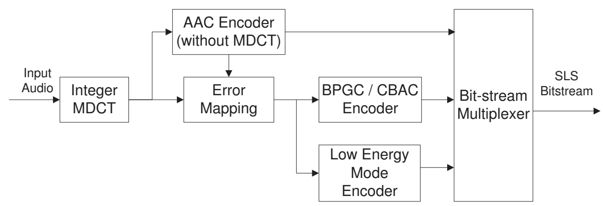


Fig. 1. Block diagram of MPEG-4 SLS encoder

2. SLS STRUCTURE

As depicted in Fig. 1, SLS consists of two separate layers - the core layer and the *lossless enhancement* (LLE) layer. In the SLS encoder, the input audio in integer PCM format is losslessly transformed into the frequency domain by using IntMDCT. The resulting coefficients are then passed to the AAC encoder to generate the core layer AAC bitstream. In order to efficiently utilize the information of the spectral data in the core layer bitstream, error-mapping procedure is employed to generate the residual spectrum coded in the LLE layer. The residual spectrum is then coded using *bit-plane Golomb coding* (BPGC) [7] combined with *context-based arithmetic coding* (CBAC) and *low energy mode coding* (LEMC) [8] to generate the scalable LLE layer bitstream.

BPGC is adopted in SLS as the major arithmetic coding scheme. It uses a probability assignment rule that is derived from the statistical properties (Laplace distributed) of the residual spectrum in SLS. The bit-plane symbol at bit-

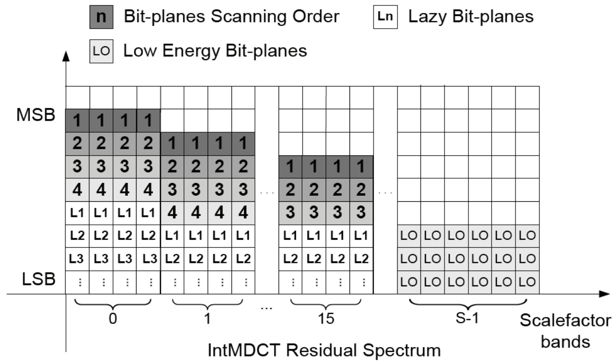


Fig. 2. The bit-plane coding sequence in MPEG-4 SLS

plane bp is coded with probability assignment given by

$$Q^{L[s]}[bp] = \begin{cases} \frac{1}{1+2^{2^{L[s]-bp}}}, & bp \leq L[s] \\ \frac{1}{2}, & bp > L[s] \end{cases} \tag{1}$$

where $s(0 \leq s < S)$ is the scalefactor band (sfb) and S denotes the total number of the sfb. $bp = 1$ denotes the bit-plane with the *most significant bit* (MSB). Since coding of binary symbol with probability assignment $1/2$ can be implemented by directly outputting input symbols to compressed bitstream, BPGC enters a *lazy* mode for bit-planes below $L[s]$. Therefore, $L[s]$ and the bit-planes below are referred as the *lazy* planes. For each sfb, $L[s]$ can be selected using the decision rule given in [7].

BPGC only delivers excellent compression performance when the sources are near-Laplacian distributed. However, for some music items, there exist some 'silence' regions where the spectral data are in fact dominated by the rounding errors of IntMDCT. In order to improve the coding efficiency, LEMC is adopted for coding signals from low energy regions. It is also possible to improve the coding efficiency of BPGC by further incorporating more sophisticated probability assignment rules that take into account the dependencies of the distribution of IntMDCT spectral data to several contexts, which can be effectively captured by using CBAC.

The overall bit-plane coding sequence is illustrated in Fig. 2 (BPGC only). The bit-plane coding in an SLS codec is performed in a sequential order that the plane of the MSB for spectral data from the lowest sfb to the highest sfb is coded first. It is followed by the subsequent bit-planes. Once the normal bit-planes are completed using either BPGC or CBAC, they are followed by the direct coding of the *lazy* bit-planes (without compression). The low energy bit-planes will be coded at last using LEMC until it reaches the plane of the *least significant bit* (LSB) for all sfbs.

Finally, the LLE bitstream is multiplexed with the core AAC bitstream to produce the final SLS bitstream.

3. TRANSPARENT BITRATE AND DATA HIDING

In the first part of this section, the relevance between the transparent bitrate and the bit-plane levels is analyzed. This information is then applied in the proposed hiding process, which will be elaborated in the second subsection.

3.1. Transparent Bitrate

As mentioned in Section 2, the bit-plane coding starts from $bp = 1$ for each sfb, and followed by $bp = 2$, $bp = 3$, and so on. Assume that after the encoding of sfb $s_t (0 \leq s_t < S)$ in $bp = T$, the distortion of all the sfbs will be just below the mask, i.e.,

$$d[s][T] < M[s], \ \forall\, 0 \leq s < S, \tag{2}$$

where $d[s][T]$ is the distortion in sfb s and $M[s]$ is the mask (allowed distortion) for s. The corresponding average bitrate needed is defined as transparent bitrate. Theoretically, SLS could provide transparent quality audio at this bitrate.

Firstly, the distortion, $d[s][T]$, is computed as follows. Given that $\bar{e}[k][T]$ is the reconstructed amount for residual frequency element k in sfb s after encoding of $bp = T$, it can be computed by

$$\bar{e}[k][T] = (2\hat{\varepsilon}[k] - 1) \cdot \sum_{bp=1}^{T} \left(b[k][bp] \cdot 2^{(m[s]-bp)} \right) \tag{3}$$

where $\hat{\varepsilon}[k]$ is the reconstructed sign symbol (0 or 1), $b[k][bp]$ is the bit symbol (0 or 1) and $m[s]$ denotes the total number of bit-planes in s.

If $T < m[s]$ and $m[s] > L[s]$, the reconstruction is further enhanced by an estimation process: although the bit-planes below the current $bp = T$ are not coded yet, they can be estimated based on the Laplacian distribution feature of the frequency elements in SLS coding. This reconstruction enhancement is performed in the SLS decoder. Specifically, the estimated extra amplitude for the following bit-planes can be computed using probability assignment in Eqn. (1) by

$$\tilde{e}[k][T] = (2\hat{\varepsilon}[k] - 1) \sum_{bp=T+1}^{m[s]} \left(Q_{bp}^{L[s]} \cdot 2^{(m[s]-bp)} \right), \tag{4}$$

The final reconstructed spectrum coefficient $\hat{e}[k][T]$ is obtained as

$$\hat{e}[k][T] = \begin{cases} \bar{e}[k][T] + \tilde{e}[k][T], & T < m[s], m[s] > L[s] \\ \bar{e}[k][T], & \text{otherwise} \end{cases} \tag{5}$$

The total distortion energy for the sfb s, $d[s][T]$ (dB) is then computed as

$$d[s][T] = 10 \log_{10} \left[\sum_{k=O[s]}^{O[s+1]-1} \left(e[k] - \hat{e}[k][T]\right)^2 \right]. \tag{6}$$

Table 1. MPEG-4 audio test excerpts (48kHz/16bits stereo)

No.	Name	No.	Name	No.	Name
1	avemaria	6	cymbal	11	haffner
2	blackandtan	7	dcymbals	12	mfv
3	broadway	8	etude	13	unfo
4	cherokee	9	flute	14	violin
5	clarinet	10	fouronsix	15	waltz

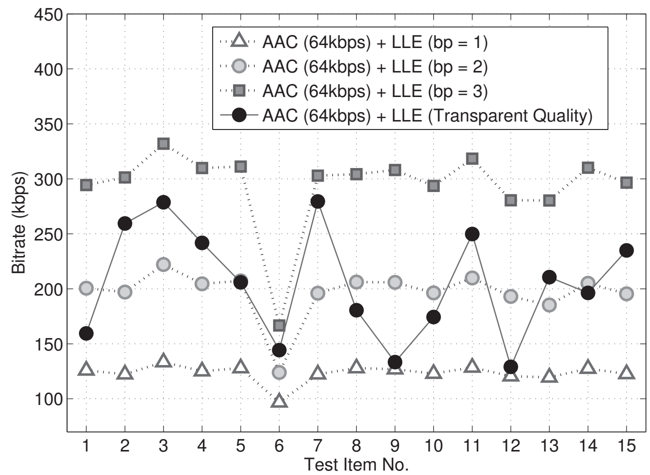


Fig. 3. The SLS transparent bitrates (total for stereo channels) for the 15 test items

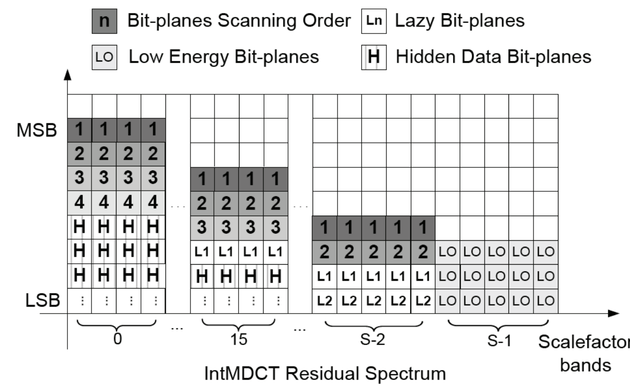


Fig. 4. The data hiding algorithm for SLS

where $O[s]$ denotes the starting frequency element number of sfb s and $e[k]$ is the original residual element.

The mask, $M[s]$ can be directly extracted from the psychoacoustic model in the AAC core encoding process. Therefore, based on the criteria defined in Eqn. (2), the transparent bitrate can be obtained.

A test on the transparent bitrate of SLS reference codec is performed using the MPEG-4 standard test sequences as listed in Table 1, with frame length of 1024, AAC core bitrate of 64kbps (total for stereo) and BPGC+LEMC only (without CBAC). Besides the transparent bitrate, the bitrate required for SLS at different bit-plane levels are also plotted for reference. Specifically, $bp = 1$ indicates the bitrate required for the first bit-plane of all the frames to be coded, and so on. It should be noted that all the bitrates mentioned in this paper are the total amount for stereo channels. The result is shown in Fig. 3.

It can be observed that for all the test sequences, the transparent quality can be obtained before the 3rd bit-plane of all the frames are coded. It is thus believed that the bit-planes below the 3rd bit-planes are not closely related to the perceptual quality. As such, generally the 4th and below bit-planes can be used for data hiding. In addition, BPGC will go to lazy mode coding (without compression) from the 5th bit-plane. It means that if the audio data is replaced by hidden data from the 5th bit-plane, the size of the bitstream will not change compared to the original bitstream. Instead, if the audio data in BPGC mode is replaced by hidden data, the bitstream size is likely to be changed due to the different probability characteristic of the hidden data compared with the audio data. By combining all the above conditions, it is concluded that all the bit-planes from the 5th and below, i.e., the lazy bit-planes, can be used for data hiding.

3.2. Data Hiding Algorithm

By applying the information obtained from the previous subsection, the data hiding for SLS can be performed as depicted in Fig. 4. For each sfb s, if $m[s] > 4$, the hidden data can be embedded from the 5th and below bit-planes. Otherwise, there will be no data embedded in the sfb. As $m[s]$ is known to the SLS decoder, no extra side information is required to extract the hidden data.

4. RESULT

According to the data hiding algorithm proposed in Section 3, the test on the amount of hidden data and the quality of the data embedded audio is performed for the same set of test sequences as listed in Table 1. The quality of the audio is evaluated using ITU-R BS.1387 p*erceptual evaluation of audio quality* (PEAQ) test method, where the performance of the audio under test is compared with that of the original audo inputs. The grading scale of 'objective difference grade' (ODG) ranges from -4 ('very annoying') to 0 ('imperceptible difference').

The test results are shown in Table 2. The comparison between the lossless bitrates and the data hiding bitrates are further plotted in Fig. 5. It is observed that the minimum amount of hidden data is around 98kbps (for item 12) and the maximum is around 532kbps (for item 7). By observing the spectrum plot (Fig. 6, where the brightness indicates the intensity of energy) of these two items, the difference in data

Table 2. Test results of the lossless bitrate (B_L), maximum data hiding bitrate (B_H), objective difference grade (ODG) and noise to mask ratio (NMR) of the data-embedded sequences

No.	B_L (kbps)	B_H (kbps)	ODG	NMR
1	608.80	199.40	0.00	-21.21
2	877.60	457.75	0.04	-20.93
3	788.27	348.79	-0.12	-18.98
4	842.40	416.25	0.06	-21.41
5	746.67	317.46	0.05	-20.76
6	493.07	125.92	-0.10	-16.60
7	956.00	532.76	-0.06	-19.24
8	654.67	234.91	0.04	-21.25
9	632.80	216.82	-0.07	-20.12
10	738.13	324.45	0.03	-20.72
11	868.53	430.71	0.06	-21.22
12	476.00	98.83	-0.10	-19.27
13	810.40	406.26	0.06	-21.27
14	763.73	335.58	0.01	-20.30
15	838.67	421.68	0.07	-21.49

hiding capacity is due to the reason that item 7 has more high frequency components and it results a larger number of bit-planes.

5. CONCLUSION

The relevance between the bit-plane levels and the transparent quality for MPEG-4 SLS is studied in this paper. The information is then applied in the data hiding process. It is proposed that all the lazy bit-planes can be used for data embedding without affecting the audio quality. The proposed hiding process has low complexity and a minimum amount of 98kbps data hiding capacity. However, the collected data

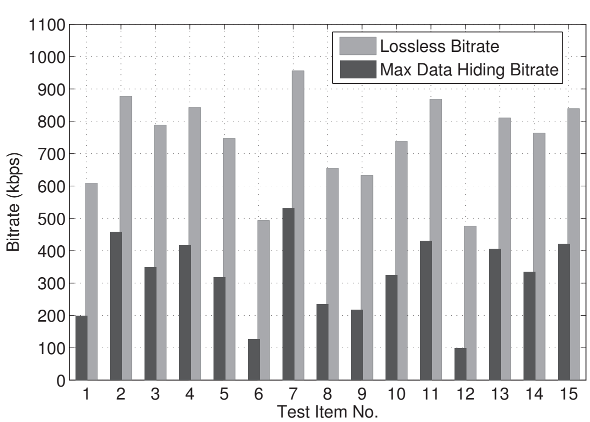
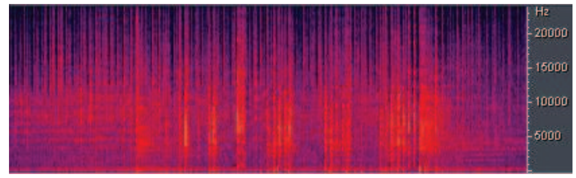
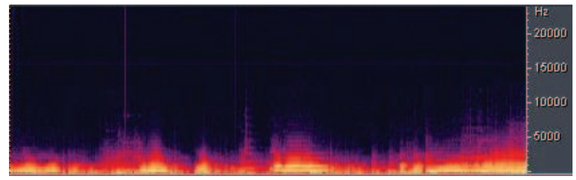


Fig. 5. The SLS lossless bitrates & maximum data hiding bitrates for the 15 test items



(a)



(b)

Fig. 6. Spectrum plot of (a) the 7th test item (b) the 12th test item

is only applicable for SLS with AAC core of 64kbps. It will be our future work to further explore the capacity of SLS bitstream with other configurations.

6. REFERENCES

[1] ISO/IEC 23003-1:2007, "Information technology – MPEG audio technologies – Part 1: MPEG Surround," 2007.

[2] W. ten Kate, L. van de Kerkhof and F. Zijderveld, "A new surround-stereo-surround coding technique," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 376–383, May 1992.

[3] J. Chou, K. Ramchandran and A. Ortega, "High capacity audio data hiding for noisy channels," in *Proc. IEEE Information Technology Coding and Computing*, Apr. 2001.

[4] R. Geiger, Y. Yokotani and G. Schuller, "Audio data hiding with high data rates based on IntMDCT," in *Proc. Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, pp. 205–208, Apr. 2006.

[5] ISO/IEC 14496-3, "Information Technology - Coding of audiovisual objects, Part 3. Audio," 1998.

[6] ISO/IEC 14496-3:2005/Amd 3, "Scalable lossless coding (SLS)," 2006.

[7] R. Yu, C.C. Ko, and S. Rahardja and X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions," in *Proc. ICASSP*, pp. 277-280, Apr. 2003.

[8] R. Yu, X. Lin, S. Rahardja, C.C. Ko, and H. Huang, "Improving coding effciency for mpeg-4 audio scalable lossless coding," in *Proc. ICASSP*, vol. 3, pp. 169–172, 2005.

[9] ITU-R BS.1387, "Method for objective measurements of perceived audio quality".