Quantization Watermarking Schemes for MPEG-4 General Audio Coding

Shu-chang Liu and Jin-tao Li

Digital Technology Lab, Institute of Computing Technology, CAS,
No.6 Kexueyuan South Road, P.O.Box 2704, 100080, Beijing, PRC
{scliu,jtli}@ict.ac.cn

Abstract. Quantization Watermarking or Quantization Index Modulation is usually used for uniform scalar quantizarion. In this article, the Quantizaion Watermarking extendedly refers to all the watermarking schemes that achieve watermark embedding during quantization process, ignoring the concrete method: scalar or vector quantization. After a brief review of audio watermarking, this paper mainly introduces the watermarking techniques, whose aim is to embed watermark in the quantized MDCT coefficients of AAC or the index value of TwinVQ. Some experimental results show that it works well, for the distortion coming from quantization watermarking is limited within the HAS thresholds, and the embedded watermark is difficult to be removed without significantly audio quality degrading and the key only known by the owner.

1 Introduction

MPEG-4 audio should become the most favorite audio stream over Internet, because of its freedom providing for the network end users. Like as other types of multimedia, copyright protection and multimedia authentication problems must be taken into account during designing its aimed Internet applications. Simultaneously, the flexibility and integration of MPEG-4 audio add the difficulties in realizing copyright protection.

Firstly, MPEG-4 audio integrates many different types of audio coding [1]. As far as we know there is no current watermarking methods can be fit for all kinds of coding tools, to define the most suitable watermarking schemes for each of them according to its characteristics is the basic assignment research on watermarking in MPEG-4 audio. The watermarking scheme may include embedding domain: time or frequency, applied techniques: spread-spectrum or quantization, robustness: robust, semi-fragile or fragile, detection mode: blind or non-blind (public or private), intention: ownership proof, owner identification, device control, transaction tracking, or copy control [2]. This is named *multi-scheme* coexistence problems. Secondly, the same MPEG-4 audio coding stream may be divided into more than one audio object, such as background music and foreground speech. This leads to the multi-ownership identification scenario, since different object could belong to distinct producers. In order to protect all rights and interests of the actors, say, composer, artist, content provider and sub

Y.-C. Chen, L.-W. Chang, and C.-T. Hsu (Eds.): PCM 2002, LNCS 2532, pp. 442-450, 2002.

scriber, one possible solution is to embed different watermarks on the different position or domain during specific coding stage. Here, we called this *multi-watermark* coexistence issues. The following question of *multi-scheme* and *multi-watermark* is mutual conflict between distinct watermarking methods. For instance quantization watermarking is useful information for some applications like as copy control, but it is indeed an incidental distortion in the viewpoint of time- and/or transform-domain watermarking used for ownership proof. Based on Intellectual Property Management and Protection, a framework should be defined to control the whole multi-watermark embedding and detection process, to deal with the mutual conflicts, to analyze and stretch the capacity of the multi-watermarking.

From the ideas listed afore, we began our research on copyright protection mechanism of MPEG-4 audio last year. In this paper watermarking schemes applying to general audio (GA) coding tools are briefly reviewed in section 2. Section 3 and 4 mainly focus on the improved quantization watermarking techniques that is implemented during quantization and inverse quantization process of MDCT coefficients. Some experimental results and evaluations are presented in section 5. And the last section gives the summaries and some future works.

2 Watermarking Schemes Applying to GA Copyright Protection

As I.J. Cox pointed out in [2] that electronic watermarking began with embedding an identification code into music for ownership proof in 1954, and the last 10 years digital watermarking also started from LSB (least significant bits) audio watermarking proposed by L.F. Turner. Although fewer watermarkers make a study of audio watermarking, lots of audio watermarking techniques have been put forward.

The earlier techniques work by placing the hiding information in some human perceptually insignificant regions [3]. For example, LSB replaces the least significant bits of randomly selected audio samples with the bits of watermark, phase coding substitutes the phase of an initial audio segments with a reference phase representing the data [4], in [5] the Fourier transform magnitude coefficients over the frequency band from 2.4kHz to 6.4kHz are replaced by the watermark coded spectral components. Other audio watermark working on some non-sensitivity of HAS should be called statistics-based technique, because their watermark embedding and/or detection are based on statistical characteristic of the time-domain samples [6,7] or the Flourier/DCT domain coefficients [8,9].

The explicitly making-based audio watermarking may include: echo hiding [4], frequency masking hiding, temporal & frequency masking hiding [10,11], if HAS masking effects can be classified into temporal, frequency and echo masking. Echo hiding works by introducing multiple echoes, which differ in three parameters: initial amplitude, decay rate and offset to represent binary one and zero respectively. Unlike echo hiding, the latter two masking-based watermarking algorithms exploit temporal and/or frequency masking to add a perceptually shaped pseudorandom sequence (watermark coded signal) to PCM samples or frequency coefficients.

After I.J. Cox introduced the spread spectrum communication theory into watermarking [12], spread-spectrum (SS) techniques are widely applied in multimedia copyright protection. It certainly includes audio watermarking such as [13], which spread each bit w_i of an SS sequence in frequency (a subband of MCLT samples x_k) and in time (T_0 consecutive MCLT frames) simultaneously. The major problem of SS watermarking is synchronization requirement between the frequencies of the pseudorandom sequences embedded in the content and that is used for detection. The methods of resisting synchronization attack include: frame synchronization [14], synchronization code [13], redundancy synchronization [15], and content synchronization [3].

Brian Chen etc and Joachim J. Eggers etc proposed Quantization Watermarking parallelly, one concentrates more on Quantization Index Modulation [16,17], and the other more on Dithered Quantization [18,19]. In other words, the former utilizes multiply quantizers to quantize the host signal, each quantizer with its associated index. Embedding is realized by modulating the quantilizers' associated index to make the quantilized value be fallen in the corresponding set. The latter adds a dither signal to cover signal before quantization, consequently the watermarking information is embedded in the quantization noise. The hypotheses test or correlation calculations can accomplish watermark detection. A scheme similar to QIM called Parity Modulation [20] was described in M.Ramkumar's PhD Thesis. Mean Quantization-based Fragile Watermarking proposed by Gwo-Jong Yu, Chun-Shien Lu etc [21] belongs to the Dithered Quantization. In fact, quantization-watermarking techniques mentioned afore all make full use the quanization noise hole between distortion perceptual threshold and compression techniques to hiding information in the quantized values, only viewing it from different points. If the extended meaning of Quantization Watermarking refers to all watermarking techniques which embed information bits in quantized value (index) or quantization noise during quantization process, then some watermarking techniques which combine watermark embedding with audio stream encoding process and watermark detection with decoding process can be viewed as Quantization Watermarking. Lintian Qiao etc in [22] introduced a method hiding information in the modulated scalefactors. Jack Lacy etc gave more general description about this method in paper [14]. In another scheme, watermark embedding is performed during vector quantization [23]. It works by changing the number of candidates used for preselection in the search procedure or changing weighting factor used for distortion measure of the conjugate vector quantization. One disadvantage of these algorithms is that they are not standard Quantization Watermarking schemes, and another is private detection.

3 AAC Quantization Watermarking

3.1 AAC Quantization of MDCT Coefficients

AAC Quantization module is divided into three levels: frame loop, rate loop, and distortion loop [1]. The quantizied data *quant*[*i*]are calculated as follows:

$$quant[i] = \text{floor}\left(\left(\frac{\text{fabs}(spectrum[i])}{quantFac[sb]}\right)^{3/4} + \text{M}\right) \quad (1)$$

Where i is the index of MDCT coefficients, sb is the index of scalefactor bands, and M is defined to 0.4054, the quantizer step size $quantFac[sb]$ follows:

$$quantFac[sb] = \text{pow}(2, \quad scalefactor[sb]/4)$$

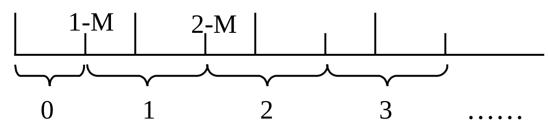

Fig. 1. AAC quantized intervals

3.2 Modulation (Embedding) Functions

There are two modulation functions to modulate the parity of $quant[i]$.

$$\text{M}(quant[i], w) = \begin{cases} \text{re}(quant[i]) & \begin{matrix}\text{pa}(quant[i])\text{=even and } w\text{=0} \\ \text{pa}(quant[i])\text{=odd and } w\text{=1}\end{matrix} \\ \text{disregard} & quant[i]\text{=0 or } quant[i]\text{=1} \\ \text{ch}(quant[i]) & \begin{matrix}\text{pa}(quant[i])\text{=even and } w\text{=1} \\ \text{pa}(quant[i])\text{=odd and } w\text{=0}\end{matrix} \end{cases} \quad (2)$$

$$\text{M}(p, quant[i], w) = \begin{cases} \text{re}(quant[i]) & \begin{matrix}\text{pa}(quant[i])\text{=pa}(p)\text{ and } w\text{=0} \\ \text{pa}(quant[i])\neq\text{pa}(p)\text{ and } w\text{=1}\end{matrix} \\ \text{disregard} & quant[i]\text{=0 or } quant[i]\text{=1} \\ \text{ch}(quant[i]) & \begin{matrix}\text{pa}(quant[i])\text{=pa}(p)\text{ and } w\text{=1} \\ \text{pa}(quant[i])\neq\text{pa}(p)\text{ and } w\text{=0}\end{matrix} \end{cases} \quad (3)$$

Where w is the current watermark bit, ch($quant[i]$) and re($quant[i]$) separately represents changing and retaining the parity of $quant[i]$, and pa($quant[i]$) returns the parity of $quant[i]$. Function (2) forces the non-zero and non-one $quant[i]$ to be an even integer when w=0, and forces it to be odd when w=1. Function (3) has one more parameter p, which means the last previous modulated $quant[]$, This function forces the parity of the non-zero and non-one $quant[i]$ to be the same one of p when w=0, while forces it to be the opposite one when w=1.

3.3 Effects of ch(*quant*[*i*])

The operations of ch(*quant*[*i*]) follows (4), and its meaning is illustrated in Figure 2

$$\mathrm{ch}(quant[i]) = \begin{cases} quant[i]+1 & quant[i]=2 \\ quant[i]-1 & qf - quant[i] < 0 \\ quant[i]+1 & qf - quant[i] \geq 0 \end{cases} \quad \textbf{(4)}$$

Where $qf = \left(\frac{\mathrm{fabs}(spectrum[i])}{quantFac[sb]}\right)^{3/4}$ is the actual quantized float data. When *qf* falls in interval (*q*-M, *q*), let *quant*[*i*] be *q*-1, and let it be *q*+1 while *qf* falls in [*q*, *q*+M], where *q* is the original quantized data (integer).

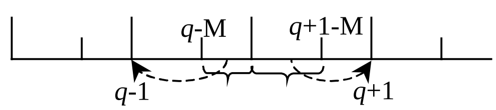

Fig. 2. Meanings of ch(*quant*[*i*])

Another explanation about effects of ch(*quant*[*i*]) comes from Dither Quantization [22]. Let *r* be the quantization errors of (fabs(*spectrum*[*i*]))^(3/4).

$$r = \left(\mathrm{fabs}(spectrum[i])\right)^{3/4} - q * \left(quantFac[sb]\right)^{3/4} = (qf - q) * (quantFac[sb])^{3/4} \quad \textbf{(5)}$$

Then *u* that denotes the dither signal adding to fabs(*spectum*[*i*])^(3/4) follows:

$$u = \begin{cases} -r & \mathrm{re}(quant[i]) \\ -r + (quantFac[sb])^{3/4} & quant[i]+1 \\ -r - (quantFac[sb])^{3/4} & quant[i]-1 \end{cases} \quad \textbf{(6)}$$

In fact *u* can be others as long as the quantized data conforms to (2) & (3).

3.4 Demodulation (Detection) Functions

According to modulation process, the demodulation functions (7) & (8) can be used for watermark detection, when the original audio is impossible to be got. The watermark detection can also be achieved by hypotheses test or correlation computation, which usually need original cover signal.

$$\mathrm{D}(quant[i]) = \begin{cases} \text{disregard} & quant[i]=0 \text{ or } quant[i]=1 \\ w=1 & \mathrm{pa}(quant[i]) = \text{odd} \\ w=0 & \mathrm{pa}(quant[i]) = \text{even} \end{cases} \quad \textbf{(7)}$$

$$\mathrm{D}(p, quant[i]) = \begin{cases} \text{disregard} & quant[i]=0 \text{ or } quant[i]=1 \\ w=1 & \mathrm{pa}(quant[i]) \neq \mathrm{pa}(p) \\ w=0 & \mathrm{pa}(quant[i]) = \mathrm{pa}(p) \end{cases} \quad \textbf{(8)}$$

4 Vector Quantization Watermarking

4.1 Vector Quantization of TwinVQ

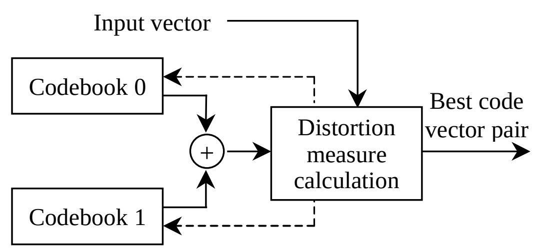

Fig. 3. Conjugate-structure vector quantization scheme

TwinVQ [24] is suitable for low-bit-rate general audio coding. It uses a conjugate-structure vector quantization scheme, which includes separate pre- and main-selection procedures. In pre-selection, a fixed number of candidate code vectors are chosen form codebook, and the best pair giving the minimum distortion measure is chosen during main-selection. Like as embedding watermark bit in *quant*[*i*], we can modulate the index pair to hide data.

4.2 Modulation (Embedding) and Demodulation (Detection) Functions

There are also two modulation functions executed during main-selection procedure in this vector quantization-watermarking scheme. Where *sum* = *index1* + *index2*, is the sum of code vector index of codebook1 (*index1*) and the one of codebook2 (*index2*).

$$\mathrm{M}(sum, w) = \begin{cases} \mathrm{pa}(sum) = \text{even} & w = 0 \\ \mathrm{pa}(sum) = \text{odd} & w = 1 \end{cases} \tag{9}$$

$$\mathrm{M}(p, sum, w) = \begin{cases} \mathrm{pa}(sum) = \text{even} & \begin{matrix} \mathrm{pa}(p) = \text{even and } w = 0 \\ \mathrm{pa}(p) = \text{odd and } w = 1 \end{matrix} \\ \\ \mathrm{pa}(sum) = \text{odd} & \begin{matrix} \mathrm{pa}(p) = \text{odd and } w = 0 \\ \mathrm{pa}(p) = \text{even and } w = 1 \end{matrix} \end{cases} \tag{10}$$

The demodulation functions list below:

$$\mathrm{D}(sum) = \begin{cases} w = 0 & \mathrm{p}a(sum) = even \\ w = 1 & \mathrm{p}a(sum) = odd \end{cases} \tag{11}$$

$$\mathrm{D}(p, sum) = \begin{cases} w = 1 & \mathrm{p}a(sum) \neq \mathrm{pa}(p) \\ w = 0 & \mathrm{p}a(sum) = \mathrm{pa}(p) \end{cases} \tag{12}$$

4.3 Effects of Modulation

Virtually, the Modulation functions change the search process in the main-selection procedure. Before calculating distortion measure [24], judge whether the variable *sum* satisfies the modulation function or not, and skip if it does not meet the condition. The real implementation of searching the best pairs is a dual-for loop. In order to get the best pair giving the minimum distortion according with the *sum* modulation, the dual-for loop need to execute two times separately taking *i_can* and *j_can* as the outer-for loop control variable. (meaning of *i_can* and *j_can* can be refered to [1]).

5 Experiments and Evaluations

The experimental environments was built based on VM of MPEG-4 audio: m4985 and w3309. The randomly selected tested audio clips list in Table 1.

Table 1. Basic information of tested audio clips

Clip Name	Type	Sample rate& Bitrate	File Size (kb)
AR001.AIF	Mono	22050, 176kbps	151
spacemusic.au	Mono	8000, 64kbps	47
San01.WAV	Mono	44100,353kbps	11,653
xuqu.wav	Mono	44100,706kbps	2,960

We modified the AAC quantization and TwinVQ vector quantization module to implement the watermarking scheme discussed above. The results about watermarking capacity of AAC_SYS mode and TVQ_SYS mode were tested in different encoding bitrate, as shown in Table 2 & Table 3. (The frame size is 1024)

Table 2. Capacity of AAC quantization watermarking

Bitrate(kb/s)	24	32	48	96	128
Capacity(b/frame)	11-15	21-35	28-58	60-110	250-370

Table 3. Capacity of TwinVQ quantization watermarking

Core Bitrate	8kps	16kps	24kps	32kps
AR001.AIF	25	55	87	117
spacemusic.au	79	165	250	335
San01.WAV	9	25	40	55
xuqu.wav	9	25	40	55

The advantages of the proposed watermarking scheme are blind-detection, more secure than pure Parity Modulation, and difficult to remove without significantly audio quality degrading. Inevitably they are fragile watermark, but this leads to its fitness for multimedia authentication (integrity assurance), under certain bit error rate of network transmissions, say 10^{-4}.

6 Conclusions

In this paper, the major problems of MPEG-4 audio copyright protection were firstly discussed, and then presented an AAC QW scheme combined with AAC quantization module and a vector QW algorithm implemented in the main-selection procedure of TwinVQ in detail. This is only the beginning of our project; there are a lot of works to do in future, such as more robust QW scheme, IPMP-integrated copyright protection framework, and capacity analysis tools for multi-watermark and multi-scheme.

References

1. ISO/IEC 14496-3. Information Technology–Coding of Audio-visual Objects–Part 3: Audio
2. I.J. Cox, M.L. Miller: The First 50 Years of Electronic Watermarking. Journal of applied Signal Processing. 2 (2002) 126-132
3. C.P. Wu, P.C. Su and C.J. Kuo: Robust and Efficient Digital Audio Watermarking Using Audio Content Analysis. Proceedings of SPIE Vol. 3971. (2000) 382-392
4. W. Bender, D. Gruhl, N. Morimoto and A. Lu: Techniques for Data Hiding. IBM Systems Journal. VOL 35, NOS 3&4 (1996) 313-336
5. J.F. Tilki and A.A. Beex: Encoding a Hidden Digital Signature onto an Audio Signal Using Psychoacoustic Masking. The 7th International Conference on Signal Processing Application & Technology, Boston MA. (1996) 476-480
6. P. Bassia and I. Pitas: Robust Audio Watermarking in the Time Domain. Proceedings of Eusipco-98. (1998) 25–28
7. P. Bassia, I. Pitas and N. Nikolaidis: Robust Audio Watermarking in the Time Domain. IEEE Transactions on multimedia. Vol.3 No.2 (2001) 232-241
8. M. Arnold: Audio Watermarking: Features, Applications and Algorithms. IEEE International Conference on Multimedia and Expo 2000, New York, NY, USA, (2000)
9. In-K. Yeo and H. J. Kim: Modified PatchWork Algorithm: a novel Audio Watermarking Scheme. Proc. of ITCC'01, Las Vegas, NV, (2001)
10. L. Boney, A.H. Tewfik and K.N. Hamdy: Digital Watermarks for Audio Signals. IEEE International Conference on Multimedia Computing and Systems. (1996) 473-480
11. M.D. Swanson, B. Zhu, A.H. Tewfik and L. Boney: Robust Audio Watermarking using Perceptual Masking. Signal Processing. Vol. 66 No.3, (1998) 337-355
12. I.J. Cox, J. Kilian, T. Leighton and T. Shamoon: Secure Spread Spectrum Watermarking for Multimedia. IEEE Transactions on Image Proceedings. Vol.6, No.12, (1997) 1673-1687

13. D. Kirovski and H.S. Malvar: Robust Covert Communication over a Public Audio Channel Using Spread Spectrum. IH Workshop, (2001)
14. J. Lacy, S.R. Quackenbush, A. Reib-man, D. shur, and J. Snyder: On Combining Watermarking with Perceptual Coding. In ICASSP, Vol.6 (1998) 3725-3728
15. R. Tachibana, S. Shimizu, T. Nakamura and S. Kobayashi: An audio watermarking method robust against time- and frequency-fluctuation. Security and watermarking of Multimedia Content, Proceedings of SPIE, Vol.4314. (2001) 104-115
16. B. Chen: Design and Analysis of Digital Watermarking, Information Embedding and Data Hiding System. PhD Thesis, MIT (2000)
17. B. Chen, G.W. Wornell: Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. IEEE Transaction on Information Theory. Vol.47, No.4 (2001) 1423-1443
18. J.J. Eggers and B. Grid: Quantizaion Watermarking. Proceedings of SPIE Vol. 3971 (2000)
19. J.J. Eggers and B. Girod: Quantization Effects on Digital Watermarks. Signal Processing. Vol.81, No.2 (2000) 239-263
20. M.Ramkumar: Data hiding in Multimedia: Theory and Application. PhD thesis, New Jersey Institute of Technology (NJIT) (2000)
21. G.J. Yu, C.S. Lu, and H.M. Liao: Mean Quantization-based Fragile Watermarking for Image Authentication. Optical Engineering, Vol.40, No.7 (2001) 1396-1408
22. L.T. Qiao and K. Nahrstedt: Non-Invertible Watermarking Methods for MPEG Audio. Research report of CS Department, UIUC, UIUCDCS-R-98-2069. (1998)
23. T. Moriya, Y. Takashima, T. Nakamura and N. Iwakami: Digital Watermarking Schemes Based on Vector Quantization. IEEE Workshop on Speech Coding for Telecomm.. (1997)
24. N. Iwakami, T. Moriya, A. Jin, T. Mori and K. Chikira: Fast Encoding Algorithms for MPEG-4 TwinVQ Audio Tool. ICASSP2001. (2001)