Adaptive Wavelet Quantization Index Modulation Technique for Audio Watermarking

Jong-Tzy Wang [1], Ming-Shan Lai[2], Kai-Wen Liang[2], and Pao-Chi Chang[2]
[1] Department of Electrical Engineering, Jin Wen Institute of Technology, Taiwan
[2] Department of Communication Engineering, National Central University, Taiwan
{jwang, mslai, kwliang, pcchang}@vaplab.ee.ncu.edu.tw

ABSTRACT

The quantization index modulation (QIM) that is often used in watermarking can provide a tradeoff between robustness and transparency. In this paper, we propose a robust audio watermarking technique which adopts the wavelet QIM method with adaptive step sizes for blind watermark extraction. Since wavelet transform offers both temporal and frequency resolutions, it is suitable for audio signal processing. The adaptive step size technique is applied to audio signals with different characteristics. This technique is designed based on the criterion that SNR must be maintained above 20 dB so that it is robust and transparent. No side information on the step sizes need to be transmitted. The experimental results show that the embedding capacity is around 4 bits/frame and the watermark is robust against MP3 compression at 64 Kbps, resampling, requantization, and Gaussian noise corruption. The NC values after attacks are all above 0.8 in the experiments so that the copyright can easily be distinguished.

Keywords: *robust watermarking, blind detection, wavelet packet, QIM*

1: Introduction

Watermark techniques that provide solutions to the copyright problem become more and more important because the distribution of digital media over Internet is getting popular. Conventional audio watermarking algorithms are performed in the time and the frequency domain. Since wavelet transform offers both temporal and frequency resolutions, it is suitable for analyzing audio signals that need different resolutions in different bands. Recently, several watermarking techniques performed in the wavelet domain were proposed [1] [2].

Chen *et al.* [3] introduced QIM technique for characterizing the inherent tradeoffs between the robustness and the rate-distortion of the embedding. The embedding scheme [1] proposed by T.-T. Lu is to use the AIM (Activity Index Modulation) technique that applies QIM to the image activity represented by the sum of the absolute pixel values for embedding. A potential problem of the scheme is that the image quality and the robustness cannot be maintained since it uses the same step size for different image signals with different characteristics.

According to the watermarking scheme [2] proposed by P. Bao and X. Ma, adaptive sizes are used for different image signals. However, the step size is required to be sent to the extraction end. This makes the overhead become very large.

We propose an audio watermarking system based on wavelet packet decomposition and psychoacoustic modeling. The audio watermarking system can deliver perceptual transparent audio quality, and it is robust against various signal processing or malicious attacks. We analyze signals in the wavelet domain and use adaptive step sizes for audio signals with different characteristics based on the criterion that the SNR must be maintained above 20dB so that it is robust and transparent. In addition, no side information on the step size needs to be sent to the extraction end. The relationship between the QIM step size and the SNR is analyzed. Both the embedding end and the extraction end can use the same formula to obtain the optimal step size.

Section 2 introduces the quantization index modulation technique. The proposed scheme is described in Section 3, Section 4 shows the experimental results, and finally the conclusion is provided in section 5.

2: Quantization Index Modulation

The determination of the step size is a tradeoff between robustness and quality of audio signals. In this section, we will discuss how the QIM step size affects the noise margin and SNR.

When the QIM is used to embed watermark, we first find the maximum value of audio signal and the difference between 0 and the maximum value is divided into intervals. Each interval will be assigned an index as 0 or 1. We define polarity of calculated signal value by the index of the interval in which it is located. In order to embed watermarks, we shift the value to the median of the interval or to a nearest median of the neighbor intervals by the relationship between polarity of signal and watermark bit. Here is an example showing in Fig. 1, if the bit of watermark and polarity are the same (right black point), they just need to be moved to the median of the same interval. If the bit of watermark and polarity are different (left black point), the value is shifted to the median of the nearest neighbor interval. The quantization error is $\pm\Delta$ at most as expressed in (1), where Δ is the step size.

$$Q(x) = x \pm \Delta . \quad (1)$$

The mean squared quantization noise power is derived as the following:

$$\langle q^2 \rangle = \int_{-\Delta}^{\Delta} \varepsilon^2 \frac{1}{2\Delta} d\varepsilon = \Delta^2 / 3 \ . \quad (2)$$

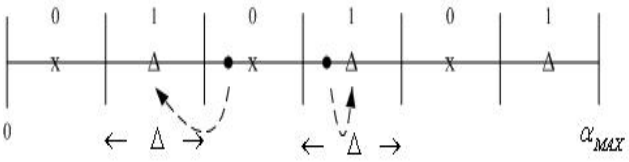

Fig. 1. Step Size vs. SNR

It shows that the mean squared quantization noise power is affected by the step size. On the other hand, the noise margin which determines the robustness of the watermark is also proportional to the step size. Therefore, the selection of the step size is basically a tradeoff between audio quality and robustness. Selecting a suitable step size is extremely important in a QIM based watermarking system. Our proposed adaptive system is capable of maintaining SNR of audio quality above 20dB while provides the good performance in robustness at the same time.

3: The Proposed Watermarking Strategy

This section describes the signal flow of the watermark embedding and the blind watermark extraction of our proposed adaptive QIM algorithm.

3.1: Embedding Algorithm

The embedding algorithm is described in Fig. 2. The original audio signal is first segmented into frames, with each frame denoted as $\mathbf{x} = \{x_n, n = 1, 2....\}$ and divided into 29 subbands via the wavelet packet decomposition. The bandwidth allocation of the subband decomposition structure is close to the critical band structure of human auditory system [4]. Watermark is embedded in the wavelet domain by the QIM method to produce watermarked audio $\mathbf{x}' = \{x'_n, n = 1, 2, \ldots\}$.

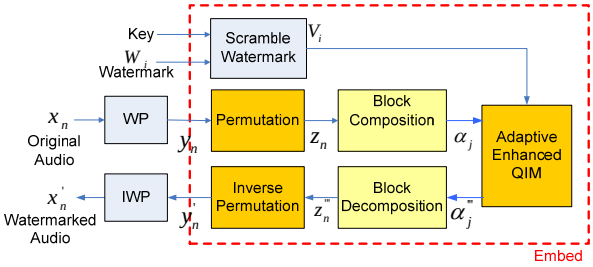

Fig. 2. Embedding diagram

3.1.1. Scramble watermark. Perform the scrambling operation on the watermark sequence W_i to obtain V_i, where T_i is the random binary sequence, and N_1 is the length of W_i in bits.

$$V_i = W_i \oplus T_i, \ i = 1 \sim N_1 \quad (3)$$

3.1.2. Selection of wavelet packet subbands. An audio signal is transformed into wavelet packet subbands:

$$\mathbf{y} = W_P(\mathbf{x}), \ \mathbf{x} = \{x_n, n = 1 \sim m_1\} . \quad (4)$$

Where $\mathbf{y} = \{y_n, n = 1 \sim m_1\}$, and m_1 is the number of samples in a frame.

According to the psychoacoustic model, middle-low subbands are used for watermark embedding to maintain the robustness and the transparency. Therefore, we choose wavelet coefficients from n_1 to n_2 out of m_1 to embed. In addition, from the experiments [5], we exploit the simplest wavelet basis, Haar wavelet, in our system.

3.1.3. Permutation and Block Composition. In order to avoid the damage from burst errors, the wavelet coefficient y_n are first processed by random permutations, $P_w(.)$, to yield $\mathbf{Z} = \{Z_n, n = n_1 \sim n_2\}$,

$$\mathbf{Z} = P_w(\mathbf{y}), \ \mathbf{y} = \{y_n, n = n_1 \sim n_2\} \quad (5).$$

The permutated wavelet coefficients in each subband are grouped into blocks with a suitable block size such as 8 coefficients and are calculated to get block mean α_j, where j is the block index, $j = 1, 2, .. \sim N$.

$$\alpha_j = \frac{1}{8} \sum |(z_n)|, \ \text{n}=n_1 \sim n_2, \ j = 1 \sim N, \ N = \left\lfloor \frac{n_2 - n_1}{8} \right\rfloor \quad (6)$$

3.1.4. Adaptive Enhanced QIM. The Adaptive Enhanced QIM that consists of two major steps, the determination of the number of intervals and the calculation of $Index\_number_j$, are described in detail as follows.

- Determination of the number of intervals (I):

First, we choose the maximum value of α_j through all blocks:

$$\alpha_{MAX} = MAX(\alpha_j), \ j = 1 \sim N . \quad (7)$$

The initial number of intervals I_i is calculated as

$$I_i = \frac{\alpha_{MAX}}{\Delta_k} \quad (8)$$

where Δ_k is the step size determined by the adaptive algorithm that will be described later.

Based on the different characteristic of audio signal in the temporal domain, we use the adaptive step size technique for different frames such that the watermark is robust while the good audio quality can be maintained. No side information on the step sizes need to be

transmitted. According to the requirement of IFPI [6], SNR of audio watermark of good quality must be above 20dB. Therefore, we start the step size from a small value and then increase it until SNR is equal to or larger than 20dB.

From the experimental result, we find that the step size Δ_k varies with a tendency that is in proportional to the mean $(\alpha_{MAX})_k$ of the block, such that we can get a formula of the step size as a linear equation (9),

$$\Delta_k = \Delta_{\mathrm{m}} + \frac{\Delta_M - \Delta_{\min}}{\beta_M - \beta_m}((\alpha_{MAX})_k - \beta_m) \tag{9}$$

where $\beta_M = MAX((\alpha_{MAX})_k)$, $\beta_m = MIN((\alpha_{MAX})_k)$, and $\Delta_{\mathrm{m}} = MIN(\Delta_k)$, $\Delta_M = MAX(\Delta_k)$, by the statistics Δ_{m}=5, β_M=1218 and β_m=50. Δ_M =102.33

It is important to maintain the same number of intervals at the extraction end to avoid watermarking decoding errors. In order to send no side information, the number of interval I is recalculated at the extraction end by (8), too. Dead zone evacuation is used to enhance the robustness of calculated I which is described as

$$\begin{aligned} \sigma_1 &= [0.5 - (I_i - \lfloor I_i \rfloor)] * \Delta_k \\ \alpha'_{MAX} &= \alpha_{MAX} + \sigma_1 \end{aligned} . \tag{10}$$

where σ_1 is the parameter used to change the original audio signals for increasing α_{MAX} to α'_{MAX} . And α'_{MAX} is then used to calculate I' so that $I' = (\lceil I_i \rceil + \lfloor I_i \rfloor)/2$.

Because the number of intervals must be an integer, we floor value to get I

$$I = \lfloor I' \rfloor . \tag{11}$$

● Calculation of $Index\_number_{\mathrm{j}}$:

Before embedding watermark, we need to calculate $Index\_number_{\mathrm{j}}$ for each block as in (12), where α_j is the average magnitude of the permuted wavelet coefficients.

$$Index\_number_{\mathrm{j}} = \left\lfloor \frac{\alpha_j}{I} \right\rfloor \bmod 2 \ , j = 1 \sim N \tag{12}$$

In our proposed algorithm, the basic concept of QIM is used and the range of means of blocks is divided into intervals according to I .We need to calculate $\mathrm{A_j}$ based on the step size Δ_k to determine the interval that α_j is located:

$$\mathrm{A_j} = \frac{\alpha_j}{\Delta_k} \ , j = 1 \sim N . \tag{13}$$

It is important to keep α_{MAX} to be maximal because α_{MAX} is used to determine the number of intervals I . In the process of embedding watermark, we need to change specific α_j to make sure that α_j is less than α_{MAX} even after embedding. Because α_j is modulated for embedding such that the value of α_j may exceed α_{MAX} . It may happen when α_j is located in the maximal interval. On the other hand, the block that α_{MAX} belongs to should be kept unchanged.

In the case that α_j is in the maximal interval, we need to decrease $\mathrm{A_j}$ by two to enhance correct extraction.

$$\mathrm{A}'_j = \mathrm{A_j} - 2 . \tag{14}$$

To get the result, we first modify the coefficients in each block.

$$\mathrm{z}'_{\mathrm{n}} = \begin{cases} \mathrm{z_n} - 2\Delta_k, & if \ \mathrm{z_n} \geq 0 \\ \mathrm{z_n} + 2\Delta_k, & if \ \mathrm{z_n} < 0 \end{cases} \tag{15}$$

Since α'_j is calculated by z'_{n}, we get

$$\alpha'_j = \alpha_{\mathrm{j}} - 2\Delta \tag{16}$$

Finally A'_j calculated from α'_j will match the desired condition in (14).

In cases that α_j is not in the maximal interval, the coefficients are kept with no change.

.

Embedding scheme:

.

If $Index\_number_{\mathrm{j}}$ and the permuted watermark sequence V_i are the same, we need to move σ_2 to the central of the interval that they belong to.

$$\begin{cases} \sigma_2 = [0.5 - (\mathrm{A}'_{\mathrm{j}} - \lfloor \mathrm{A}'_{\mathrm{j}} \rfloor)] \times \Delta_k \\ \alpha''_j = \alpha'_j + \sigma_2 \end{cases} . \tag{17}$$

If $Index\_number_{\mathrm{j}}$ and the permuted watermark sequence V_i are different, we need to move σ_2 to the central of the nearest neighbor interval.

$$\begin{cases} \sigma_2 = \begin{cases} [1.5 - (\mathrm{A}'_{\mathrm{j}} - \lfloor \mathrm{A}'_{\mathrm{j}} \rfloor)] \times \Delta_k, & if \ (\mathrm{A}'_{\mathrm{j}} - \lfloor \mathrm{A}'_{\mathrm{j}} \rfloor) \geq 0.5 \\ [(\lceil \mathrm{A}'_{\mathrm{j}} \rceil - \mathrm{A}'_{\mathrm{j}}) - 1.5] \times \Delta_k, & if \ (\mathrm{A}'_{\mathrm{j}} - \lfloor \mathrm{A}'_{\mathrm{j}} \rfloor) < 0.5 \end{cases} . \\ \alpha''_j = \alpha'_j + \sigma_2 \end{cases} \tag{18}$$

$$z_n^{''} = \begin{cases} z_n^{'} + \sigma_2, & if \ \ z_n^{'} \geq 0 \\ z_n^{'} - \sigma_2, & if \ \ z_n^{'} < 0 \end{cases}. \qquad (19)$$

3.1.5. Block Decomposition and Inverse Permutation. In the process of adaptive enhanced QIM, z_n is changed to $z_n^{'}$. $y_n^{'}$ that is the watermarked coefficient of wavelet packet, is then generated by inverse permutation.

$$\mathbf{Z}^{'} = P_w^{-1}(\mathbf{y}^{'}) \qquad (20)$$

3.1.6. Inverse Wavelet Packet Transform (IWP). Finally, IWP of the wavelet coefficients generates the watermarked audio signal($x_n^{'}$).

$$\mathbf{x}^{'} = W_p^{-1}(\mathbf{y}^{'}) \qquad (21)$$

After performing all the above steps, the watermarked audio signal is generated.

3.2: Blind Watermark extraction

In the extraction process, a watermarked audio signal ($\tilde{x}_n^{'}$) is performed by the same procedure as in the embedding process to finally generate $\tilde{I}ndex\_number_j$.

According to the following criterion, we can determine $\tilde{V}_i$:

(A) if $\tilde{I}ndex\_number_j = 0$,then $\tilde{V}_i = 0$.

(B) if $\tilde{I}ndex\_number_j = 1$,then $\tilde{V}_i = 1$.

Since the watermark sequences W_i are randomly permuted in the embedding end, the extracted watermark ($\tilde{W}_i$) must be inverse permuted as in (22).

$$\tilde{W}_i = P_w^{-1}(\tilde{V}_i), \ \ i = 1 \sim N_1 \qquad (22)$$

4: Experimental Results

We test our algorithm on 24 sequences of 16-bit signed mono audio signals sampled at 44.1 kHz in PCM format. NC (Normalized Correlation) and SNR are used as the performance criteria of the proposed algorithm. The NC, defined as in (23) is used to calculate the similarity between the extracted and the original watermarks. A watermark pattern (W) with the size *28 pixe lx28 pixel*, is used in simulations.

$$\text{Nc} = \frac{\sum_i \sum_j W(i,j)\tilde{W}(i,j)}{\sum_i \sum_j [W(i,j)]^2} \qquad (23)$$

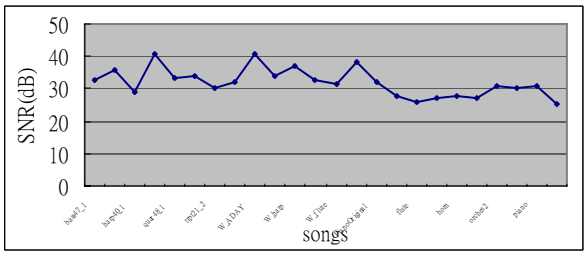

Fig. 3. SNR of original audio signals and watermarked audio signals

In these experiments, we use LAME MP3 encoder/decoder for re-encoding attacks, we reduce the original sampling rate from 44.1 kHz to 22.05 kHz and raise it back to 44.1 kHz by using interpolation for re-sampling attacks. Similarly, requantization changes the number of bits needed from 16 bits to 8 bits first, and then raises it back to 16 bits. White Gaussian noise with a constant level of 28 and 36 dB is added to the watermarked signals. The results of the performance after attacks are shown in Fig. **4** that shows the watermark is robust against MP3 compression at 64 Kbps, resampling, requantization, and Gaussian noise corruption. In all cases, NC values are above 0.8 that the copyright can clearly be distinguished.

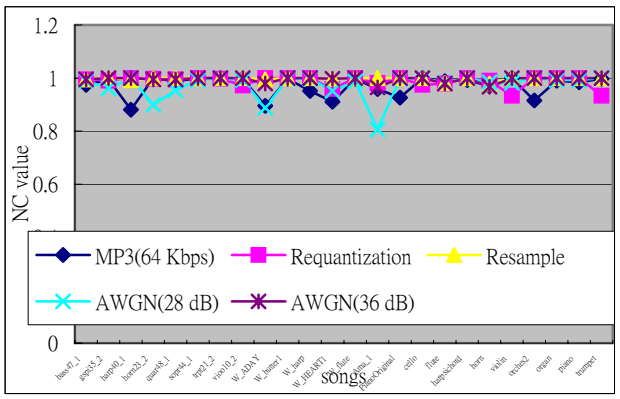

Fig. 4. Nc value after audio-processing.

A false positive rate is a detection of a watermark in a piece of media that does not actually contain that watermark. For convenience, only the host signal is extracted. If the NC of the host signal is very small, it proves that there is no false positive error. **Fig. 5** shows that the false positive rate is very low because the NC of the original (unwatermarked) signal is nearly 0.

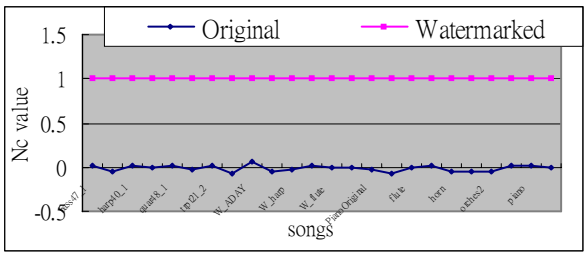

Fig. 5. NCs of original audio signals and watermarked audio signals.

Fig. 6 shows the step sizes by running the adaptive algorithm at the embedding side as well as the extraction side. Both curves are almost indistinguishable that shows it is not necessary to send the side information on the step size.

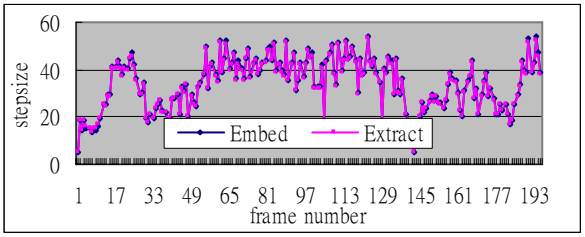

Fig. 6. The step size of trpt21_2 at the embedding and extraction end.

5: Conclusion

In this paper, we present a robust audio watermarking system based on wavelet quantization index modulation. Original audio signals are not needed for extraction. In addition, an adaptive step size technique is utilized for getting the balance of the robustness and the transparency. The capacity is around 4 bits/frame. The simulation results show excellent performance against various attacks. The focus of future work will be the enhancement of robustness against more attacks, such as random dropping, random inserting, and random cropping.

References:

[1].Lu, T. T.: Featured-based block-wise processing applied to image and video compression and watermarking systems. Ph.D. dissertation. Elect. Eng. Dept. Univ. of National Central. Taoyuan. Taiwan (2003)

[2].Bao, P. , Ma, X.: Image adaptive watermarking using wavelet domain singular value decomposition. IEEE Trans. on Circuits and Systems for Video Technology. Vol.15. No. 1. Jan. (2005) 96-102

[3].Chen, B., Wornell, G. W.: Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inf. Theory. Vol.47. No.4. May (2001) 1423-1443

[4].ISO/IEC 11172-3 : Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 M bits/s - Part 3: Audio (MPEG-1) . (1992)

[5].Wu, S., Huang, J. D., Shi, Y. Q.: Self-synchronized audio watermark in DWT domain," IEEE Int. Symposium on Circuits and Systems. Vancouver Canada. May23-26 (2004) 23- 26.

[6].Katzenbeisser, S., Petitcolas, F. A. P.: Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, Inc. Canton Street Norwood MA. (2000)