Journal of Computers Vol. 28, No. 4, 2017, pp. 63-73
doi:10.3966/199115592017082804007

Improving DWT-DCT-Based Blind Audio Watermarking Using Perceptually Energy-Compensated QIM

Hwai-Tsu Hu[1*], Szu-Hong Chen[1], and Ling-Yuan Hsu[2]

[1] Department of Electronic Engineering, National I-Lan University
Yi-Lan 26041, Taiwan, ROC
hthu@niu.edu.tw*; red055132@hotmail.com

[2] Department of Information Management, St. Mary's Junior College of Medicine, Nursing and Management,
Yi-Lan 26644, Taiwan, ROC
hsulingyuan@gmail.com

Received 6 July 2015; Revised 19 February 2016; Accepted 27 December 2016

Abstract. A scheme for energy compensation is proposed to remedy the deficiency of a DWT-DCT-based scheme while applying quantization index modulation (QIM) to perform blind audio watermarking. Our experimental results show that both compensated and uncompensated DWT-DCT schemes can achieve satisfactory robustness and imperceptibility at a payload capacity as high as 516.80 bps. However, because of the exploitation of the auditory masking effect, the perceptual quality attained by the compensated DWT-DCT scheme is even higher than that by the uncompensated one. With the employment of energy compensation, not only a 100% recovery of the watermark is guaranteed for non-attack situations but the survival rate is substantially improved in the case of extremely lowpass filtering. Furthermore, in a comparison with four other recently developed methods, the proposed DWT-DCT scheme observably exhibits a superior performance in imperceptivity and payload capacity while its robustness is comparable with others.

Keywords: blind audio watermarking, DWT-DCT based scheme, perceptually energy-compensated QIM

1 Introduction

The advancement of information and Internet technology has made the reproduction and dissemination of digital data much easier than ever before. People around the world keep creating and spreading a mass amount of multimedia data every day. Unfortunately, the illegal use of multimedia data is also rampant in the digital age. The protection against intellectual property violation appears an important issue nowadays. Digital watermarking technology has been considered a promising means to resolve this issue. It is a technique of hiding proprietary information into multimedia data and later extracting such information for copyright protection, content authentication, ownership verification, etc.

Watermarking schemes are often evaluated from four aspects, namely, security, robustness, imperceptibility and payload capacity. The embedded watermark shall remain secure during data transmission and endure various intentional attacks or unintentional modifications. The quality of the watermarked signal is required to be as close to the original as possible. Furthermore, the watermark capacity needs to be sufficiently enough to contain all necessary information.

For audio data, watermarking can be implemented in either the time domain [1-3] or transform domains such as spectrum [4-6], discrete cosine transform (DCT) [7-10], discrete wavelet transform (DWT) [7, 11-12], cepstrum [13-15], singular value decomposition (SVD) [9, 16-18]. Transform-domain

* Corresponding Author

techniques are generally more efficient because they can take advantage of signal characteristics and auditory properties [19].

The DWT-DCT scheme developed by Wang et al. [7, 20] was shown to be very efficient in many aspects. With the choice of appropriate embedding strength, the DWT-DCT achieves high capacity embedding with excellent perceptual quality and yet the resultant watermark is robust against common digital signal processing attacks. However, despite its superiority in all measures, the original formulation of the DWT-DCT exhibits a fundamental deficiency. That is, the host signals are modified without considering the consequential influence to watermark extraction. As a result, there is no guarantee that the watermark can be fully recovered even when the attacks are absent. In the following a scheme based on perceptual watermarking is introduced to amend this deficiency.

The rest of the paper is organized as follows. Subsequent to the introduction, Section 2 discusses in detail about the techniques involved in the proposed watermarking scheme. This section has been divided into several subsections including the auditory masking, perceptual-based QIM, restraint of energy compensation in watermark embedding, frame synchronization and watermark extraction. Section 3 presents the performance evaluation in comparison with other recently developed schemes. Finally, Section 4 draws up concluding remarks.

2 DWT-DCT Based Watermarking

The proposed watermarking scheme is performed in the DWT-DCT domain, where the targeted objects are the DCT coefficients, termed c_k's, derived from the approximation coefficients of the 5^{th} level DWT of the audio signal. In our design, the DWT-DCT coefficients have been partitioned into frames of length 128 to facilitate the subsequent formulation. As the energy of the audio signal is mostly centered at low frequencies, our focus is particularly placed on the first 64 coefficients that roughly correspond to a spectral span from 0 to $f_s/128$ with f_s denoting the sampling rate.

2.1 Exploitation of Auditory Masking

According to the auditory masking theory [21], the signal alteration due to watermarking will be inaudible providing the altered energy falls below the masking threshold in a specific critical band. Here we take the middle of a frequency band as the representative frequency termed f_{rep} and convert it to a Bark scale via

$$z_{rep} = 13\tan^{-1}\left(0.00076 f_{rep}\right) + 3.5\tan^{-1}\left((f_{rep}/7500)^2\right). \quad \textbf{(1)}$$

The auditory masking threshold for a specific band can be assessed using

$$a(z) = \lambda a_{tmn}(z) + (1-\lambda) a_{nmn}(z)\,[\mathrm{dB}], \quad \textbf{(2)}$$

where λ denotes a tonality factor, $a_{tmn}(z)$ is the tone-masking noise index estimated as $a_{tmn}(z) = -0.275z - 15.025$, and $a_{nmn}(z)$ is the noise-masking noise index usually fixed as $a_{nmn}(z) = -9$. Since $a(z) \geq a_{tmn}(z)$ no matter what λ is, we can regard $a_{tmn}(z_{rep})$ as the maximum tolerable level of energy variation. In theory, the embedded watermark will be imperceptible if the energy variation does not exceed E_{mask}, which is defined as

$$E_{mask} = 10^{\frac{a_{tmn}(z_{rep})}{10}} \times E_c; \quad E_c = \sum_{i=0}^{63} c_i^2. \quad \textbf{(3)}$$

2.2 Perceptual-based QIM

To embed a binary bit w_b into a selected c_k, we resort to the QIM rule [22] such that

$$\breve{c}_k = \begin{cases} \left\lfloor \frac{c_k}{\Delta} + 0.5 \right\rfloor \Delta, & \text{if } w_b = 0; \\ \left\lfloor \frac{c_k}{\Delta} \right\rfloor \Delta + \frac{\Delta}{2}, & \text{if } w_b = 1, \end{cases} \tag{4}$$

where $\breve{c}_k$ denotes the quantized version of c_k. $\lfloor \bullet \rfloor$ stands for the floor function. Δ is the quantization step size. Employing a larger Δ can enhance robustness but degrade audio quality. On the other hand, using a smaller Δ avails imperceptibility but impairs robustness. Our solution to this dilemma is to raise Δ to the maximum level that is tolerable by the human auditory system. Apparently, establishing a link between E_{mask} and Δ is of paramount importance.

In our formulation, the 64 coefficients are first categorized into two groups, namely G_1 and G_2, containing the indexes of $L_{G_1}(=48)$ and $L_{G_2}(=16)$ coefficients respectively.

$$G_1 = \left\{ n \middle| n = 0,1,\cdots, L_{G_1} + L_{G_2} - 1 \right\} - G_2 \tag{5}$$

$$G_2 = \left\{ 4n-1 \middle| n = 1,2,\cdots, L_{G_2} \right\} \tag{6}$$

We insert an exact amount of L_{G_1} bits into the coefficients in G_1. Given that f_s = 44.1 kHz, the resultant payload capacity is 516.80 bps. The coefficients in G_2 are reserved to maintain energy balance. Owing to the formulation shown in Equation (4), the difference between $\breve{c}_k$ and c_k in G_1 generally exhibits a uniform distribution over $[-\Delta/2, \Delta/2]$. Similarly, we restrict the magnitude change to be less than $\Delta/2$ for each coefficient in G_2. This condition is analogous to the effect caused by the QIM. In the worst scenario where all the modified coefficients deviate from their original values by $\Delta/2$ in G_2, the overall energy deviation becomes

$$\begin{aligned} E_{dev} &= L_{G_1} \times \underset{k \in G_1}{\mathrm{E}} \left[(\breve{c}_k - c_k)^2 \right] + L_{G_2} \times \left(\frac{\Delta}{2} \right)^2 \\ &= 48 \frac{\Delta^2}{12} + 16 \frac{\Delta^2}{4} \\ &= 8\Delta^2, \end{aligned} \tag{7}$$

where $\underset{k \in G_1}{\mathrm{E}} [\bullet]$ denotes the expectation for samples drawn from G_1. By letting E_{dev} equal E_{mask}, we have

$$8\Delta^2 = 10^{\frac{a_{tmn}(z_{rep})}{10}} \times E_c. \tag{8}$$

Or equivalently,

$$\Delta = \sqrt{10^{\frac{a_{tmn}(z_{rep})}{10}} \times E_c / 8}. \tag{9}$$

Theoretically, using the above derived Δ will make the embedded watermark imperceptible.

2.3 Restraint of Energy Compensation in Watermark Embedding

Because of the magnitude constraint of each c_k, the total energy in group G_2 can only vary between ρ_{dec} and ρ_{inc}:

$$\rho_{inc} = \sum_{k \in G_2} \left[\left(|c_k| + \frac{\Delta}{2} \right)^2 - c_k^2 \right]; \tag{10}$$

$$\rho_{dec} = \sum_{k \in G_2} \left[\left(\max\left\{ |c_k| - \frac{\Delta}{2}, 0 \right\} \right)^2 - c_k^2 \right]. \tag{11}$$

This implies that the energy variation due to the QIM in G_1 must satisfy the inequality

$$-\rho_{inc} \le \sum_{k \in G_1} \left(\breve{c}_k^2 - c_k^2 \right) \le -\rho_{dec}. \tag{12}$$

Note that the QIM shown in Eq. (4) aims at minimizing $\sum_{k \in G_1} \left(\breve{c}_k - c_k \right)^2$ instead of $\sum_{k \in G_1} \left(\breve{c}_k^2 - c_k^2 \right)$. In case Inequality (12) does not hold, some of the coefficients in G_2 will undergo excessive adjustments in order to compensate the energy variation of the coefficients in G_1. Hence an algorithm is developed in the following to ensure the validity of Inequality (12) while performing the QIM in G_1. Let us first define a pair of modulated amplitudes

$$\begin{bmatrix} \breve{c}_{k,\{1\}} \\ \breve{c}_{k,\{2\}} \end{bmatrix} = \begin{cases} \begin{bmatrix} \breve{c}_k \\ \breve{c}_k + \Delta \end{bmatrix} & \text{if } \breve{c}_k < c_k; \\ \begin{bmatrix} \breve{c}_k \\ \breve{c}_k - \Delta \end{bmatrix} & \text{if } \breve{c}_k \ge c_k. \end{cases} \quad \text{for } k \in G_1 \tag{13}$$

where $\breve{c}_{k,\{1\}}$ is the direct outcome of the QIM and $\breve{c}_{k,\{2\}}$ is the suboptimal alternative of the QIM in terms of squared error. The individual energy variation, termed g_k, due to the replacement of $\breve{c}_{k,\{1\}}$ by $\breve{c}_{k,\{2\}}$ for the k^{th} coefficient is thereby

$$g_k = \breve{c}_{k,\{2\}}^2 - \breve{c}_{k,\{1\}}^2, \text{ for } k \in G_1. \tag{14}$$

The algorithm starts with an initial setup of involved variables:

$$\hat{c}_k = \breve{c}_{k,\{1\}}; \tag{15}$$

$$\eta^{(0)} = \sum_{k \in G_1} \left(\hat{c}_k^2 - c_k^2 \right) = \sum_{k \in G_1} \breve{c}_{k,\{1\}}^2 - \sum_{k \in G_1} c_k^2, \tag{16}$$

where $\eta^{(0)}$ denotes the energy gap. The superscript (0) indicates the iteration number. The following part is an iterative procedure consisting of three steps.

Step 1. In the j^{th} iteration, we end the algorithm whenever $-\rho_{inc} \le \eta^{(j)} \le -\rho_{dec}$. If either $\eta^{(j)} > -\rho_{dec}$ or $-\rho_{inc} > \eta^{(j)}$ occurs, we search for the coefficient c_K that mostly reduces the energy gap, i.e.

$$K = \arg\min_k \left| \eta^{(j)} + g_k \right|. \tag{17}$$

Step 2. A new energy gap is subsequently obtained by

$$\eta^{(j+1)} = \eta^{(j)} + g_K. \tag{18}$$

Step 3. The value of $\eta^{(j+1)}$ is examined. If $\left| \eta^{(i+1)} \right| < \left| \eta^{(i)} \right|$, then we assign $\hat{c}_K = \breve{c}_{K,\{2\}}$ and $g_K = \infty$ before returning to Step 1. Otherwise, the algorithm is terminated.

By using the foregoing iterative algorithm, the inequality condition $-\rho_{inc} \le \eta^{(J)} \le -\rho_{dec}$ is often achieved within few iterations. We then adjust the magnitudes of the coefficients in G_2 to counteract the energy deviation emerging from the QIM process in G_1. Our strategy here is to evenly distribute the energy gap $\eta^{(J)}$ to the coefficients in G_2. Note that the way used to deal with the negative $\eta^{(J)}$ is

somewhat different from that with the positive $\eta^{(J)}$. In a situation where $\eta^{(J)} < 0$, the amplitude for the m^{th} coefficient can be simply modified as

$$\hat{c}_m = \mathrm{sgn}\left(c_m\right)\left(c_m^2 - \frac{\eta^{(J)}}{L_{G_2}}\right)^{1/2} \text{ for } m \in G_2, \tag{19}$$

where $\mathrm{sgn}(x)$ is the sign function defined as

$$\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ -1, & \text{if } x < 0. \end{cases} \tag{20}$$

When $\eta^{(J)} > 0$, every coefficient magnitude in G_2 is supposedly decreased by a certain amount. As the maximum permissible reduction for each coefficient is limited by its own magnitude, a simple algorithmic procedure is proposed below to resolve the difficulty. The entire procedure consists of only two steps. First, we sort the coefficient magnitudes in G_2 such that

$$\left|c_{m_0}\right| \leq \left|c_{m_1}\right| \leq \cdots \leq \left|c_{m_{L_{G_2}-2}}\right| \leq \left|c_{m_{L_{G_2}-1}}\right|, \; m_k \in G_2. \tag{21}$$

Next, we derive the corresponding coefficients one by one with the indexes counting from c_{m_0} to $c_{m_{L_{G_2}-1}}$:

$$\eta_{m_k}^{(J)} = \eta_{m_k}^{(J)} - \left(c_{m_k}^2 - \hat{c}_{m_k}^2\right); \tag{22}$$

$$\hat{c}_{m_k} = \mathrm{sgn}\left(c_m\right)\left(\max\left\{0, c_{m_k}^2 - \frac{\eta_{m_k}^{(J)}}{L_{G_2} - k}\right\}\right)^{1/2}. \tag{23}$$

For each frame, the embedding procedure ends whenever all the $\hat{c}_k$'s in G_1 and G_2 are properly modified. Throughout the modifications by either Eq. (19) or (23), the overall energy for the 64 DWT-DCT coefficients remains intact, i.e.

$$\sum_{k \in G_1} \hat{c}_k^2 + \sum_{k \in G_2} \hat{c}_k^2 = E_c = \sum_{k=0}^{63} c_k^2. \tag{24}$$

Eventually, with the energy compensation, the quantization step Δ derived from $\hat{c}_k$'s remains the same as that from c_k's. To summarize the foregoing discussion, we depicted a flowchart in Fig. 1 to provide a better understanding of the embedding process.

2.4 Frame Synchronization and Watermark Extraction

Just like many other watermarking methods, the proposed DWT-DCT-based scheme has been equipped with a synchronization technique [23-24] to withstand the de-synchronization attacks. The procedure for extracting watermark bits from a watermarked audio is rather simple. Prior to watermark extraction, we identify the position where the watermark is embedded using the standard synchronization technology of digital communications. Once the DWT-DCT coefficients, termed $\tilde{c}_k$'s, are obtained as the manner described at the beginning of Section 2, the quantization step $\tilde{\Delta}$ in each frame is acquired using Equation (9). The bit $\tilde{w}_b$ residing in each designated coefficient $\tilde{c}_k$ is determined by

$$\tilde{w}_b = \begin{cases} 1, & \text{if } \left|\tilde{c}_k / \tilde{\Delta} - \left\lfloor \tilde{c}_k / \tilde{\Delta} \right\rfloor - 0.5\right| < 0.25; \\ 0, & \text{otherwise.} \end{cases} \text{ for } k \in G_1. \tag{25}$$

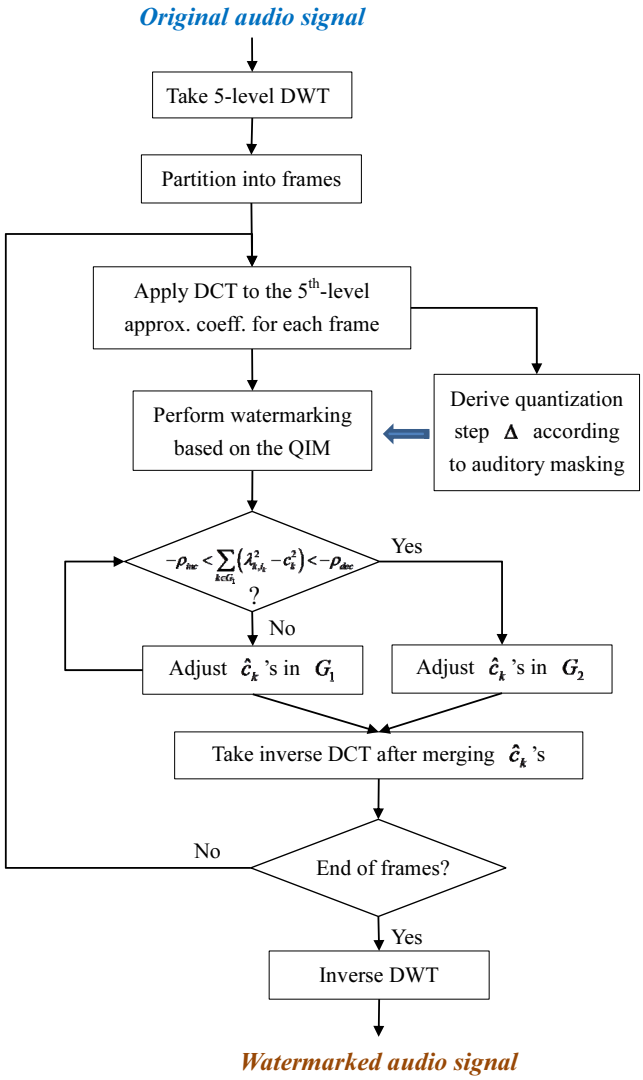

Fig. 1. The watermark embedding procedure of the compensated DWT-DCT scheme

3 Performance Evaluation

The primal DWT-DCT approach introduced in [20] was employed as the baseline for comparison. Similar to the manner adopted by the proposed scheme, we performed a 5^{th} level DWT over the audio signal and divided the 5^{th} level approximation and detail coefficients into frames of length 128. After taking DCT of the approximation and detail subbands in each frame, the watermark embedding was carried out by applying the QIM to the first 48 DWT-DCT coefficients in the approximation subband. The quantization step S was computed as

$$S = \eta \times \frac{\left\lfloor \left(\left| \overline{A(i)} \right| + \left| \overline{D(i)} \right| \right) \times 1000 + 0.5 \right\rfloor}{1000}, \tag{26}$$

where $\left| \overline{A(i)} \right|$ and $\left| \overline{D(i)} \right|$ represent the mean values of the magnitude DWT-DCT coefficients in the 5^{th} level approximation and detail subbands, respectively. η was tentatively chosen as 0.375 to reach a satisfactory tradeoff between robustness and imperceptibility.

In addition to the comparison with the primal DWT-DCT, the proposed scheme was compared in capacity, imperceptibility and robustness with four other recently developed methods, which were named in abbreviated form as SVD-DCT [9], DWT-SVD [16], DWT-norm [26] and LWT-SVD [25]. Following

their original specifications, the payload capacity for the SVD-DCT, DWT-SVD, DWT-norm and LWT-SVD are 43, 45.56, 102.4 and 170.67 bits per second (bps), respectively.

In our experiments, the variables α and β used in the SVD-DCT to derive the linear model of the frequency mask were set to 0.125 and 0.1 respectively. In the DWT-SVD method, the minimum and maximum values for quantization steps were Δ_m=0.6 and Δ_M=0.9 respectively. The two user-defined weight parameters were S_{mean}=0.1 and S_{std}=0.6. In the implementation of the DWT-norm, the variables α_1 and α_2 used to control quantization steps were assigned as 0.4 and 0.2 respectively and the variable "$attack\_SNR$" was set as 20 dB. For the LWT-SVD, the decomposition level of the lifting wavelet transform was 3 and the quantization step size was 0.45. All the foregoing parameters were selected to render adequate signal-to-noise ratios (SNR) so that the resulting performance can be appraised at a comparable basis.

The test materials comprised twenty 30-second music clips collected from various CD albums, including vocal arrangements and ensembles of musical instruments. All audio signals were sampled at 44.1 kHz with 16-bit resolution. The watermark bits for the test were a series of alternate 1's and 0's long enough to cover the entire host signal. Such an arrangement is particularly useful when we want to perform a fair comparison for watermarking methods with different capacities.

The quality of the watermarked audio signals is evaluated using the SNR defined in Equation (27) along with the perceptual evaluation of audio quality (PEAQ) [27].

$$SNR = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1}\breve{s}^2(n)}{\sum_{n=0}^{N-1}\left(\breve{s}(n)-s(n)\right)^2}\right). \tag{27}$$

The PEAQ renders an objective difference grade (ODG) between -4 and 0, signifying a perceptual impression from "very annoying" to "imperceptible". Table 1 provides a general interpretation with respect to typical ODG scores. In this study, we adopted the program released from the TSP Lab in the Department of Electrical and Computer Engineering at McGill University [27]. Because the final outcome is derived from an artificial neural network that simulates the human auditory system, the PEAQ may come up with a value higher than 0.

Table 1. Impairment grades of the PEAQ

Impairment description	ODG
Imperceptible	0.0
Perceptible, but not annoying	−1.0
Slightly annoying	−2.0
Annoying	−3.0
Very annoying	−4.0

According to the statistical results shown in Table 2, the differences between the proposed perceptually energy-compensated scheme and the baseline are subtle. The average SNR's for both schemes are above 20 dB, which is a level recommended by the International Federation of the Phonographic Industry (IFPI) [19]. Basically, both the compensated and uncompensated schemes can achieve transparent watermarking since the resultant ODG's are very near 0. The proposed scheme renders a mean ODG of 0.005 with a small standard deviation of 0.088, suggesting that the resultant audio quality is not only exceptionally high but also remarkably stable. By contrast, the quality impairments resulting from the LWT-SVD and DWT-norm are within the acceptable range, whereas the DWT-SVD and SVD-DCT are merely on the fringe of acceptability.

As for the robustness test, this study examines the bit error rates (BER) between the original watermark W and recovered watermark $\tilde{W}$.

$$BER\left(W,\tilde{W}\right) = \frac{\sum_{m=0}^{M-1} W(m)\oplus\tilde{W}(m)}{M} \tag{28}$$

Table 2. Statistics of the measured SNR's and ODG's. The data in the second and third columns are interpreted as "mean [±standard deviation]"

Watermarking schemes	SNR in decibel	ODG	Payload (bps)
SVD-DCT	29.751 [±2.465]	-1.535 [±1.444]	43
DWT-SVD	22.403 [±2.584]	-1.871 [±1.359]	45.56
DWT-norm	24.629 [±2.253]	-0.305 [±0.565]	102.4
LWT-SVD	20.030 [±2.788]	-0.767 [±0.856]	170.67
Uncompensated DWT-DCT	20.990 [±0.665]	-0.014 [±0.206]	516.80
Compensated DWT-DCT	20.058 [±1.849]	0.005 [±0.088]	516.80

where $\oplus$ stands for the exclusive-OR operator and M is the length of the watermark bit sequence. The attacks types consist of resampling, requantization, amplitude scaling, noise corruption, filtering, AD/DA conversion, echo addition, jittering and MPEG3 compression. In this study, signal jittering was done by randomly deleting or adding one sample for every 100 samples within each frame. The DA/AD conversion was a process of converting a digital audio file to an analog signal and then resampling the analog signal at 44.1 kHz. Following the experimental setup in [28], the DA/AD conversion was performed through an onboard Realtek ALC892 audio codec, of which the line-out was connected to the line-in using a cable line during playback and recording.

Table 3 shows the BER's under various attacks. It is observed that the proposed scheme has rectified the rudimentary deficiency of the DWT-DCT scheme. When the attack is absent, the watermark recovered by the proposed compensation scheme reaches a 100% accuracy rate. The proposed scheme also demonstrates a perfect performance for the resampling, amplitude scaling and lowpass filtering with a cutoff frequency of 4 kHz. It is pointed out that the compensated DWT-DCT and SVD-DCT are the two methods completely surviving the amplitude scaling attack. For the DWT-SVD, LWT-SVD and DWT-norm methods, amplitude scaling can easily ruffle the embedded watermarks. Note also that the DA/AD conversion is equivalent to the composite effect of time-scaling, amplitude scaling and noise corruption. Based on our observation, the time-scaling effect caused by the ALC892 codec is not obvious. Most alterations come from the amplitude scaling. Consequently, the schemes incompetent to resist the amplitude scaling also fail to recover the watermarks in the case of DA/AD conversion.

Table 3. Averaged BER's obtained from the uncompensated and compensated DWT-DCT schemes

Attack type \ Method	SVD-DCT	DWT-SVD	DWT-norm	LWT-SVD	Uncompensated DWT-DCT	Compensated DWT-DCT
None	0.00%	0.00%	0.00%	0.00%	0.17%	0.00%
Resampling (44.1 kHz→22.05 kHz→44.1 kHz)	0.12%	0.29%	0.00%	0.00%	0.17%	0.00%
	0.00%	0.00%	0.00%	0.00%	0.17%	0.03%
Requantization (16 bits→8 bits→16 bits)	0.00%	53.32%	76.730%	60.47%	0.17%	0.00%
Amplitude sccaling (85%)	0.00%	0.00%	0.000%	0.00%	0.19%	0.21%
Noise corruption (SNR = 30 dB)	0.12%	0.05%	0.000%	0.00%	0.57%	0.92%
Noise corruption (SNR = 20 dB)	0.54%	3.47%	0.820%	0.00%	0.17%	0.00%
Lowpass filtering (@ 4 kHz)	46.17%	49.18%	50.020%	50.08%	42.69%	26.21%
Highpass filtering (@ 4 kHz)	49.75%	50.38%	49.140%	35.68%	45.94%	3.60%
Lowpass filtering (@ 500 Hz)	0.00%	47.10%	46.970%	47.53%	0.43%	0.40%
DA/AD conversion						
Echo addition (delay: 50ms; decay: 5%)	0.00%	0.77%	0.930%	0.78%	4.32%	4.10%
jitter (1/100)	0.36%	0.01%	0.000%	0.02%	0.35%	0.86%
MPEG3 128(kbps)	0.05%	0.00%	0.000%	0.00%	0.19%	0.01%
MPEG3 64(kbps)	0.56%	0.63%	0.990%	1.34%	2.32%	2.96%

One obvious advantage of the perceptually energy-compensated scheme is that it survives the extreme lowpass filtering. This is not surprising at all, since the watermark is embedded in the frequency band below 345 Hz. What surprises us is that the proposed scheme possesses certain resistance against highpass filtering. It appears that the proposed scheme can still retrieve some watermark bits from the filtered residual as long as the low frequency components are not completely obviated by highpass filtering. Nevertheless, a further inspection reveals that both the uncompensated and compensated DWT-DCT schemes may still suffer slight imperfection in the presence of noise corruption. The reason can be ascribed to the imperfect quantization step sizes retrieved from the noise-corrupted watermarked audio signal. Moreover, the additive white Gaussian noise will incidentally cause excessive alteration for few DCT coefficients, thus leading to erroneous judgment on embedded bits. An analogous explanation can be applicable to the results observed in the cases of echo addition and 64 kbps MPEG3.

4 Conclusion

A scheme is developed to compensate the energy variation due to the QIM watermarking at a specified frequency band in the DWT-DCT domain. This scheme offers a high payload capacity of 516.80 bps. During watermark embedding, the alterations due to the QIM and energy compensation are both constrained below the auditory masking threshold. The PEAQ scores confirm that the watermarked audio signal obtained from the energy-compensated DWT-DCT scheme is perceptually indistinguishable from the original audio signal. Our experimental results show that the energy compensation successfully remedies the imperfection in the previous design of the DWT-DCT framework. The watermark can be retrieved with 100% accuracy when no attack is present. Compared with the other four recently developed methods, the proposed DWT-DCT scheme demonstrates a significant better performance in imperceptibility and payload capacity, while its robustness against malicious attacks is comparable with, if not better than, others. Moreover, the proposed scheme can survive the extremely lowpass filtering and amplitude scaling attacks.

It is pointed out that the idea of perceptual QIM and energy-compensation methods is applicable to the rest part of the DWT-DCT coefficients, leading to the possibility that the payload capacity can be further increased. The robustness can also be enhanced by grouping multiple coefficients into a vector and performing the QIM based on the vector. Many of these issues will be explored in our future research.

Acknowledgement

This research work was supported by the Ministry of Science and Technology, Taiwan, ROC under grants MOST 103-2221-E-197 -020 and MOST 105-2221-E-197-019.

References

[1] P. Bassia, I. Pitas, N. Nikolaidis, Robust audio watermarking in the time domain, IEEE Trans. Multimedia 3(2)(2001) 232-241.

[2] W.-N. Lie, L.-C. Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification, IEEE Trans. Multimedia 8(1)(2006) 46-59.

[3] H. Wang, R. Nishimura, Y. Suzuki, L. Mao, Fuzzy self-adaptive digital audio watermarking based on time-spread echo hiding," Applied Acoustics 69(10)(2008) 868-874.

[4] L. Wei, X. Xiangyang, L. Peizhong, Localized audio watermarking technique robust against time-scale modification, IEEE Trans. Multimedia 8(1)(2006) 60-69.

[5] R. Tachibana, S. Shimizu, S. Kobayashi, T. Nakamura, An audio watermarking method using a two-dimensional pseudo-random array, Signal Processing 82(10)(2002) 1455-1469.

[6] D. Megías, J. Serra-Ruiz, M. Fallahpour, Efficient self-synchronised blind audio watermarking system based on time domain

and FFT amplitude modification, Signal Processing 90(12)(2010) 3078-3092.

[7] X.-Y. Wang, H. Zhao, A novel synchronization invariant audio watermarking scheme based on DWT and DCT, IEEE Trans. Signal Processing 54(12)(2006) 4835-4840.

[8] I.-K. Yeo, H.J. Kim, Modified patchwork algorithm: a novel audio watermarking scheme, IEEE Trans. Speech and Audio Processing 11(4)(2003) 381-386.

[9] B.Y. Lei, I.Y. Soon, Z. Li, Blind and robust audio watermarking scheme based on SVD–DCT, Signal Processing 91(8)(2011) 1973-1984.

[10] B. Lei, I.Y. Soon, F. Zhou, Z. Li, H. Lei, A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition, Signal Processing 92(9)(2012) 1985-2001.

[11] X.-Y. Wang, P.-P. Niu, H.-Y. Yang, A robust digital audio watermarking based on statistics characteristics, Pattern Recognition 42(11)(2009) 3057-3064.

[12] S. Wu, J. Huang, D. Huang, Y.Q. Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission, IEEE Trans. Broadcasting 51(1)(2005) 69-76.

[13] X. Li, H.H. Yu, Transparent and robust audio data hiding in cepstrum domain, in: Proc. IEEE Int. Conf. Multimedia and Expo, 2000.

[14] S.C. Liu, S.D. Lin, BCH code-based robust audio watermarking in cepstrum domain, Journal of Information Science and Engineering 22(3)(2006) 535-543.

[15] H.-T. Hu, W.-H. Chen, A dual cepstrum-based watermarking scheme with self-synchronization, Signal Processing 92(4)(2012) 1109-1116.

[16] V. Bhat K, I. Sengupta, A. Das, An adaptive audio watermarking based on the singular value decomposition in the wavelet domain, Digital Signal Processing 20(6)(2010) 1547-1558.

[17] M. Steinebach, F.A.P. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, S. Seibel, N. Fates, L.C. Ferri, StirMark benchmark: audio watermarking attacks, in: Proc. Int. Conf. on Information Technology: Coding and Computing, 2001.

[18] J.J.K.Ò. Ruanaidh, T. Pun, Rotation, scale and translation invariant spread spectrum digital image watermarking, Signal Processing 66(3)(1998) 303-317.

[19] S. Katzenbeisser, F.A.P. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking/Stefan Katzenbeisser, Artech House, Boston, 2000.

[20] X. Wang, W. Qi, P. Niu, A new adaptive digital audio watermarking based on support vector regression, IEEE Trans. on Audio, Speech, and Language Processing 15(8)(2007) 2270-2277.

[21] X. He, M.S. Scordilis, An enhanced psychoacoustic model based on the discrete wavelet packet transform, Journal of the Franklin Institute 343(7)(2006) 738-755.

[22] B. Chen, G.W. Wornell, Quantization index modulation: a class of provably good methods for digital watermarking and information embedding, IEEE Trans. Information Theory 47(4)(2001) 1423-1443.

[23] H.-T. Hu, C. Yu, A perceptually adaptive QIM scheme for efficient watermark synchronization, IEICE Trans. Information and Systems E95-D(12)(2012) 3097-3100.

[24] H.-T. Hu, L.-Y. Hsu, H.-H. Chou, Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity, Digital Signal Processing 31(2014) 115-123.

[25] B. Lei, I. Yann Soon, F. Zhou, Z. Li, H. Lei, A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition, Signal Processing 92(9)(2012) 1985-2001.

[26] X. Wang, P. Wang, P. Zhang, S. Xu, H. Yang, A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform, Signal Processing 93(4)(2013) 913-922.

[27] P. Kabal, An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality, TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, 2002.

[28] S. Xiang, Audio watermarking robust against D/A and A/D conversions, EURASIP Journal on Advances in Signal Processing 2011(1)(2011) 3.