Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **5**, September **2015**

Optimal SNR of Audio Watermarking by Wavelet and Compact PSO Methods

Ming Zhao[1], Jeng-Shyang Pan[1,2*],and Shuo-Tsung Chen[3]

[1] Innovative Information Industry Research Center, Harbin Institute of Technology
Shenzhen Graduate School, Shenzhen 518055, P.R. China
hitmzhao@gmail.com

[2] College of Information Engineering and Science, Fujian University of Technology
Fuzhou, China

[3] Department of Mathematics, Tunghai University, Taichung 40704, Taiwan (R.O.C.)
*Corresponding author

Received February 2015; revised May 2015

Abstract. *Discrete wavelet transform (DWT) provides a useful platform for hidden digital information and copyright protection; therefore numerous DWT-based algorithms have been proposed in recent years. Signal-to-noise ratio (SNR) and bit-error-rate (BER) are commonly utilized performance indexes in measuring the fidelity, robustness and quality of watermarking algorithms. However, there is a tradeoff relationship between audio quality and robustness. To overcome the drawback, this paper aims in proposing an optimization-based watermarking scheme using optimal multi-coefficients quantization in the wavelet domain. Compact PSO, which has excellent performance with less hardware requirement, plays an essential role to obtain the optimal scaling factors. Experimental results confirm that the embedded audio in the proposed method has high SNR and low BER, indicating strong robustness against various attacks, such as re-sampling, amplitude modification, and mp3 compression.*

Keywords: Discrete wavelet transform; Signal-to-noise ratio; Bit-error-rate; Compact PSO; Optimal scaling; Almost invariant feature.

1. **Introduction.** An audio watermarking scheme generally fulfills three IFPI (International Federation of Phonographic Industry) requirements [1-2]: (1) Watermarks have to be imperceptible in the embedded audio. (2) The embedding design should offer more than 20dB signal-to-noise ratio (SNR) and 20 bps (bits per-second) embedding capacity for watermarked audio versus original one. (3) The embedded watermark should be able to resist common attacks, such as re-sampling, filtering, amplitude modification, time-scaling manipulation, and mp3 compression and so on.

Most audio watermarking techniques can be classified according to the type of watermark being used or the domain in which the watermark is applied. A number of methods reported in literature involve in inserting watermarks in either time domain [3-12] or frequency domain [2, 13-20]. Lie et al. [7] adopted the amplitude modification to improve robustness in time domain; however, the capacity and SNR were extremely low. Huang et al. [13] embedded watermark into discrete cosine transform (DCT) coefficients and hid barcode (the user-defined information) in the time domain as synchronization codes. Due to low embedding strength in the time domain, the synchronization codes are not adequately robust. On the other hand, for those synchronization codes hidden in the frequency domain, the computational complexity and cost increase explosively. To compensate the drawback, Wu et al. [2] used quantization index modulation method to embed synchronization codes and watermarks into low frequency coefficients in the discrete wavelet transform (DWT) domain. This technique achieves better robustness against common signal processing procedures and noise corruptions; however, it is still vulnerable to amplitude modification and time-scaling manipulation attacks due to the single-coefficient quantization. Xiang et al. [16] modified Lies method in wavelet domain and obtained slightly improved results.

There is a tradeoff relationship between fidelity and robustness of an embedded audio, which are typically assessed by the signal-to-noise ratio (SNR) and bit error ratio (BER). Chen et al. [18] proposed an optimization-based scheme to obtain the best embedded-audio quality using fixed scaling DWT coefficients. Even though the hidden data are robust against some common attacks in their approach, audio quality worsens with variation of modification factors and the embedded watermarks are inadequately robust to amplitude modification attack. The purpose of this paper is to propose an optimization-based scheme with the inclusion of scaling DWT coefficients in wavelet domain. This novel method is designed to ensure the best fidelity of an embedded audio.

Experimental results confirm that such optimization-based scaling scheme indeed provides good audio quality and adequate robustness against common attacks.

The paper is organized as following. Overview of DWT and mathematical clarification of Lagrange principle is presented in Section 2. In Section 3, introduction of the proposed optimization-based scaling scheme, the wavelet-based functional connecting the multi-coefficients quantization equation and the performance index are illustrated. The optimization process for DWT coefficients and their corresponding scaling factors are enlightened as well. The extraction technique and the almost invariant feature of the optimal scaling factors under amplitude scaling attack are also derived in this section. Experimental results are presented in Section 4; some remarks are concluded in Section 5.

2. RELATED WORKS.

2.1. Discrete Wavelet Transform.

The DWT has been extensively employed in many digital watermarking applications. In this section, we will briefly introduce concepts in DWT. The wavelet transform maps a function in $L^2(R)$ onto a scale-space plane. Wavelets are obtained by a single prototype function (mother wavelet) $\psi(x)$ which is regulated with a scaling parameter and a shift parameter [21, 22]. The discrete normalized scaling and wavelet basis function are defined as

$$\varphi_{i,n}(t) = 2^{i/2} h_i \varphi(2^i t - n), \tag{1}$$

$$\psi_{i,n}(t) = 2^{i/2} g_i \psi(2^i t - n), \tag{2}$$

where i and n are the dilation and translation parameters; h_i and g_i are the low-pass and high-pass filters. Orthogonal wavelet basis functions not only provide simple calculation in coefficients expansion but also span $L^2(R)$ in signal processing. As a result, any audio signal $S(t) \in L^2(R)$ can be expressed as a series expansion of orthogonal scaling functions and wavelets. More specifically,

$$S(t) = \sum_{\ell} c_{j_0}(\ell)\varphi_{j_0,\ell}(t) + \sum_{k} \sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(t), \tag{3}$$

where $c_j(\ell) = \int_R S(t)\varphi_{j,\ell}(t)dt$ and $d_j(k) = \int_R S(t)\psi_{j,k}(t)dt$ denote the sequences of low-pass and high-pass coefficients, respectively; j_0 be the integer to define an interval on which $S(t)$ is piecewise constant. Throughout this paper, the host digital audio signal , $S(n), n \in N$, denoting samples of the original audio signal $S(t)$ at the nth sample time, is cut into segments where DWT will be preformed. This can be done by exploiting orthogonal basis to implement DWT through filter bank. Fig. 1 demonstrates how the input digital audio signal $S(n)$ is segmented into eight non-overlapping multi-resoluton sub-bands by the seven-level DWT decomposition. For the consideration of robustness of low-pass filtering, the synchronization codes and watermarks are embedded into the seventh level low-frequency sub-band coefficients (i.e., the sub-band A7 in Fig. 4), which will be referred as the lowest-frequency DWT coefficients in this paper.

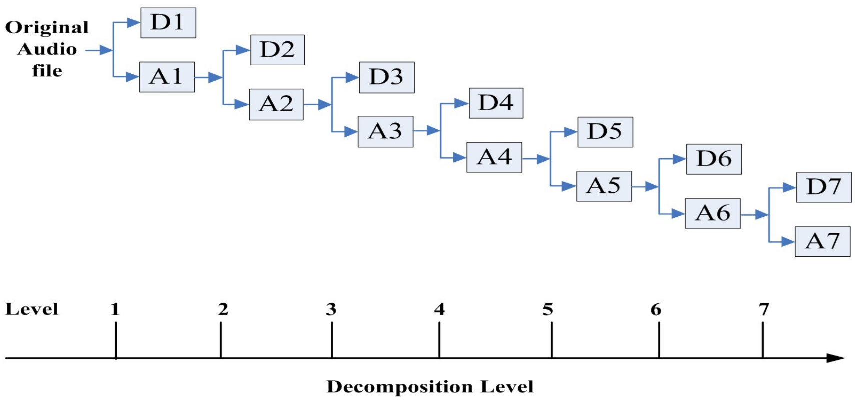

Figure 1. Seven-level discrete wavelet transformation.

2.2. **Compact PSO.** Compact PSO [24] was proposed by Neri et al. in 2013. This optimization algorithm was proposed based on novel search logic. It has features of standard Particle Swarm Optimization (PSO) algorithms, but it is unlike classical PSO algorithms, cPSO employs a probabilistic model to represent the swarms solution set. Neither the positions nor the velocities were not stored. And only a particle was used in the whole algorithm. Thus, a modest memory space is required. It was well suited for the embedded equipment with limited hardware.

In cPSO, the probabilistic model is a Perturbation Vector PV_{lb}, it consists of μ and σ, which are respectively, mean and standard deviation values of a Gaussian Probability Density Function (PDF) truncated within the interval [-1, 1] for each dimension of the solution. The PV_{lb} is used to generated a new solution which is considerate as the local best x_{lb}. So cPSO is based on virtual population. μ and σ is updated by formula (4) and (5). The updated μ and σ will help to generated more efficient solution because of this novel updating rule.

$$\mu^{t+1}(i) = \mu^t(i) + \frac{1}{N_p}\left[winner(i) - loser(i)\right], \tag{4}$$

$$\left[\sigma^{t+1}(i)\right]^2 = \left[\sigma^t(i)\right]^2 + \left[\mu^t(i)\right]^2 - \left[\mu^{t+1}(i)\right]^2 + \frac{1}{N_p}\left\{\left[winner(i)\right]^2 - \left[loser(i)\right]^2\right\}, \tag{5}$$

Where N_p is the size of the virtual population, and *winner* and *loser* are individuals with better fitness and worse fitness between existed position and x_{lb} or x_{gb}.

After new solution (particle) is generated, the movement of the particle in cPSO is the same to the stand PSO algorithm, and the position x and velocity v are updated by formula (6) and formula (7):

$$v^{t+1} = \phi_1 v^t + \phi_2 \left(x_{lb}^t - x^t\right) + \phi_3 \left(x_{gb}^t - x^t\right), \tag{6}$$

$$x^{t+1} = x^t + v^{t+1}, \tag{7}$$

where, as mentioned in [5], x^t indicates the current position of the particle, x_{lb} is the best position visited by the particle. the vector v^t is also generated by the Perturbation Vector, ϕ_1, ϕ_2, and ϕ_3 are three weight factors which might be constant or randomized. The details for cPSO are given below:

Step 1: initialize μ and σ of probabilistic model vector PV_{lb}.

Step 2: generate the global best solution x_{gb} by means of perturbation vector PV_{lb}.

Step 3: randomly generate x^t and v^t for the particle.

Step 4: generate the local best solution x_{lb} from PV_{lb}.

Step 5: updating x^t and v^t according to the standard PSO algorithm updating rule.

Step 6: compare x_{lb} and x^{t+1}, let be the one with better fitness, and the is the other one.

Step 7: updating μ and σ according to formula (6) and formula (7).

Step 8: compare x_{lb}, x_{gb} and x^{t+1} , assign the best solution to x_{gb}.

Step 9: if it is not meet the termination condition go to step 4.

Step 10: output the global best solution x_{gb}.

The experimental results of this method show that cPSO display a pretty performance than the corresponding population-based algorithms and compact algorithms. Because of its modest memory requirement, it is used to solve the optimization problem in those hard ware environments is limited due to cost and space limitations.

3. THE PROPOSED OPTIMIZATION-BASED EMBEDDING AND EXTRACTION.

In this section, the proposed optimization-based embedding with optimal scaling on DWT coefficients and extraction are presented.

3.1. Embedding Technique.

Since the watermarked audio may suffer the attack of shifting or cropping of a certain segment inside the audio signal, synchronization codes need to be embedded jointly with the watermark. Through examination of the synchronization codes, position of the embedded watermark can also be recognized. Prior to the embedding process, synchronization codes and watermark are arranged into a binary pseudorandom noise (PN) sequence $\{\beta_i\}, \beta_i = 0$ or 1. To ensure strong robustness, embedding is not applied to all DWT coefficients; that is, $\{\beta_i\}$ is only embedded successively into the lowest-frequency sub-band coefficients of each segment. Assemble every N consecutive absolute DWT coefficients and put them in the vector form $\mathbf{C}_N = [\ |c_1| \quad |c_2| \quad \cdots \quad |c_N|\ ]^T$. The rule for embedding $\{\beta_i\}$ becomes:

- if the bit $\beta_i = 1$ is embedded in C_N, the group-amplitude $\sum\limits_{j=1}^{N} |c_j|$ is quantized to

$$\gamma_1 = \left\lfloor \sum_{j=1}^{N} |c_j| \Big/ Q \right\rfloor Q + \frac{3}{4}Q; \tag{8}$$

- if the bit $\beta_i = 0$ is embedded in C_N, the group-amplitude $\sum_{j=1}^{N} |c_j|$ is quantized to

$$\gamma_0 = \left\lfloor \sum_{j=1}^{N} |c_j| \Big/ Q \right\rfloor Q + \frac{1}{4}Q, \tag{9}$$

where $\{c_j\}$ be the sequence of DWT lowest-frequency sub-band coefficients of the original signal; $\lfloor\rfloor$ indicates the floor function, and Q is the quantization parameter.

3.2. **Amplitude Scaling Factors.** Consider a positive scalar matrix $\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \end{bmatrix}$ whose entries are arbitrarily assigned by an encoder; we need to determine unknown values of watermarked lowest-frequency absolute DWT coefficients, $\bar{\mathbf{C}}_N = \begin{bmatrix} |\bar{c}_1| & |\bar{c}_2| & \ldots & |\bar{c}_N| \end{bmatrix}^T$, which relate to the original DWT coefficients C_N, such that

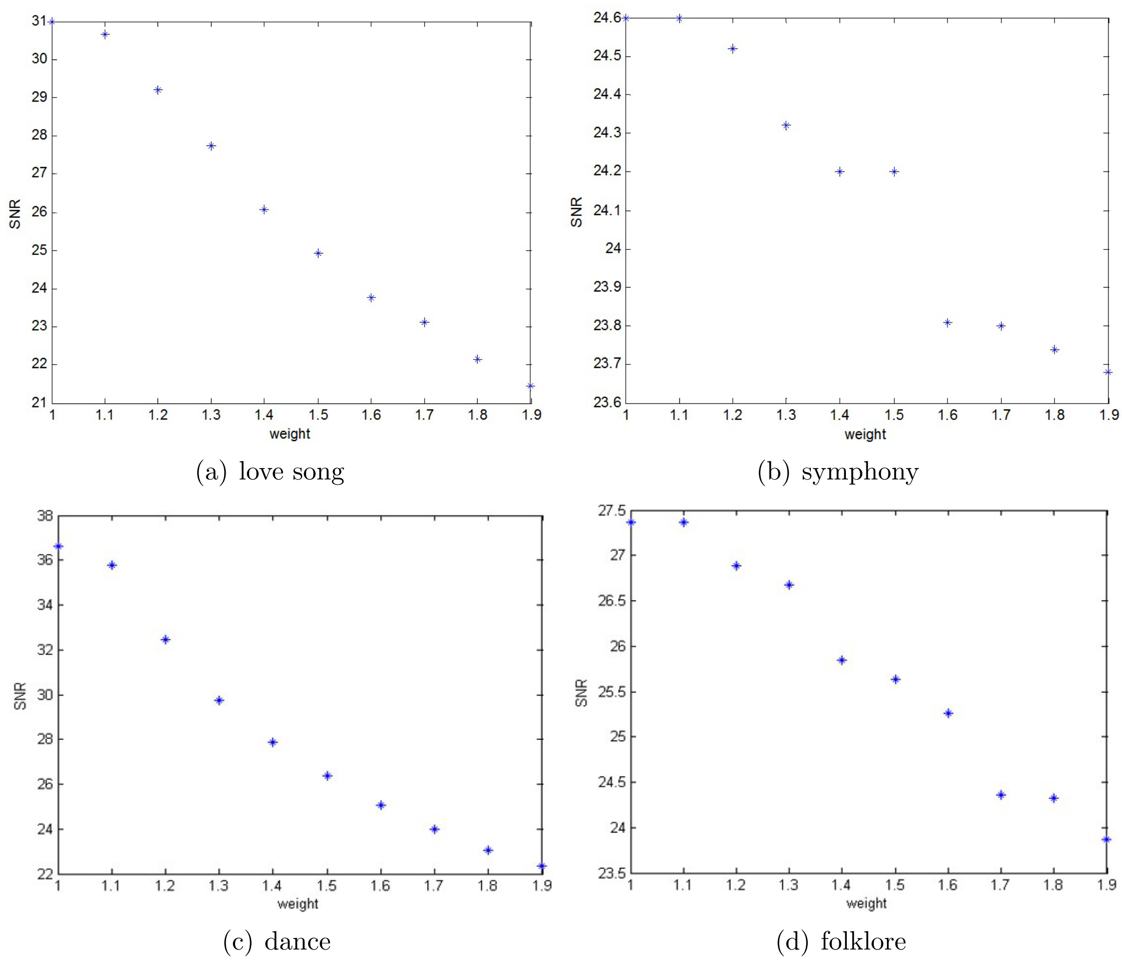

FIGURE 2. The variation of SNR versus a_2 for four types of songs when N=2 and M=2.

$$\sum_{j=1}^{N} a_j |\bar{c}_j| = \begin{cases} \gamma_1, \text{ if} \beta_i = 1 \\ \gamma_0, \text{ if} \beta_i = 0. \end{cases} \tag{10}$$

To avoid the situation that some entries in A may become arbitrarily large, without loss of generality, we assume all entries of A sum up to a constant M. For example, $A = \begin{bmatrix} 0.9 & 1.2 & 1.2 & 0.7 \end{bmatrix}$ represents a suitable selection for $N = 4$ and $M = 4$.

To better illustrate the effect of scaling factors adjustment on SNR_L, a case with $a_1 + a_2 = 2$ and their relationship are shown in Fig. 2. It is clear that the SNR_L drops as a_2 decreases and is maximized when $a_2 = 2$, which corresponds to $a_1 = 2$ as well. For that reason, it is crucial to investigate the scaling factors so to ensure the SNR_L attains its maximum.

3.3. **Optimal Embedding.** The SNR between the original audio $S(n)$ and the watermarked audio $\bar{S}(n)$ is formulated as formula (11).

$$SNR = -10 \cdot \log \left(\frac{\sum\limits_n \left\| \bar{S}(n) - S(n) \right\|_2^2}{\sum\limits_n \left\| S(n) \right\|_2^2} \right). \tag{11}$$

By Parseval Theorem, the square sum of $S(n)$ over all the samples is the same as the square sum of all the corresponding DWT coefficients. Due to the fact that the DWT coefficients are implemented with orthogonal wavelet bases

$$SNR = -10 \cdot \log \left(\frac{\left\| \bar{\mathbf{C}}_N - \mathbf{C}_N \right\|_2^2}{\left\| \mathbf{C}_N \right\|_2^2} \right) = -10 \cdot \log \left(\frac{\sum\limits_{i=1}^{N} \left(|\bar{c}_i| - |c_i| \right)^2}{\sum\limits_{i=1}^{k} |c_i|^2} \right). \tag{12}$$

Obviously, the formula (refE12) is a strictly deceasing function, the *SNR* is maximized when $\sum_{i=1}^{N} \left(|\bar{c}_i| - |c_i| \right)^2 \Big/ \sum_{i=1}^{N} |c_i|^2$ is minimized. The expression (13) is subject to the constraint (14).

$$\text{minimize} \qquad \frac{\sum\limits_{\mathrm{i=1}}^{\mathrm{N}} \left(|\bar{\mathrm{c}}_{\mathrm{i}}| - |\mathrm{c}_{\mathrm{i}}| \right)^2}{\sum\limits_{\mathrm{i=1}}^{\mathrm{k}} |\mathrm{c}_{\mathrm{i}}|^2} \tag{13}$$

$$\text{subjectto} \quad \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{a_i} \left| \bar{\mathrm{c}}_{\mathrm{i}} \right| = \gamma_1 \ \ and \ \ \sum\limits_{i=1}^{N} a_i = M \ . \tag{14}$$

3.4. **The design for the compact PSO.** The minimization of formula (14) is a continuous optimization problem; in addition, the application domain is MP3 player with limited hardware. So compact PSO [24] can solve this problem well. We can use formula (14) as fitness function of cPSO, and the designed variable is wavelet coefficient . Thus, the problem can be described as follows:

A designed particle $(x_1, x_2, ...x_i)$ with fitness function $\sum_{i=1}^{N} \left(|\overline{x}_i| - |x_i| \right)^2 \Big/ \sum_{i=1}^{N} |x_i|^2$ is subjected to $\sum\limits_{i=1}^{N} a_i \left| \overline{x}_i \right| = \gamma_1$ and $\sum\limits_{i=1}^{N} a_i = M$. The global best particle x_{gb} is required in this case.

3.5. **Extraction Technique.** Since the proposed scheme is a blind watermarking, the original audio may not be available in actual scenarios. The extraction process is simply a direct reverse of the embedding procedure. We first divide the test audio into segments and perform DWT on each section. Let $\hat{\mathbf{C}}_N = [|\hat{c}_1||\hat{c}_2| \cdots |\hat{c}_N|]^T$ denote the coefficients of N consecutive absolute DWT low-frequency sub-band coefficients in the seventh level of each segment from the test signal. The binary PN sequence including the watermark $\{\hat{\beta}_i\}$

will be extracted from $\hat{\mathbf{C}}_N$ and optimal amplitude scaling factors a_j^*, which are obtained during the embedding process, according to the rules:

- If $\sum_{j=1}^{N} a_j^* |\hat{c}_j| - \left\lfloor \sum_{j=1}^{N} a_j^* |\hat{c}_j| \Big/ Q \right\rfloor Q \geq \frac{Q}{2}$, then $\hat{\beta}_i = 1$;
- If $\sum_{j=1}^{N} a_j^* |\hat{c}_j| - \left\lfloor \sum_{j=1}^{N} a_j^* |\hat{c}_j| \Big/ Q \right\rfloor Q < \frac{Q}{2}$, then $\hat{\beta}_i = 0$.

4. **EXPERIMENTAL RESULTS.** The evaluation of the proposed optimization-based amplitude quantization for audio watermarking is discussed in this section. Four types of audio signals, love song, symphony, dance, and folklore, are to be tested. These audio signals are 16-bit mono-type of length 11.6 seconds and sampling rate 44.1 kHz. Prior to the embedding process, the original audio is cut into four segments and without loss of generality, set $M = N$. Apply seven-level decomposition DWT to embed synchronization codes and watermarks into the lowest-frequency sub-band coefficients. Accordingly, the embedding capacities are set to be 2000, 1000, and 500 bits for $N = 2, 4, 8$, respectively. The SNR in Wus approach [2] is measured under single-coefficient amplitude quantization of size $Q = 6500$. In order to make a contrast, the values of Q in this study are set to be 13000, 26000 and 52000 for $N = 2, 4, 8$, respectively.

4.1. **Quality Evaluation of Watermarked Audios.** Table I contains information of the embedding domain (DWT level), embedding capacity, and SNR for various methods. Due to the fact that every N lowest DWT coefficients are optimized, one can see that higher SNR appears in the proposed scheme. Even though such scheme is initially designed to maximize the SNR, the data shows that it turns out to provide higher SNR comparing to other approaches as well. Although the increase in group size of DWT coefficients can give us higher SNR, but there is a limit. We note that the embedding capacity of our scheme satisfies the IFPI requirement - providing at least 20 bps embedding capacity and if the group size greater than 16 this requirement is violated. Hence in our study only three kinds of group sizes, i.e., $N = 2, 4, 8$, are considered.

4.2. **Robustness Measurement.** After the embedding process, some common attacks are applied to test the robustness which will be measured by the bit error rate (BER). The BER, the ratio of bit errors to the total transferred errors during a tested time interval, is usually expressed in percentage and can be formulated as

$$\text{BER} = \frac{B_{error}}{B_{total}} \times 100\%, \tag{15}$$

where B_{error} and B_{total} denote the numbers of error binary bits and total binary bits during a tested period. Five types of attacks that will be applied to the audio signals, including re-sampling, low-pass filtering, amplitude modification, time-scaling manipulation and mp3 compression, will be introduced below.

1) *Re-sampling*: The sampling rate of the audio signal in the re-sampling procedure can be increased or decreased by a factor in three steps: (i) down-sample (ii) interpolation (iii) up-sample. In this study, the watermarked audio was down-sampled from 44.1kHz to 22.05kHz, and then increased back to 44.1kHz in the form of linear interpolation filter. Similarly, the sampling rates were down-sampled from 44.1kHz to 11.025kHz and 8kHz, and then moved back up to 44.1kHz following the steps (i)-(iii). The BER results of re-sampling manipulation to four types of audios are presented in Table II. These experimental results confirm that the proposed scheme has lower BER, which corresponds to higher robustness, than those by Wu et al, [2] and Chen et al. [18]. In

addition, comparing with works by Xiang et al [16] and Chen et al. [17], the proposed method shows comparable robustness as but is definitely more robust in the case of $N = 8$.

2) *Low-pass filtering*: Table III shows the effect of adopting a low-pass filter with the cutoff frequencies 3 kHz and 6kHz. The proposed method is more resistant to such attack than Wus work in [2] and has slightly lower robustness than Xiangs [16], Chens[17], and Chens [18] even when $N = 8$, except for the audio type symphony. As we expected that as the cutoff frequency increases the BER is getting better.
3) *Amplitude modification*: Since a bigger modification factor results in saturation, the amplitude modification factor τ is set as 0.5, 0.8, 1.1, and 1.2 in the study. The comparison data in Table IV clearly illustrate better performance and strong robustness provided by the proposed scheme than Wus[2], Xiangs [16] and Chens [18] but slightly less robustness than Chens [17]. This is simply due to the almost invariance property those obtained optimal scaling factor holds.
4) *Time-scaling manipulation*: Experimental results of the watermarked audios under time-scaling attack of ±2% and ±5% are listed in Table V. These outcomes evidently indicate slightly better robustness our scheme presents than those in the approaches of Wus [2], Chens [17] and Chen [18]. When comparing with Xiangs work, it is slightly less robust with positive time-scaling attack but offers somewhat higher robustness with negative time-scaling attack.
5) *MP3 compression*: MP3 compression is an audio compression method. Table VI show the experiments of applying MP3 compression at different bit rates to the watermarked audio. We find the proposed scheme has good performance and is more robust than other methods.

5. **Conclusions.** This study presents an optimization-based quantization scheme with optimal scaling on wavelet coefficients. To enhance the robustness, the watermarks are embedded into grouped lowest-frequency DWT coefficient. These DWT coefficients, along with corresponding scaling factors, are all optimized. Experimental results confirm that the embedded audios have high SNR, small SDG, and better robustness against signal processing and common attacks, such as re-sampling, amplitude modification and mp3 compression. For the low-pass filtering the result of our scheme is similar to those results from related wavelet based schemes. However, similar to other previously developed wavelet domain methods in literature, the proposed scheme is not that robust to the time-scaling manipulation. Our future work will aim in proposing an invariant feature in the wavelet domain to improve the low robustness from low-pass filtering attack and time-scaling manipulation.

REFERENCES

[1] S. Katzenbeisser, F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, London, 1999.

[2] S. Wu, J. Huang, Y. Q. Shi, Efficiently self-synchronized audio watermarking for assure audio data transmission, *IEEE Transactions on Broadcasting*, vol 51, no.1, pp. 69-76, 2005.

[3] M. A. Gerzon, P. G. Graven, A high-rate buried-data channel for audio CD, *Journal of the Audio Engineering Society*, vol 43, no 1/2, pp. 3-22, 1995.

[4] D. Gruhl, A. Lu, Echo hiding, *the 1st Information Hiding Workshop*, LNCS 1174, Springer, Berlin, Germany, pp.295-315, 1996.

[5] H. J. Kim, Y. H. Choi, J. W. Seok, J. W. Hong, Audio watermarking techniques, *Pacific Rim Workshop on Digital Steganography*, Kitakyushu, Japan, pp.30-40, 2003.

[6] H. Alaryani, A. Youssef, A novel audio watermarking technique based on frequency components, *the Seventh IEEE International Symposium on Multimedia*, Irvine, California, pp.668-673, 2005.

[7] W. N. Lie, L. C. Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. *IEEE Transactions on Multimedia*, vol 8, no 1, pp.46-59, 2006.

[8] L. Boney, Digital watermarks for audio signals, *International Conference on Multimedia Computing and Systems*, Hiroshima, Japan, pp.473-480, 1996.

[9] T. Ciloglu, S. U. Karaaslan, An improved all-pass watermarking scheme for speech and audio, *IEEE International Conference on Multimedia Computing and Exposition*, New York, USA, pp.1017-1020, 2000.

[10] H. Kim, Stochastic model based audio watermark whitening filter for improve detection, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, pp.1971-1974, 2000.

[11] P. Bassia, I. Pitas, N. Nikolaidis, Robust audio watermarking in the time domain, *IEEE Transaction on Multimedia*, vol 3, no 2, pp.232-241, 2001.

[12] B. S. Ko, Time-spread echo method for digital audio watermarking using PN sequence, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.2001-2004, 2002.

[13] J. Huang, S. Wang, Y. Q. Yang, A blind audio watermarking algorithm with self-synchronization, *IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, pp.627-630, 2002.

[14] W. L. Lyu, C. C. Chang, T. S. Nguyen, C. C. Lin, Image Watermarking Scheme Based on Scale-Invariant Feature Transform, *KSII Transactions on Internet and Information Systems*, vol 8, no 10, pp.3351-3606, 2014.

[15] C. P. Wu, Robust frequency domain audio watermarking based on audio content analysis, *International Symposium on Multimedia Information Processing*, Wuhan, China, 37-45, 1999.

[16] S. Xiang, J. Huang, Robust audio watermarking against the D/A and A/D conversions, arXiv:0707.0397v1, 2007.

[17] S. T. Chen, H. N. Huang, C. J. Chen, G. D. Wu, Energy-proportion based scheme for audio watermarking, *IET Proceedings on Signal Processing*, vol 4, no 5, pp.576-587, 2010.

[18] S. T. Chen, G. D. Wu, H. N. Huang, Wavelet-domain audio watermarking scheme using optimization-based quantization, *IET Proceedings on Signal Processing*, vol 4, no 6, pp. 720-727, 2010.

[19] C. Y. Yang, C. H. Lin, W. C. Hu, Reversible Data Hiding By Adaptive IWT-coefficient Adjustment, *Journal of Information Hiding and Multimedia Signal Processing*, vol 2, no 1, pp.24-32, 2011.

[20] R. M. Noriega, M. Nakano, B. Kurkoski, K. Yamaguchi, High Payload Audio Watermarking: toward Channel Characterization of MP3 Compression, *Journal of Information Hiding and Multimedia Signal Processing*, vol 2, no 2, pp.91-107, 2011.

[21] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transaction on Pattern Anal and Machine Intelligence*, pp. 674-693, 1989.

[22] C. S. Burrus, R. A. Gopinath, H. Gao, *Introduction to Wavelet Theory and Its Application*, Prentice-Hall, New Jersey, USA, 1998.

[23] F. Neri, E. Mininno, G. Lacca, Compact Particle Swarm Optimization, *Information Science*, vol 23, no 9, pp.96-121, 2013.

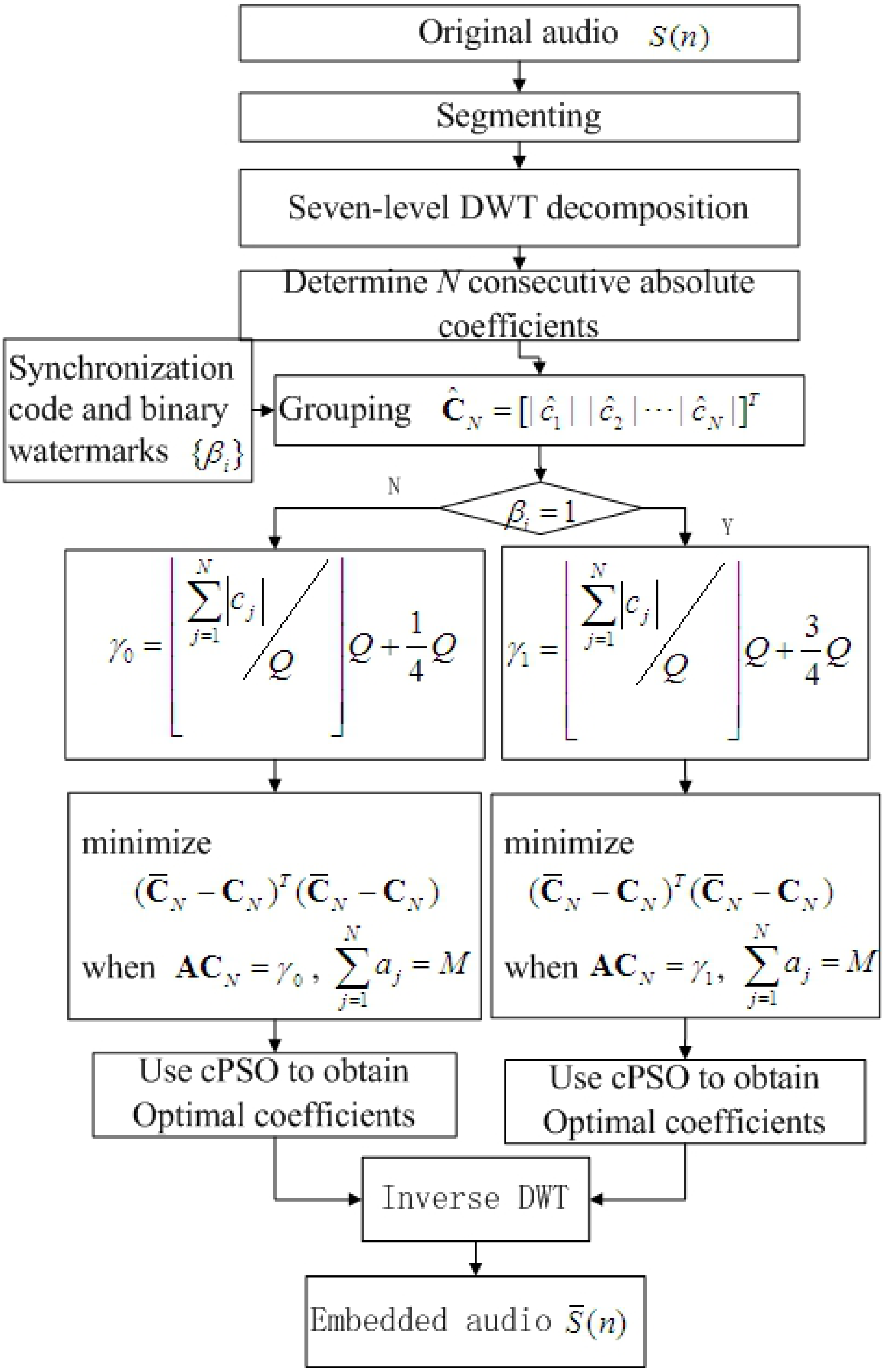

FIGURE 3. Optimal watermark embedding process.

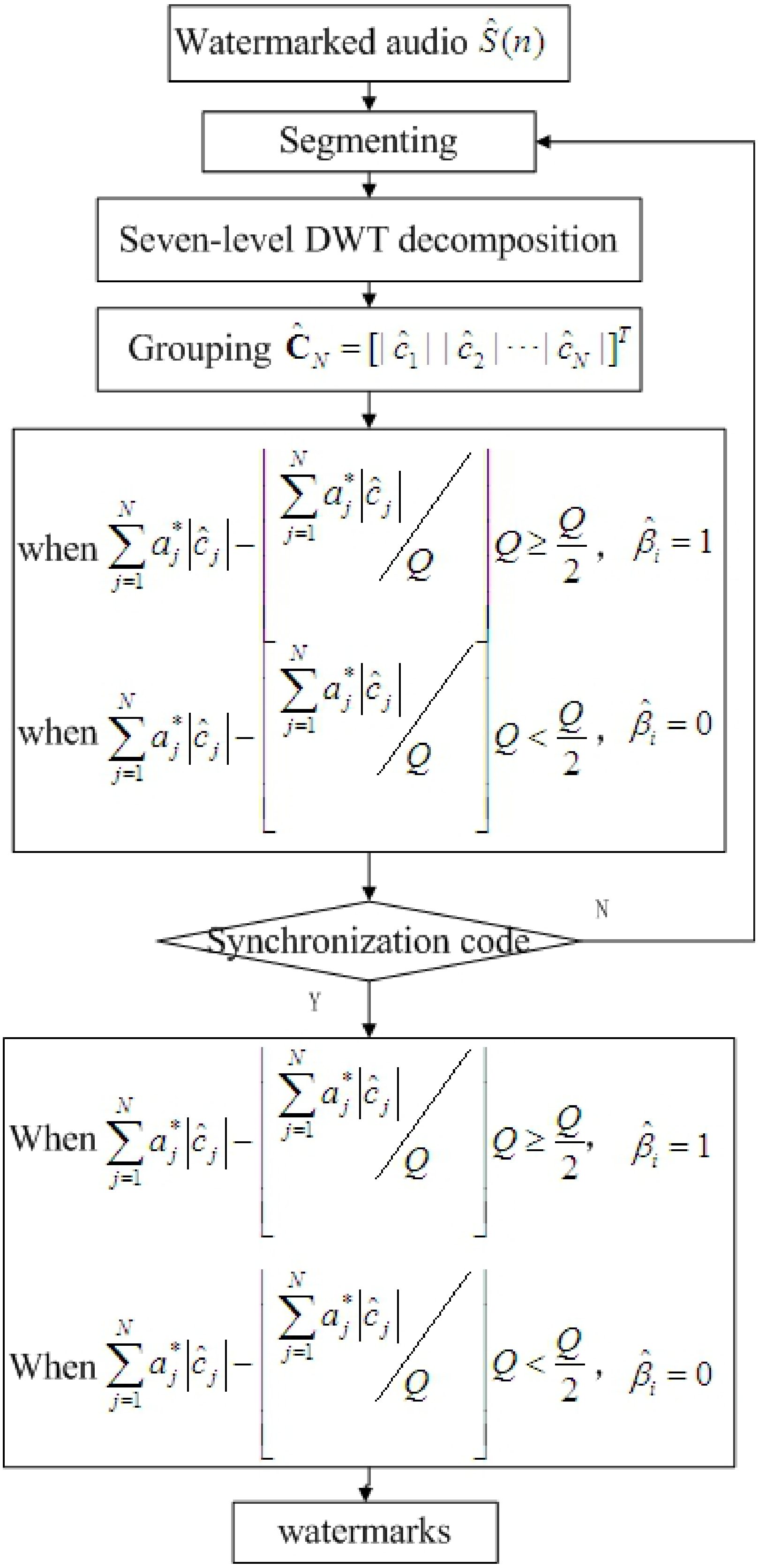

FIGURE 4. Watermark extraction process.

Table 1. PSNR and payload values of the proposed scheme with different codebook sizes.

		Domain	Embedding capacity	SNR(dB)	
Wu et al. [2]		DWT 8-level	172 bps (2000bits/11.6seconds)	33.6 (love song) 36.4 (dance)	26.7 (symphony) 27.2 (folklore)
Xiang et al. [16]		DWT 7-level	57 bps (666bits/11.6seconds)	17.8 (love song) 16.5 (dance)	18.2 (symphony) 17.4 (folklore)
Chen et al. [17]		DWT 7-level	115 bps (1333bits/11.6seconds)	17.8 (love song) 15.6 (dance)	17.2 (symphony) 17.3 (folklore)
Chen et al. [18]		DWT 7-level	172 bps (2000bits/11.6seconds)	35.4 (love song) 33.7 (dance)	26.8 (symphony) 24.5 (folklore)
Proposed Method	$N = 2$	DWT 7-level	172 bps (2000bits/11.6seconds)	35.4 (love song) 39.6 (dance)	26.8 (symphony) 28.9 (folklore)
	$N = 4$	DWT 7-level	86 bps (1000bits/11.6seconds)	41.5 (love song) 36.5 (dance)	33.6 (symphony) 34.5 (folklore)
	$N = 8$	DWT 7-level	43 bps (500bits/11.6seconds)	41.4 (love song) 36.4 (dance)	33.9 (symphony) 34.4 (folklore)

Table 2. BER (%) IN THE RE-SAMPLING ATTACK.

Audio Type		long song			symphony			dance			folklore		
Re-sampling Rate(kHz)		22.5	11.03	8	22.05	11.03	8	22.05	11.03	8	22.05	11.03	8
Wu et al. [2]		7.75	19.04	18.28	0.80	1.28	1.32	14.25	28.35	27.55	0.85	10.65	10.20
Xiang et al. [16]		1.24	7.82	8.25	0.42	1.02	1.26	3.21	8.92	9.03	0.44	2.06	2.42
Chen et al. [17]		1.42	3.50	3.35	4.23	4.84	5.92	1.02	1.62	1.24	0.83	2.02	1.95
Chen et al. [18]		4.86	11.37	11.68	0.72	1.31	1.32	9.02	14.48	14.46	0.78	4.56	4.54
Proposed Method	N=2	4.82	11.26	0.65	0.65	1.26	1.26	9.02	14.48	0.78	0.78	4.56	4.54
	N=4	1.30	8.43	0.51	0.51	1.23	1.22	2.36	8.01	0.20	0.20	1.24	1.26
	$N = 8$	0.20	5.41	0.40	0.40	1.00	1.00	1.21	3.23	0	0	0	0.01

TABLE 3. BER (%) IN THE LOW-PASS FILTERING ATTACK.

Audio Type		long song		symphony		dance		folklore	
Cut-off frequency		3 kHz	6 kHz	3 kHz	6 kHz	3 kHz	6 kHz	3 kHz	6 kHz
Wu et al. [2]		40.64	25.13	23.24	8.52	36.24	21.22	30.30	15.75
Xiang et al. [16]		20.62	6.28	7.81	0.48	21.08	6.38	10.54	1.34
Chen et al. [17]		18.90	3.36	28.23	13.12	11.45	0.41	15.45	1.68
Chen et al. [18]		38.34	23.34	23.11	8.24	34.92	19.92	28.56	13.56
Proposed Method	N=2	38.34	23.34	23.11	8.16	34.92	19.92	28.56	13.13
	N=4	33.40	18.40	6.38	0.00	33.02	17.38	20.23	5.62
	$N = 8$	24.60	9.60	2.24	0.00	27.56	12.27	11.80	4.78

TABLE 4. BER (%) IN THE AMPLITUDE MODIFICATION ATTACK.

Audio Type		long song				symphony				dance				folklore			
Amplitude modification factor		0.5	0.8	1.1	1.2	0.5	0.8	1.1	1.2	0.5	0.8	1.1	1.2	0.5	0.8	1.1	1.2
Wu et al. [2]		51.62	43.02	31.32	40.54	51.48	17.93	8.63	13.76	47.26	41.61	41.06	42.86	51.35	36.97	22.52	31.21
Xiang et al. [16]		2.47	2.03	1.98	1.99	2.06	2.01	1.84	1.88	4.65	4.46	4.12	4.32	1.68	1.44	1.22	1.56
Chen et al. [17]		0.11	0.10	0.10	0.10	0.46	0.36	0.24	0.38	0.15	0.12	0.12	0.12	0.08	0.07	0.06	0.07
Chen et al. [18]		47.25	45.55	41.40	43.85	48.00	38.72	23.63	24.54	43.12	41.40	40.15	40.84	45.90	43.52	42.54	42.86
Proposed Method	N=2	1.98	1.12	1.08	1.12	1.67	0.96	1.41	1.44	2.83	1.75	1.83	2.01	1.64	1.28	0.92	1.31
	N=4	0.95	0.85	0.83	0.91	1.12	0.86	0.90	0.96	2.01	1.58	0.98	1.94	1.03	0.82	0.82	0.86
	$N = 8$	0.92	0.82	0.82	0.89	1.10	0.85	0.89	0.96	1.98	1.57	0.92	1.94	1.02	0.81	0.81	0.85

Table 5. BER (%) IN THE TIME-SCALING ATTACK.

Audio Type		long song				symphony				dance				folklore			
Time-scaling (%)		-5	-2	2	5	-5	-2	2	5	-5	-2	2	5	-5	-2	2	5
Wu et al. [2]		50.14	48.92	50.23	50.84	51.81	49.23	51.13	52.36	51.62	49.61	51.75	52.86	51.65	49.67	51.72	52.81
Xiang et al. [16]		47.24	43.21	41.06	44.56	42.63	42.14	45.87	45.43	43.76	43.22	43.58	43.23	43.22	41.74	41.23	42.16
Chen et al. [17]		48.53	43.46	45.86	48.25	42.48	40.36	40.29	41.37	45.92	43.67	44.91	47.32	48.68	44.92	48.74	48.72
Chen et al. [18]		47.11	42.89	43.87	46.32	42.74	37.82	46.42	46.19	45.18	40.21	46.58	47.98	43.18	39.12	46.35	47.43
Proposed Method	N=2	47.22	42.02	43.63	45.12	42.33	37.68	45.21	46.18	45.34	40.12	46.54	46.46	43.12	38.92	46.45	47.38
	N=4	46.45	40.05	44.83	46.58	43.03	37.82	46.29	46.26	44.11	39.58	44.92	47.98	42.23	38.22	46.43	47.36
	$N = 8$	46.45	36.78	44.42	44.35	41.42	37.40	46.29	46.14	42.25	39.14	44.36	47.24	41.25	38.21	46.38	46.61

Table 6. BER (%) IN THE TIME-SCALING ATTACK.

Audio Type		long song				symphony				dance				folklore			
Bit rate (kbps)		128	112	96	80	128	112	96	80	128	112	96	80	128	112	96	80
Wu et al. [2]		0.84	3.14	4.26	7.78	0.41	2.34	3.27	6.13	0.91	3.21	4.46	7.52	0.38	2.12	3.08	6.06
Xiang et al. [16]		0.68	2.56	2.78	2.98	0.12	0.16	2.23	3.88	0.64	2.36	2.48	2.72	0.26	1.96	2.04	2.08
Chen et al. [17]		1.22	2.98	3.42	5.83	0.37	2.37	3.32	5.79	1.21	3.42	3.68	6.36	0.44	2.37	3.54	5.12
Chen et al. [18]		0.75	2.62	2.90	3.24	0.15	0.15	2.40	3.93	0.82	2.43	2.68	3.11	0.42	2.05	2.82	3.41
Proposed Method	N=2	0.74	2.62	2.88	3.24	0.15	0.14	2.27	3.88	0.81	2.42	2.57	2.68	0.43	2.02	2.74	3.28
	N=4	0.66	2.14	2.23	2.22	0.12	0.11	1.92	2.08	0.76	2.12	2.44	2.46	0.36	1.88	1.90	1.92
	$N = 8$	0.47	1.86	1.88	1.89	0.09	0.10	1.68	1.69	0.35	1.64	1.62	1.63	0.31	1.63	1.60	1.61