EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Incorporation of perceptually adaptive QIM with singular value decomposition for blind audio watermarking

Hwai-Tsu Hu[1*], Hsien-Hsin Chou[1], Chu Yu[1] and Ling-Yuan Hsu[2]

Abstract

This paper presents a novel approach for blind audio watermarking. The proposed scheme utilizes the flexibility of discrete wavelet packet transformation (DWPT) to approximate the critical bands and adaptively determines suitable embedding strengths for carrying out quantization index modulation (QIM). The singular value decomposition (SVD) is employed to analyze the matrix formed by the DWPT coefficients and embed watermark bits by manipulating singular values subject to perceptual criteria. To achieve even better performance, two auxiliary enhancement measures are attached to the developed scheme. Performance evaluation and comparison are demonstrated with the presence of common digital signal processing attacks. Experimental results confirm that the combination of the DWPT, SVD, and adaptive QIM achieves imperceptible data hiding with satisfying robustness and payload capacity. Moreover, the inclusion of self-synchronization capability allows the developed watermarking system to withstand time-shifting and cropping attacks.

Keywords: Singular value decomposition; Discrete wavelet packet transform; Adaptive quantization index modulation; Auditory masking threshold; Frame synchronization

1 Introduction

In recent years, copyright protection of multimedia data has been of great concern to content owners and service providers. Digital watermarking technology received much attention for resolving such a concern because this technology could hide information into the multimedia object (e.g., images, audio, and video) for applications like intellectual property protection, content authentication, and fingerprinting.

An audio watermarking scheme generally takes into consideration four aspects, namely, imperceptibility, security, robustness, and capacity. The developed schemes shall ensure the security and inaudibility of the embedded information, but still possess the ability of withstanding malicious attacks. The payload capacity must be large enough to accommodate necessary information. Different methods were attempted on various domains, such as time [1-5], Fourier transform [6-8], cepstral transform [9-13], discrete cosine transform (DCT) [14-17], and discrete wavelet transform (DWT) [14,16,18-23].

Compared with transform domain methods, the time-domain approach is rather easier to implement and requires less computation. The watermark is usually a pseudo noise added to the host signal. Alternatively, binary information can be converted to a noise-like signal through the spread spectrum technique. The existence of the watermark can be verified by measuring the correlation function between the pseudo noise and watermarked signal. The time-domain methods are usually less robust to digital signal processing attacks unless a long segment along with adequate embedding strength is adopted. In contrast, quantization index modulation (QIM) has been proven to be a promising technique [24]. The time-domain data embedding is achieved by quantizing the parameters derived from the time series. Though the QIM generally outperforms the spread spectrum in the time domain, it still needs a long segment for reliable detection. As a consequence, the time-domain QIM was mainly used for frame synchronization in many watermarking systems [14,20,21,24]. Being aware of the limitation of the time-

* Correspondence: hthu@niu.edu.tw
[1]Department of Electronic Engineering, National I-Lan University, Yi-Lan 26041, Taiwan
Full list of author information is available at the end of the article

domain approach, many researchers thus turned to the transform domains where signal characteristics could be better explored. The embedding intensity as well as position of the watermark can be selected based upon the features extracted in the transform domains [1,14,21].

Singular value decomposition (SVD) is a powerful tool for image processing applications [25,26]. Because the SVD can adapt to various transform domains, it has been extensively applied in audio watermarking [5,8,17,22,27]. For instance, Abd El-Samie [5] utilized a twofold strategy to embed the watermark. After applying the first SVD to a 2-D matrix formed by the audio signal, he blended the intended watermark with the diagonal matrix holding singular values and then performed the second SVD on the modified matrix. In his design, the matrices containing left- and right-singular vectors must be conserved in order to extract the watermark. Al-Nuaimy et al. [27] further extended the twofold strategy and applied it to the audio signals transmitted over network systems on a segment-by-segment basis.

Bhat et al. [22] presented a SVD-based blind watermarking scheme operated in the DWT domain. The watermark bits were embedded into the audio signals using QIM, of which the quantization steps were adaptively determined according to the statistical properties of the involved DWT coefficients. The authors claimed that their scheme was the first adaptive audio watermarking scheme exploring both DWT and SVD and had a high payload and superior performance against MP3 compression. Lei et al. [17] also attempted to embed a binary watermark into the high-frequency band of the SVD-DCT block. They attained a performance generally better than the previous SVD-based methods. Most recently, Lei et al. [28] integrated lifting wavelet transform (LWT), SVD, and QIM to achieve a very good tradeoff among the robustness, imperceptibility, and payload. Apart from the abovementioned methods, there are other audio watermarking schemes applicable to different domains in the literature [29,30].

Audio watermarks are supposed to be transparent to human ears, by what means the modification due to watermarking is virtually inaudible. One way to enhance the embedding efficiency is to exploit the auditory characteristics so that the embedding strength is sufficiently high to withstand attacks without introducing audible distortion. The methods presented in [16,17,22] demonstrated the benefit of exploiting the signal characteristics, but they relied on heuristic rules to decide the embedding strength. In these methods, even though some attention was paid to adjust relevant parameters to reach optimal performance, the connection between multiple transform domains and human auditory properties has not been thoroughly addressed.

Because the DWPT possesses multi-resolution capacity and is more computationally efficient than the Fourier transform, it may cooperate with the psychoacoustic model to render an estimate of auditory masking thresholds [31,32]. Hence, our aim in this study is to explore all useful properties of the DWPT, SVD, and QIM for audio watermarking such that the issues of robustness, imperceptibility, and payload capacity can be resolved altogether. In particular, the primary interest is placed on the blind watermarking, which does not require the original audio signal to extract the watermark.

2 Derivation of auditory masking threshold in the DWPT domain

Auditory masking is the effect when a sound is inaudible due to the presence of a louder sound. There are two types of auditory masking. One is spectral masking (sometimes referred to as simultaneously masking), which is the characteristic of the human auditory system when a sound signal is masked by a masker with a different frequency. The other is temporal masking (or non-simultaneous masking), which is the masking effect occurring before and after a sudden stimulus sound.

While studying spectral masking, critical bands are of great importance because they can be employed to elucidate the properties of frequency selectivity [32,33]. Based upon the theory of perceptual entropy [31-35], this study derives the auditory masking threshold in terms of signal power for each critical band. The derivation begins with the utilization of the DWPT to approximate the critical bands. The procedures for deriving spectral masking thresholds are briefly summarized as follows:

1. Segment the host audio signal into frames, each of 4,096 samples in length.
2. Decompose the audio signal using the DWPT according to the specification given in Table 1, in which each packet node approximately corresponds to a critical band. The decomposition is carried out using the Daubechies-8 wavelet. Let $c_i^{(n)}$ denote the ith DWPT coefficient in the nth band with a length of $N^{(n)}$.
3. Compute the short-term spectrum $X_i^{(n)}$ in each band by applying the fast Fourier transform (FFT) to $c_i^{(n)}$, i.e., $X_i^{(n)} = \mathrm{FFT}\{c_i^{(n)}\}$.
4. Estimate the tonality factor τ to see whether the band is noise-like or tone-like.

$$\tau = \min\left\{\frac{10\log_{10}\left(PM_{\mathrm{g}}\left(\left|X_i^{(n)}\right|^2\right)/PM_{\mathrm{a}}\left(\left|X_i^{(n)}\right|^2\right)\right)}{-25},\ 1\right\}, \tag{1}$$

where $PM_{\mathrm{g}}(|X_i^{(n)}|^2)$ and $PM_{\mathrm{a}}(|X_i^{(n)}|^2)$ stand for the geometric and arithmetic means of $|X_i^{(n)}|^2$, respectively.

5. Adjust the masking level according to the tonality factor.

Table 1 The arrangement of DWPT decomposition

Band number	DWPT {depth, index}	Approximate boundary (Hz)
1	{8,0}	86
2	{8,1}	172
3	{8,3}	258
4	{8,2}	345
5	{8,7}	431
6	{8,6}	517
7	{7,2}	689
8	{7,7}	861
9	{7,6}	1,034
10	{7,4}	1,206
11	{7,5}	1,378
12	{6,7}	1,723
13	{6,6}	2,067
14	{6,4}	2,412
15	{6,5}	2,756
16	{5,7}	3,445
17	{5,6}	4,134
18	{5,4}	4,823
19	{5,5}	5,513
20	{4,7}	6,891
21	{4,6}	8,269
22	{3,7}	11,025
23	{3,6}	13,781
24	{3,2}	16,538
25	{3,4}	19,294
26	{3,5}	22,050

$$D_z(n) = \left(\frac{1}{N^{(n)}} \sum_{i=0}^{N^{(n)}-1} \left(c_i^{(n)} \right)^2 \right) 10^{\frac{a(n)}{10}}, \tag{2}$$

where $a(n)$ signifies the permissible noise floor relative to the signal in the nth band, and it is formulated as

$$a(n) = \tau(-0.275n - 15.025) + (1-\tau) \times (-9.0) \quad \text{(expressed in dB)}. \tag{3}$$

6. Extend the masking effect to the adjacent bands by convolving the adjusted masking level with a spreading function $SF(n)$, namely $C_z(n) = D_z(n) \otimes 10^{\mathrm{SF}(n)/10}$, with $\mathrm{SF}(n)$ defined as

$$\mathrm{SF}(n) = p + \frac{u+v}{2}(n+y) - \frac{v-u}{2}\sqrt{h + (n+y)^2} \quad \text{(expressed in dB)}, \tag{4}$$

where $p = 15.242$, $y = 0.15$, $h = 0.3$, $u = -25$, and $v = 30$.

7. Compare the masking threshold $C_z(n)$ with the absolute threshold of hearing in quiet state, termed $T(n)$ in decibel. The maximum of the two is selected as the masking threshold, i.e.,

$$\eta(n) = \max\left\{ C_z(n),\ 10^{\frac{\mathrm{T}(n)}{10}} \right\}. \tag{5}$$

The masking threshold obtained through the above procedure is designated as $\eta(n)$, which represents the noise power level not detectable by human ears in the nth band.

3 Frame synchronization

One of the weaknesses of the existing watermarking methods consists in the vulnerability to time shifting and cropping [14]. The frame synchronization is perhaps the most prevailing counterstrategy to deal with such an issue. Many watermarking systems considered dividing the audio signal into two sorts of segments, namely, one for synchronization and the other for watermarking. This study resorts to the idea of frequency division which uses non-overlapping frequency bands to hide the synchronous codes and information bits separately. Figure 1 illustrates the idea of frequency division, where the synchronous code is placed in the frequencies below 172 Hz and the information bits are allowed to hide in the critical bands above 172 Hz.

To synchronize the frames, this study utilizes a time-domain QIM that was developed in [36] but is modified to suit the requirements here. The audio signal is deliberately partitioned into frames of length $L_\mathrm{f} = 8192$ (twice the amount for mask threshold derivation), and each frame is further divided into $N_\mathrm{s} = 32$ Subsections. A 32-bit Barker code '1111101110100111-0100101001001000' [37] is employed for the synchronization task because this code has low correlation with a time-shifted version of itself. Each binary bit is first converted into bipolar form, termed $S_\mathrm{b}(k) \in \{-1, 1\}$, and then embedded into a subsection spreading over $L_\mathrm{s}(\triangleq L_\mathrm{f}/N_\mathrm{s} = 256)$ samples by

$$\hat{m} = \begin{cases} \lfloor m/D \rfloor D + D/4 & \text{if } S_\mathrm{b}(k) = -1 \\ \lfloor m/D \rfloor D + 3D/4 & \text{if } S_\mathrm{b}(k) = 1 \end{cases} \text{ for } k = 0, 1, \cdots, L_\mathrm{s}-1, \tag{6}$$

where m and $\hat{m}$ denote, respectively, the original and modified mean values of the Subsection. D is the quantization step supposedly yielding no perceptible distortion.

To achieve the goal of imperceptivity, the quantization step at sample i, designated as D_i, is obtained by referring

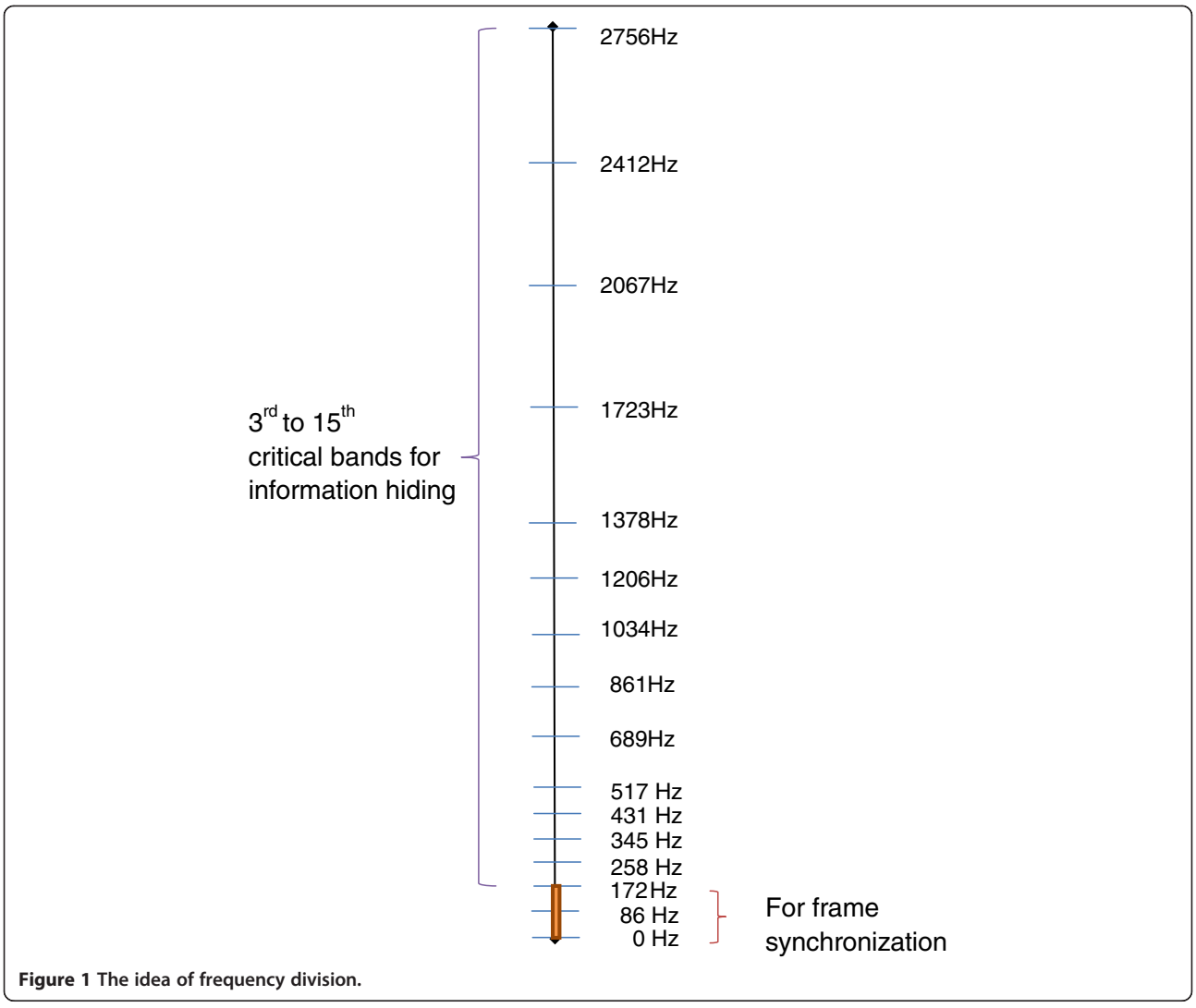

Figure 1 The idea of frequency division.

to the root-mean-square of N_p past lowpass-filtered samples:

$$D_i = \left(\frac{1}{N_p} \sum_{n=1}^{N_p} x_{lp}^2(i-n) \right)^{\frac{1}{2}} \times 10^{-10/20}, \quad (7)$$

where $x_{lp}(i)$ is the output of feeding the audio signal through a fourth order Butterworth lowpass filter with the cutoff frequency set at 172 Hz. N_p is chosen as 1,536. The scaling factor $10^{-10/20}$ aims at attenuating the signal power by 10 dB. The purpose of using $x_{lp}(i)$ is twofold. First, it provides an estimate of the signal power for frequency components below 172 Hz. Second, it excludes the disturbance from high-frequency bands where the information bits are located.

Following the derivation of a new mean, the proposed time-domain QIM modifies the audio samples in each subsection using

$$\hat{x}(k) = x(k) + (\hat{m}-m)M(k) \text{ for } k = 0, 1, \cdots, L_s-1, \quad (8)$$

where $M(k)$ is a function designed to have a flat top in the middle but descend to zero on both ends, i.e.,

$$M(k) = v \times \begin{cases} 0.5-0.5\cos(2\pi k/63), & k = 0, 1, ..., 31; \\ 1, & k = 32, ..., L_s-33; \\ 0.5-0.5\cos(2\pi(k-192)/63), & k = L_s-32, ..., L_s-1. \end{cases} \quad (9)$$

The variable v in Equation (9) is a scaling factor used to attain a mean of unity for $M(k)$, i.e., $\frac{1}{L_s}\sum_{k=0}^{L_s-1} M(k) = 1$. Based on the analysis given in [21], the QIM via Equation (8) introduces a noise with a power level of $7D_i^2/48$, which is 8.36 dB lower than D_i^2. The window $M(k)$ contributes about −0.46 dB to the signal-to-noise ratio (SNR). Combining with the 10 dB given in Equation (7),

the overall SNR resulting from the watermarking is around 17.9 dB. According to the theory of auditory entropy [31,34], the masking threshold for the frequency components below 172 Hz is approximately −16 dB below the signal power regardless of signal tonality. Consequently, the purposely reserved 17.9-dB SNR is sufficient to ensure the imperceptibility of the embedded synchronous code.

The detection of the synchronization code requires the preparation of a bit sequence $\tilde{b}(i)$, which is of the same length as the watermarked audio signal and can be derived as

$$\tilde{b}(i) = 2\left[\left(\tilde{m}_i - \lfloor \tilde{m}_i/\tilde{D}_i \rfloor \tilde{D}_i\right) \overset{?}{>} 0.5\tilde{D}_i\right] - 1, \tag{10}$$

where $\tilde{m}_i$ denotes the mean computed over a subsection starting from the ith sample. $\tilde{D}_i$ corresponds to the −10-dB RMS of previous N lowpass-filtered samples. After acquiring $\tilde{b}(i)$, the existence of a synchronous code can be identified by examining the cross-correlation between the Barker code $S_b(k)$ and a decimated version of $\tilde{b}(i)$:

$$r(i) = \sum_{k=0}^{N_s-1} S_b(N_s-1-k)\tilde{b}(i-kL_s). \tag{11}$$

As Equation (11) places the synchronous code in a backward direction, the largest $r(i)$ over an interval of 8,192 samples indicates a salient demarcation between the frames. This synchronization marker can be more prominent by adding up two other cross-correlation functions that are 8,192 samples away from the current one.

$$\hat{r}_3(i) = \sum_{j=-1}^{1} r(i + 8192j). \tag{12}$$

The position of the marker, termed I, is identified simply by picking the largest peak of $\hat{r}_3(i)$ in each interval:

$$I = \arg\max_i \{\hat{r}_3(i) | i_{\text{start}} \le i < i_{\text{start}} + 8192\}, \tag{13}$$

where i_{start} denotes the starting index.

4 Watermarking via SVD

An advantage of the SVD-based watermarking is that large singular values change very little for most types of attacks. The proposed watermarking scheme thus takes such an advantage by applying the QIM to the gap between two principal singular values. For each packet node of the DWPT, the N coefficients c_i's in a frame are organized as a $2 \times N/2$ matrix $\mathbf{M}$ in the following manner:

$$\mathbf{M} = \begin{bmatrix} c_1 & c_3 & \cdots & c_{N-1} \\ c_2 & c_4 & \cdots & c_N \end{bmatrix}_{2\times N/2}. \tag{14}$$

Without loss of generality, the superscript (n) previously used to signify a specific band has been removed in the expression. Taking SVD of $\mathbf{M}$ results in $\mathbf{M} = \mathbf{USV}^T$, where $\mathbf{U}$ is a 2×2 real unitary matrix, $\mathbf{S}$ is a $2 \times N$ / 2 diagonal matrix with non-negative real diagonal values λ_i's in decreasing order, and $\mathbf{V}^T$ (the transpose of V) is an N / $2 \times N$ /2 real unitary matrix. Alternatively, the matrix $\mathbf{M}$ can be written as

$$\begin{aligned} \mathbf{M} &= [\,\mathbf{u}_1 \quad \mathbf{u}_2\,] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \end{bmatrix} [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_{N/2}\,]^T \\ &= \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T, \end{aligned} \tag{15}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith columns of the matrices $\mathbf{U}$ and $\mathbf{V}$. The total energy of the N DWPT coefficients is the squared sum of all the elements in $\mathbf{M}$, i.e.,

$$E_c = \sum_{i=1}^{N} c_i^2. \tag{16}$$

The same result can be obtained using

$$E_c = \text{trace}\left(\mathbf{MM}^T\right) = {\lambda_1}^2 + {\lambda_2}^2. \tag{17}$$

It is recalled that the procedure described in Section 2 provides a masking threshold η, which is the maximum tolerable power variation unperceivable by human ears. The derived threshold can guide us devise a robust and transparent watermarking scheme. This study proposes embedding a watermark bit w_b into the matrix $\mathbf{M}$ by manipulating λ_1 and λ_2 subject to three criteria. First, the overall energy shall remain unchanged. That is

Criterion 1

$$\lambda_1'^2 + \lambda_2'^2 = {\lambda_1}^2 + {\lambda_2}^2, \tag{18}$$

where λ_1' and λ_2' denote the adjusted results of λ_1 and λ_2, respectively. Second, the gap between λ_1' and λ_2', termed $g' = \lambda_1' - \lambda_2'$, must comply with the QIM rule according to w_b:

Criterion 2

$$g' = \lambda_1' - \lambda_2' = \begin{cases} \left\lfloor \frac{\lambda_1 - \lambda_2}{\Delta} \right\rfloor \Delta + \frac{\Delta}{4}, & \text{if } w_b = 0; \\ \left\lfloor \frac{\lambda_1 - \lambda_2}{\Delta} \right\rfloor \Delta + \frac{3\Delta}{4}, & \text{if } w_b = 1, \end{cases} \tag{19}$$

where $\lfloor \cdot \rfloor$ represents the floor function. As for the third criterion, the signal power variation shall not exceed the auditory masking threshold η.

Let $\mathbf{M}'$ denote the matrix restored by substituting the modified eigenvalues into $\mathbf{S}$ such that

$$\mathbf{M}' = \begin{bmatrix} c_1' & c_3' & \cdots & c_{N/2-1}' \\ c_2' & c_4' & \cdots & c_{N/2}' \end{bmatrix} = \mathbf{U}\begin{bmatrix} \lambda_1' & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2' & 0 & \cdots & 0 \end{bmatrix}\mathbf{V}^T. \tag{20}$$

Because of the constraint imposed by Equation (19), the adjustment of these two eigenvalues thus holds the inequality

$$\left|(\lambda_1'-\lambda_2')-(\lambda_1-\lambda_2)\right| \le \frac{\Delta}{2}; \tag{21}$$

and the resulting error energy E_{error} becomes

$$\begin{aligned} E_{\text{error}} &= \sum_{i=1}^{N}(c_i'-c_i)^2 \\ &= \text{trace}\left((\mathbf{M}'-\mathbf{M})(\mathbf{M}'-\mathbf{M})^T\right) \\ &= (\lambda_1'-\lambda_1)^2 + (\lambda_2'-\lambda_2)^2. \end{aligned} \tag{22}$$

It is readily seen from Equation (21) that

$$(\lambda_1'-\lambda_1)^2 + (\lambda_2-\lambda_2')^2 \le \frac{\Delta^2}{4}. \tag{23}$$

Ideally, if the error power, i.e., E_{error}/N, falls beneath the masking threshold η, the signal alteration due to watermarking will be inaudible. Such a condition can be expressed as

Criterion 3

$$\frac{E_{\text{error}}}{N} \le \frac{\Delta^2}{4N} \le \eta. \tag{24}$$

Let $\Delta_{\max} = 2\sqrt{N\eta}$ denote the maximum step size used to quantize the gap between the two eigenvalues without causing perceivable distortion. The modifications with respect to λ_1' and λ_2' are denoted as $\lambda_1' = \lambda_1 + \delta_1$ and $\lambda_2' = \lambda_2 - \delta_2$. Then, the derivation of λ_1' and λ_2' based on the three criteria becomes very straightforward. Following the replacement of $\Delta_{\max}$ for Δ in Equation (19), an equation with variables δ_1 and δ_2 is formed:

$$\delta_1 + \delta_2 = g' - \lambda_1 + \lambda_2 = \rho. \tag{25}$$

In combination with Equation (18), δ_1 can be solved from a quadratic equation like

$$2\delta_1^2 + (2\lambda_1 + 2\lambda_2 - 2\rho)\delta_1 + \left(-2\lambda_2\rho + \rho^2\right) = 0. \tag{26}$$

The relationship among all involved variables is illustrated in Figure 2. After obtaining δ_1, δ_2 is acquirable using Equation (25). As Equation (26) usually comes up with two solutions for δ_1, this study chooses the one with a smaller magnitude. Nevertheless, Equation (26) may also render complex roots when $(g')^2 > E_c$. Hence, a preventive measure is taken to ensure the obtainment of real roots. It is noted from Equation (19) that the minimum possible value of g' is $3\Delta_{\max}/4$ for $w_b = 1$. In an extreme case where $\lambda_1' = g' = 3\Delta_{\max}/4$ and $\lambda_2' = 0$, $\Delta_{\max}$ must satisfy

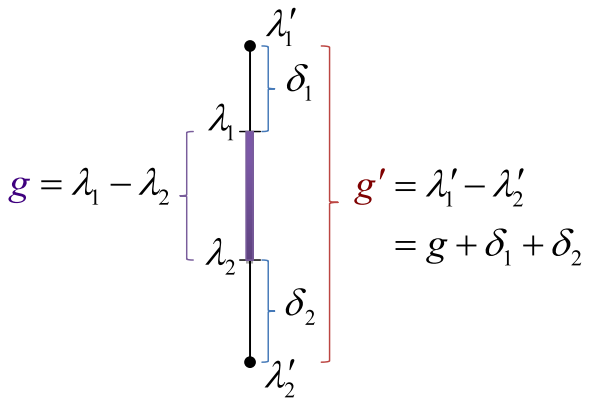

Figure 2 Relationship among all the variables involved in the SVD-based watermarking.

$$E_c = \lambda_1'^2 + \lambda_2'^2 = g'^2 = (3\Delta_{\max}/4)^2. \tag{27}$$

Consequently, the preventive measure examines the inequality whether $\Delta_{\max} < \frac{4}{3}\sqrt{E_c}$ and substitutes $\Delta_{\max}$ with $\frac{4}{3}\sqrt{E_c}$ if the inequality does not hold. This substitution, in turn, guarantees an outcome of non-negative λ_1' and λ_2'.

With the fulfillment of the three criteria, namely Equations (18), (19), and (24), the audio signal can maintain its segmental power while executing the QIM. The key factor of the entire process turns out to be η, which subsequently determines $\Delta_{\max}$, λ_1' and λ_2'. Putting the derived λ_1' and λ_2' into Equation (20) renders a modified matrix $\mathbf{M}'$ with new DWPT coefficients. Once the processes in all the involved critical bands are completed, the watermarked signal is attained by taking inverse DWPT with respect to the modified DWPT coefficients.

The watermark extraction from the watermarked signal is rather simple. Analogy to the procedures adopted for watermark embedding, the extraction process starts with taking the DWPT of the watermarked audio and then deriving the masking threshold $\tilde{\eta}$ for each packet node. Following the derivation of $\tilde{\Delta}_{\max}$ from $\tilde{\eta}$, the watermark bit $\tilde{w}_b$ can be verified by first calculating

$$\gamma = \frac{\tilde{\lambda}_1 - \tilde{\lambda}_2}{\tilde{\Delta}_{\max}} - \left\lfloor \frac{\tilde{\lambda}_1 - \tilde{\lambda}_2}{\tilde{\Delta}_{\max}} \right\rfloor. \tag{28}$$

$\tilde{w}_b$ is '1' if $\gamma \ge 0.5$, and is '0' otherwise.

5 Further enhancement

The main challenge of the adaptive QIM lies in the presupposition that the quantization steps must be accurately recovered from the watermarked signal. As seen in Section 4, the quantization step is correlated to the masking threshold, of which the formulation involves the tonality and power deduced from the signal. During the watermark embedding, the process of QIM inevitably varies the tonality and therefore causes difficulties in retrieving the quantization steps for watermark extraction. A simple way to overcome this problem is to take advantage of SVD.

It is recalled from Equation (15) that the SVD decomposes the signal into two parts, namely, $\lambda_1 \mathbf{u}_1 \mathbf{v}_1^T$ and $\lambda_2 \mathbf{u}_2 \mathbf{v}_2^T$. These two parts become $\lambda_1' \mathbf{u}_1 \mathbf{v}_1^T$ and $\lambda_2' \mathbf{u}_2 \mathbf{v}_2^T$, respectively, after applying QIM. As λ_1' is always larger than λ_2', $\lambda_1' \mathbf{u}_1 \mathbf{v}_1^T$ can be regarded as the predominant part of the watermarked signal. If the tonality is merely derived from the predominant part, i.e., $\lambda_1 \mathbf{u}_1 \mathbf{v}_1^T$ in the original signal and $\lambda_1' \mathbf{u}_1 \mathbf{v}_1^T$ in the watermarked signal, the results remain identical because the two scalars, λ_1 and λ_1', do not affect the tonality. Hence, our first enhancement to the proposed DWPT-SVD scheme is to use $\mathbf{u}_1 \mathbf{v}_1^T$ to compute for the tonality.

Another important factor in the derivation of the masking threshold is the signal power. Despite that the signal power has been deliberately maintained during watermark embedding, the attacks such as MP3 compression and noise contamination may alter the segmental power. To alleviate the problem of power alteration, our second enhancement adopts a lowpass 2-D filter to smoothen the quantization steps distributing over a plane formed by critical band numbers and frame indices. Figure 3 illustrates the idea of filter smoothing. The filter coefficients are obtained from a rotationally symmetric Gaussian function with the variance being 0.5. The filter size is tentatively chosen as 3×3 since it offers satisfactory results. It is particularly noted in the end that the quantization steps computed at the embedding stage shall also be processed by the filter when the second enhancement takes effect. The reason for this arrangement is to ensure an exact restoration of the quantization steps from the watermarked signal.

6 Integration of the entire watermarking system

Figure 4 presents the configuration of the developed watermarking system. The watermark can be an arbitrary

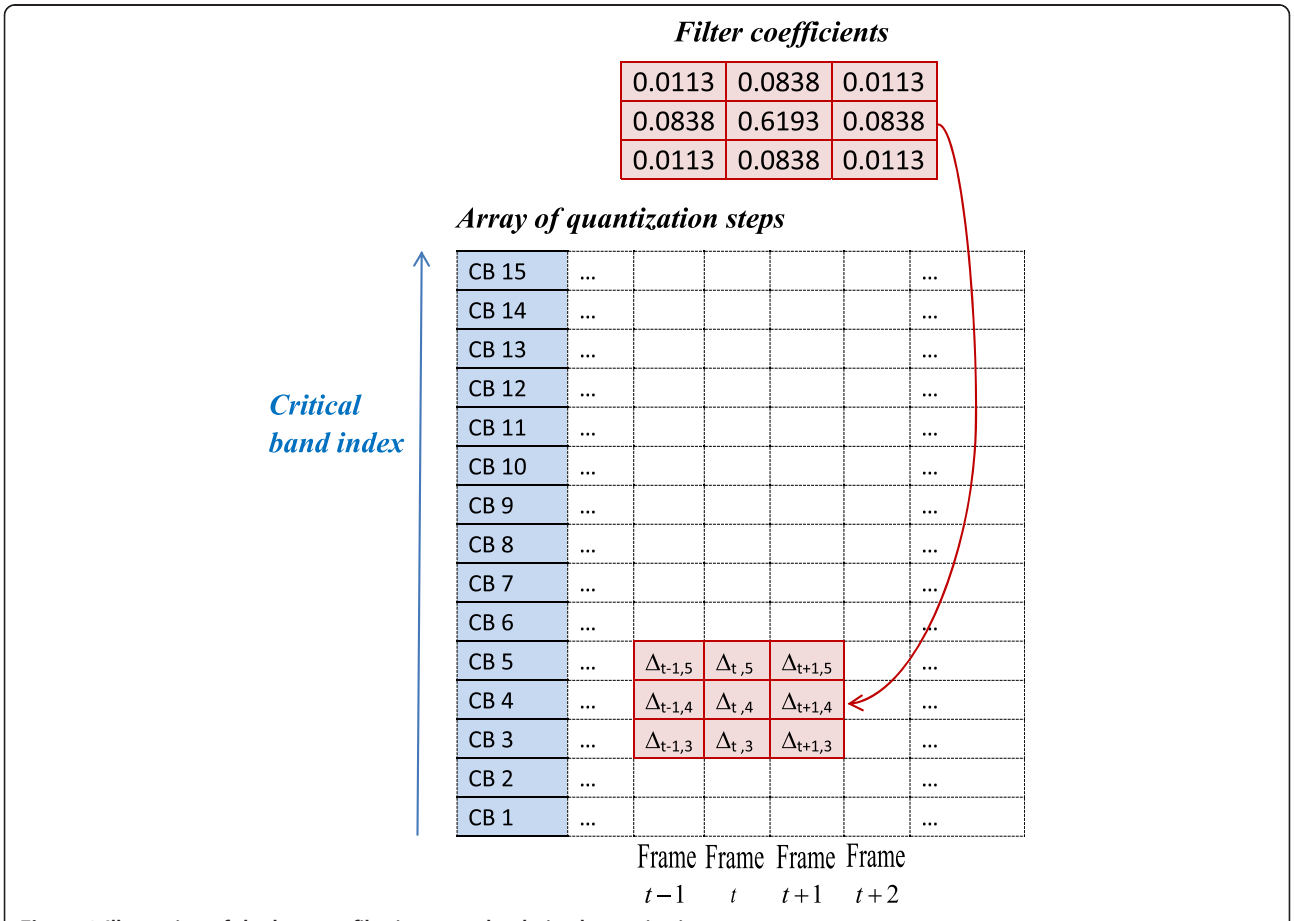

Figure 3 Illustration of the lowpass filtering over the derived quantization steps.

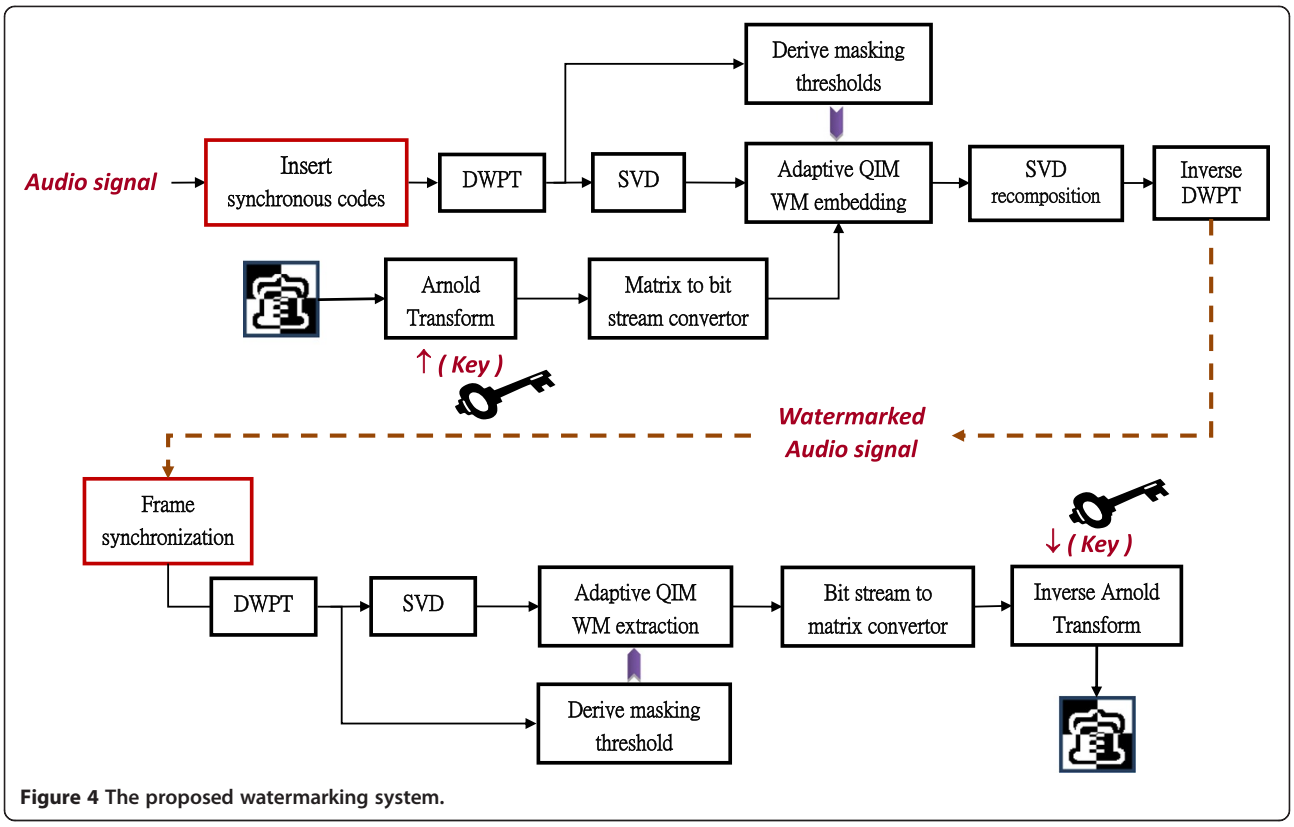

Figure 4 The proposed watermarking system.

binary bit sequence. Just for the purpose of illustration, we adopt a binary image $W(i, j)$ of size 32×32, which contains an equal amount of 0's and 1's. The procedures for embedding the watermark are as follows:

1. Maintain security by scrambling the image watermark using the Arnold transform [38].
2. Convert the scrambled image into a bit stream.
3. Partition the audio signal into frames of size 4,096 samples.
4. Insert the synchronization codes into the audio signal using the time-domain adaptive QIM presented in Section 3.
5. For the third to the fifteenth critical bands in each frame
   a. Compute the DWPT coefficients.
   b. Apply SVD to the matrix formed by the DWPT coefficients.
   c. Derive the quantization step.
   d. Embed one binary bit by quantizing the gap between two principal singular values of SVD.
   e. Recompose the DWPT coefficients.
6. Perform inverse DWPT to obtain the watermarked audio signal.

The watermark extraction is a reverse process. The procedural steps are the following:

1. Align the frame by tracing the synchronous markers.
2. For the third to the fifteenth critical bands in each frame
   a. Compute the DWPT coefficients.
   b. Apply SVD on the matrix formed by the DWPT coefficients.
   c. Derive the quantization step.
   d. Quantize the gap between two singular values.
   e. Translate the quantized value into a binary bit.
3. Gather the bits from all frames.
4. Convert the bit sequence into an image matrix.
5. Take the inverse Arnold transform to restore the watermark image, termed $\tilde{W}(i, j)$.

7 Performance evaluation

The test subjects comprised ten pieces of 30-s music recordings clipped from randomly chosen CD albums, including vocal arrangements and ensembles of musical instruments. All audio signals were sampled at 44.1 kHz with 16-bit resolution. The performance evaluation comprises three aspects: payload capacity, quality assessment, and robustness test.

To understand the influences of the two enhancements mentioned in the previous section, the test of the proposed DWPT-SVD-adaptive QIM consists of three phases, namely, the proposed one solely, the one with enhancement 1, and the one with enhancements 1 and 2.

Three recently developed SVD-based methods, denominated as 'adaptive DWT-SVD' [22], 'SVD-DCT' [17], and 'LWT-SVD' [28], are employed for performance comparison as they represent other ways to exploit the SVD for audio watermarking in transform domains. The minimum and maximum quantization steps in the adaptive DWT-SVD are 0.6 and 0.9 respectively, which are the typically suggested values. The parameters α and β for controlling the embedding strength in the SVD-DCT are assigned as 0.125 and 0.1, respectively. For the LWT-SVD method, the decomposition level of the lifting wavelet transform is chosen as 4 and the quantization step size is 0.6. The other parameters used in these three methods follow original specifications [17,22,28].

7.1 Payload

The theoretical payload capacities for the methods under investigation are presented in Table 2. The LWT-SVD holds the highest number in comparison to others. The capacity of the proposed scheme is $13 \times 44{,}100/4{,}096 = 139.97$ bps, which is lower than that of the LWT-SVD. However, this quantity is already three times more than that achieved by the adaptive DWT-SVD and SVD-DCT. It is worth pointing out that the payload capacities listed in Table 2 are computed without considering the demand of synchronous codes. In general, these numbers will drop if the watermarking methods need to allocate extra segments for frame synchronization. One advantage of the proposed synchronization technique is that it only affects the spectrum centralized in the first two critical bands, thus leaving the rest critical bands available for information hiding.

7.2 Quality assessment

The quality disturbance resulting from watermark embedding is assessed using the SNR and perceptual evaluation of audio quality (PEAQ) [39,40]. The SNR is defined as

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_n s^2(n)}{\sum_n \left(s(n)-\tilde{s}(n)\right)^2}, \tag{29}$$

where $s(n)$ and $\tilde{s}(n)$ are the original and watermarked audio signals, respectively. Since the auditory quality is a fundamentally subjective concept that does not necessarily correspond to the measured SNR, this study also resorts to the PEAQ to measure the perceived quality. The PEAQ algorithm aims at simulating human perceptual properties and integrates multiple model output variables into a single metric. It renders an objective difference grade (ODG) between −4 and 0, signifying a perceptual impression from 'very annoying' and 'imperceptible'.

Table 2 also provides the measured SNRs and ODGs for all kinds of watermarked audio signals. The SVD-DCT generally renders the largest SNR value, while the proposed scheme produces the lowest. Despite that the SNRs do not show any favor for the proposed scheme, the resulting ODGs suggest that our scheme indeed achieves the best perceived quality. In fact, the average ODG is around 0 for our scheme, implying that the watermarked signal is nearly indistinguishable from the original one. The average ODGs for the adaptive DWT-SVD and SVD-DCT are slightly above 1, indicating that the distortion caused by watermarking may still be perceivable. On the other hand, the quality degradation by the LWT-SVD seems to be minor, as the corresponding average ODG is just −0.4. Nevertheless, the ODGs resulting from these three methods are not comparable with ours.

7.3 Robustness test

The robustness test consists of two categories: one is focused on frame synchronization, and the other is concerned with watermark recovery. The attack types considered in this study include the following:

A. Resampling: conducting down-sampling to 11,025 Hz and then upsampling back to 44,100 Hz.
B. Requantization: quantizing the watermarked signal to 8 bits/sample and then back to 16 bits/sample.
C. Amplitude scaling: scaling the amplitude of the watermarked audio signal by 0.85.

Table 2 Statistics of the measured SNRs and ODGs, along with the payload capacities

Watermarking schemes	Specifications	SNR	ODG	Payload (bps)
Adaptive DWT-SVD	Reference [22]	23.872 [±2.337]	−1.030 [±1.411]	45.56[a]
SVD-DCT	Reference [17]	29.679 [±1.447]	−1.053 [±1.563]	43[a]
LWT-SVD	Reference [28]	22.025 [±2.763]	−0.400 [±1.036]	170.67[a]
The proposed scheme	DWPT-SVD-adaptive QIM	20.327 [±0.375]	−0.037 [±0.182]	139.97
The proposed scheme	+Enhancement 1	20.498 [±0.402]	−0.034 [±0.180]	139.97
The proposed scheme	+Enhancements 1 and 2	20.889 [±0.331]	−0.062 [±0.215]	139.97

The data in each cell is interpreted as 'mean [±standard deviation]'. [a]These numbers will drop if the watermarking methods need to allocate extra segments for frame synchronization.

D. Noise corruption: adding zero-mean white Gaussian noise to the watermarked audio signal with SNR = 30 dB.
E. Noise corruption: adding zero-mean white Gaussian noise to the watermarked audio signal with SNR = 20 dB.
F. Lowpass filtering: applying a lowpass filter with a cutoff frequency of 8 kHz.
G. Echo addition: adding an echo signal with a delay of 50 ms and a decay of 5% to the watermarked audio signal.
H. Jittering: randomly deleting or adding one sample for every 100 samples within each frame.
I. 128-kbps MPEG compression: compressing and decompressing the watermarked audio signal with a MPEG layer III coder at a bit rate of 128 kbps.
J. 64-kbps MPEG compression: compressing and decompressing the watermarked audio signal with a MPEG layer III coder at a bit rate of 64 kbps.
K. Time shifting: shifting the watermarked audio signal by an amount of 50% relative to the frame length.

The efficiency of the proposed synchronization scheme is demonstrated via the statistical means and standard deviations of $\hat{r}_3(i)'s$ discussed in Section 3, along with the misdetection counts of the synchronization markers. As revealed from the results in Table 3, the detectability of the synchronous marks is always reliable, indicating that common attacks do not impose any threat to the watermarking system equipped with such a synchronization technique.

Table 3 Statistical results of the estimated correlation functions for the time-domain synchronization scheme

Attack type	Sync_code absent	Sync_code present	Misdetection
0 (none)	−0.01 [±10.40]	96.00 [±0.00]	0
A	−0.01 [±10.40]	95.99 [±0.08]	0
B	−0.01 [±10.40]	96.00 [±0.00]	0
C	−0.01 [±10.40]	96.00 [±0.00]	0
D	−0.01 [±10.39]	95.98 [±0.09]	0
E	−0.01 [±10.35]	94.33 [±2.86]	0
F	−0.01 [±10.40]	96.00 [±0.00]	0
G	−0.01 [±10.05]	68.39 [±10.97]	0
H	−0.01 [±10.37]	94.38 [±3.28]	0
I	−0.01 [±10.39]	95.88 [±0.33]	0
J	−0.01 [±10.22]	86.91 [±5.75]	0
K	−0.01 [±10.40]	96.00 [±0.00]	0

The results are in terms of mean [±standard deviation] and misdetection rate.

The robustness of the proposed watermarking technique in the presence of various attacks is evaluated using the bit error rate (BER), which is defined as

$$\mathrm{BER}\left(W,\tilde{W}\right)=\frac{\sum_{i=1}^{M}\sum_{j=1}^{M}W(i,j)\oplus\tilde{W}(i,j)}{M\times M},\qquad(30)$$

where $\oplus$ stands for the exclusive-or operator. Table 4 gives the BERs obtained from the watermarked audio signals under the attacks.

Generally speaking, all the SVD-based methods manifest certain robustness against most attacks. However, the adaptive DWT-SVD and LWT-SVD appear vulnerable to amplitude scaling. The reason can be ascribed to the fact that some of the controlling parameters in both methods are fixed. A minor change in amplitude may therefore result in a disastrous consequence. In contrast, the SVD-DCT and the proposed scheme do not exhibit such deficiency, as both of them are designed to adapt to amplitude variation. Besides amplitude scaling, the adaptive DWT-SVD also suffers from the attack of resampling. The reason is due to the altered statistical distribution of the DWT coefficients that eventually leads to inaccurate watermark extraction.

As shown in Table 4, the proposed scheme generally retains very high accuracy under all sorts of attacks, but it seldom reaches 100% correctness. This is because the masking threshold derived from the watermarked signal may somewhat differ from the original one. To ameliorate such drawback, two enhancements have been proposed in Section 5. The first enhancement rectifies the inconsistency in the derivation of tonality. As a consequence, the proposed scheme comes up with a perfect accuracy if no attack is present. Excellent robustness is also observed for attacks like resampling, amplitude scaling, and lowpass filtering. The second enhancement tends to mitigate the power alterations caused by the attacks. After being equipped with the second enhancement, the proposed scheme gains noticeable improvements for all kinds of attacks. More importantly, the changes in SNR and ODG are slight, meaning that the improvement is not obtained at the cost of perceived quality.

7.4 Security

There are several possible ways to promote the watermark security. In [17,28], the synchronous code was chaotically permutated and the watermark data were scrambled. A similar strategy is certainly applicable to our system. Here, the Arnold transform is chosen to shuffle the watermark image since this technique has been widely utilized in digital image encryption. Aside from data scrambling, the controlling parameters (e.g., the frame length, the arrangement of the matrix in Equation (14), and/or the selected

Table 4 Averaged bit error rates of the watermarking schemes under various attacks

Attack type	Adaptive DWT-SVD (%)	SVD-DCT (%)	LWT-SVD (%)	Proposed (%)	Proposed + enhancement 1 (%)	Proposed + enhancements 1 and 2 (%)
0 (none)	0.000	0.000	0.000	0.121	0.000	0.000
A	3.170	0.310	0.000	0.121	0.000	0.000
B	0.000	0.000	0.000	0.199	0.082	0.063
C	31.160	0.000	74.010	0.121	0.000	0.000
D	0.000	0.000	0.000	0.312	0.199	0.158
E	0.010	0.090	0.000	1.294	1.205	1.016
F	0.760	0.110	0.000	0.121	0.000	0.000
G	0.610	0.010	0.300	0.158	0.019	0.004
H	0.010	0.670	0.040	0.845	0.730	0.537
I	0.000	0.000	0.000	0.130	0.002	0.000
J	1.770	1.140	2.180	1.616	1.582	1.168
K	0.000	0.000	0.000	0.121	0.000	0.000

critical bands) can be utilized as secret keys. It would be difficult, if not impossible, to detect the watermark without knowing the exact parameters.

8 Error analysis

There are two types of errors during the search of watermarks. The false-positive error (FPE) is the probability of declaring an unwatermarked audio signal as a watermarked one, whereas the probability of the opposite condition (classifying a watermarked audio signal as an unwatermarked one) is known as the false-negative error (FNE).

Following the basic assumption and derivative rules given in [22], the FPE P_{fp} can be computed as

$$\begin{aligned} P_{\mathrm{fp}} &= P\left\{H\left(W,\tilde{W}\right)\geq T \,\middle|\, \text{without watermark}\right\} \\ &= \sum_{k=T}^{N_{\mathrm{w}}} \binom{N_{\mathrm{w}}}{k} (P_{\mathrm{e}})^{k} (1-P_{\mathrm{e}})^{N_{\mathrm{w}}-k}, \end{aligned} \tag{31}$$

where $H\left(W,\tilde{W}\right)$ denotes the number of matched bits in a total of N_{w} bits, and T is the threshold for claiming the existence of the watermark. $\binom{N_w}{k}$ stands for the binomial coefficient. P_{e} is the probability that the extracted bits match with the original watermark bits. Since the unwatermarked bits are either 0 or 1 with pure randomness, P_{e} is therefore assumed to be 0.5. As a result, Equation (31) can be further simplified as

$$P_{\mathrm{fp}} = \frac{1}{2^{N_{\mathrm{w}}}} \sum_{k=T}^{N_{\mathrm{w}}} \binom{N_{\mathrm{w}}}{k}. \tag{32}$$

If $N_{\mathrm{w}} = 1024$ and $T = \lceil 0.8 \times N_{\mathrm{w}} \rceil = 820$, then $P_{\mathrm{fp}} = 2.62 \times 10^{-88}$, which means that FPE can rarely happen.

Analogy to the discussion in the derivation of FPE, the FNE P_{fn} can be computed as

$$\begin{aligned} P_{\mathrm{fn}} &= P\left\{H\left(W,\tilde{W}\right) < T \,\middle|\, \text{with watermark}\right\} \\ &= \sum_{k=0}^{T-1} \binom{N_{\mathrm{w}}}{k} (1-\mathrm{BER})^{k} \times (\mathrm{BER})^{N_{\mathrm{w}}-k} \\ &= \sum_{k=T}^{N_{w}} \binom{N_{\mathrm{w}}}{k} (\mathrm{BER})^{k} \times (1-\mathrm{BER})^{N_{\mathrm{w}}-k}. \end{aligned} \tag{33}$$

Taking the worst case (where BER = 0.012) in our experiments as an example, the FNE of the proposed scheme is virtually zero.

9 Conclusion

This paper presents an efficient audio watermarking technique, which integrates the DWPT, SVD, and adaptive QIM subject to the auditory masking effect. While the DWPT decomposes the audio signal into critical bands, the exploration of perceptual entropy leads to the derivation of auditory masking thresholds. The thresholds, in turn, determine the quantization steps required by the QIM. In virtue of the robustness of the SVD technique, the proposed watermarking scheme first assembles the DWPT coefficients into a matrix and then manipulates the singular values to satisfy three criteria. As a result, the embedded watermark is guaranteed to restrain underneath the perceptible level. To further improve the overall performance, this study introduces two auxiliary enhancement measures to ensure the recovery of quantization steps.

Apart from the scheme for data embedding, the developed watermarking system is equipped with a competent frame synchronization technique to withstand the time-shifting attacks. The experimental results reveal that the proposed DWPT-SVD-adaptive QIM scheme performs

very well against many attacks such as resampling, requantization, amplitude scaling, lowpass filtering, jittering, echo addition, white noise contamination, and MP3 compression. The comparison with the other SVD-related watermarking methods indicates that our scheme is comparable to, if not better than, the selected methods. Most importantly, the resulting average ODGs of the proposed scheme are around 0, implying that the embedded watermarks and synchronous codes are virtually inaudible by human ears. All these merits can be attributed to the incorporation of the perceptually adaptive QIM with SVD in the DWPT domain.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work was supported by the National Science Council, Taiwan, ROC, under grants NSC101-2221-E-197-033 and NSC102-2221-E-197-020.

Author details
[1]Department of Electronic Engineering, National I-Lan University, Yi-Lan 26041, Taiwan. [2]Department of Information Management, St. Mary's Medicine, Nursing and Management College, Yi-Lan 26644, Taiwan.

Received: 3 November 2013 Accepted: 14 January 2014
Published: 28 January 2014

References
1. MD Swanson, B Zhu, AH Tewfik, L Boney, Robust audio watermarking using perceptual masking. Signal Process. **66**(3), 337–355 (1998)
2. P Bassia, I Pitas, N Nikolaidis, Robust audio watermarking in the time domain. IEEE Trans. Multimedia **3**(2), 232–241 (2001)
3. 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), 46–59 (2006)
4. AN Lemma, J Aprea, W Oomen, L van de Kerkhof, A temporal domain audio watermarking technique. IEEE Trans. Signal Processing **51**(4), 1088–1097 (2003)
5. F Abd, El-Samie. An efficient singular value decomposition algorithm for digital audio watermarking. Int. J. Speech. Technol. **12**(1), 27–45 (2009)
6. W Li, X Xue, P Lu, Localized audio watermarking technique robust against time-scale modification. IEEE Trans. Multimedia **8**(1), 60–69 (2006)
7. R Tachibana, S Shimizu, S Kobayashi, T Nakamura, An audio watermarking method using a two-dimensional pseudo-random array. Signal Process. **82**(10), 1455–1469 (2002)
8. D Megías, J Serra-Ruiz, M Fallahpour, Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification. Signal Process. **90**(12), 3078–3092 (2010)
9. X Li, HH Yu, Transparent and robust audio data hiding in cepstrum domain. ICME **1**, 397–400 (2000)
10. S Li, L Cui, J Choi, X Cui, An audio copyright protection schemes based on SMM in cepstrum domain. Lect. Notes Comput. Sc. **4109**, 923–927 (2006)
11. SC Liu, SD Lin, BCH code-based robust audio watermarking in cepstrum domain. J. Inf. Sci. Eng. **22**(3), 535–543 (2006)
12. SK Lee, Y-S Ho, Digital audio watermarking in the cepstrum domain. IEEE T. Consum. Electr. **46**(3), 744–750 (2000)
13. H-T Hu, W-H Chen, A dual cepstrum-based watermarking scheme with self-synchronization. Signal Process. **92**(4), 1109–1116 (2012)
14. X-Y Wang, H Zhao, A novel synchronization invariant audio watermarking scheme based on DWT and DCT. IEEE Trans. Signal Processing **54**(12), 4835–4840 (2006)
15. I-K Yeo, HJ Kim, Modified patchwork algorithm: a novel audio watermarking scheme. IEEE Trans. Speech and Audio Processing **11**(4), 381–386 (2003)
16. X Wang, W Qi, P Niu, A new, adaptive digital audio watermarking based on support vector regression. IEEE T. Audio Speech **15**(8), 2270–2277 (2007)
17. BY Lei, IY Soon, Z Li, Blind and robust audio watermarking scheme based on SVD–DCT. Signal Process. **91**(8), 1973–1984 (2011)
18. X He, MS Scordilis, Efficiently synchronized spread-spectrum audio watermarking with improved psychoacoustic model. Research Letter in Signal Process (2008). 10.1155/2008/251868
19. S Xiang, HJ Kim, J Huang, Audio watermarking robust against time-scale modification and MP3 compression. Signal Process. **88**(10), 2372–2387 (2008)
20. X-Y Wang, P-P Niu, H-Y Yang, A robust digital audio watermarking based on statistics characteristics. Pattern Recognition **42**(11), 3057–3064 (2009)
21. S Wu, J Huang, D Huang, YQ Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission. IEEE Trans. Broadcast. **51**(1), 69–76 (2005)
22. V Bhat, K, I Sengupta, A Das. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. Digit. Signal Process. **20**(6), 1547–1558 (2010)
23. S-T Chen, G-D Wu, H-N Huang, Wavelet-domain audio watermarking scheme using optimisation-based quantisation. IET Signal Process. **4**(6), 720–727 (2010)
24. B Chen, GW Wornell, Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inform. Theory **47**(4), 1423–1443 (2001)
25. L Ruizhen, T Tieniu, An SVD-based watermarking scheme for protecting rightful ownership. IEEE Trans. Multimedia **4**(1), 121–128 (2002)
26. P Bao, M Xiaohu, Image adaptive watermarking using wavelet domain singular value decomposition. IEEE Trans. Circuits Syst. Video Technol. **15**(1), 96–102 (2005)
27. W Al-Nuaimy, MAM El-Bendary, A Shafik, F Shawki, AE Abou-El-azm, NA El-Fishawy, SM Elhalafawy, SM Diab, BM Sallam, FE Abd, El-Samie, HB Kazemian. An SVD audio watermarking approach using chaotic encrypted images. Digit. Signal Process. **21**(6), 764–779 (2011)
28. 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 Process. **92**(9), 1985–2001 (2012)
29. R Zezula, J Misurec, Audio digital watermarking algorithm based on SVD in MCLT domain. ICONS , 140–143 (2008)
30. A Dhawan, SK Mitra, Hybrid audio watermarking with spread spectrum and singular value decomposition. INDICON , 11–16 (2008)
31. B Carnero, A Drygajlo, Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms. IEEE Trans. Signal Processing **47**(6), 1622–1635 (1999)
32. X He, MS Scordilis, An enhanced psychoacoustic model based on the discrete wavelet packet transform. J. Franklin Inst. **343**(7), 738–755 (2006)
33. T Painter, A Spanias, Perceptual coding of digital audio. Proc. IEEE **88**(4), 451–515 (2000)
34. JD Johnston, Estimation of perceptual entropy using noise masking criteria. Proc. IEEE Int. Conf. Acoust. Speech Signal Process. **2525**, 2524–2527 (1988)
35. JD Johnston, Transform coding of audio signals using perceptual noise criteria. IEEE J. Select. Areas Commun. **6**(2), 314–323 (1988)
36. H-T Hu, C Yu, A perceptually adaptive QIM scheme for efficient watermark synchronization. IEICE T. Inf. Syst. **E95-D**(12), 3097–3100 (2012)
37. SM Gentry, *Detection Optimization Using Linear Systems Analysis of a Coded Aperture Laser Sensor System: Sandia Report* (Sandia National Laboratories, Albuquerque, 1994)
38. VI Arnold, A Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968)
39. ITU Radiocommunication Sector (ITU-R), *Recommendation BS.1387: Method for Objective Measurements of Perceived Audio Quality* (International Telecommunication Union, Geneva, 1998)
40. P Kabal, *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality, TSP Lab Technical Report* (Department of Electrical and Computer Engineering, McGill University, Montréal, 2002)

doi:10.1186/1687-6180-2014-12
Cite this article as: Hu *et al.*: Incorporation of perceptually adaptive QIM with singular value decomposition for blind audio watermarking. *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:12.