Multimed Tools Appl
https://doi.org/10.1007/s11042-018-5813-z

A blind audio watermarking technique based on a parametric quantization index modulation

Younes Terchi[1] · **Saad Bouguezel**[1]

Received: 10 May 2017 / Revised: 24 December 2017 / Accepted: 19 February 2018

Abstract In this paper, we propose an efficient transform-based blind audio watermarking technique by introducing a parametric quantization index modulation (QIM). Theoretical expressions for the signal to watermark ratio and probability of error are derived and then used in an optimization technique based on the Lagrange multipliers method to find the optimal values for the parameters of the parametric QIM that ensure the imperceptibility while maximizing the robustness under an additive white Gaussian noise (AWGN) attack. Moreover, a fast scheme for the implementation of the proposed watermarking technique is developed and an efficient procedure is suggested to find the interval for the best selection of the watermark embedding positions that provide a good trade-off between the effects of high and low pass filtering attacks. The parameters of the resulting optimal parametric QIM coupled with the embedding positions constitute a highly robust secret key for the proposed watermarking technique. We also carry out several experiments to show the usefulness of the theoretical analysis presented in the paper and compare the proposed technique with other existing QIM-based watermarking techniques by considering known attacks such as AWGN, re-quantization, resampling, low/high pass filtering, amplitude scaling and common lossy compressions.

Keywords Blind audio watermarking · Method of Lagrange multipliers · Particle swarm optimization · Quantization index modulation

✉ Saad Bouguezel
bouguezel_saad@yahoo.com

Younes Terchi
terchi.younes@gmail.com

[1] Laboratoire de Croissance et Caractérisation de Nouveaux Semiconducteurs, Département d'Electronique, Faculté de Technologie, Université Ferhat Abbas Sétif-1, 19000 Sétif, Algeria

Published online: 05 March 2018

1 Introduction

The recent explosion of communication systems and Internet as collaborative mediums has opened the door for companies or people who want to share or sell their multimedia products. Nonetheless, the advantages of such open mediums can lead to very serious problems for digital media owners who do not want their products to be distributed without their consent. This is due to the ease of illegal reproduction, manipulation and distribution of digital media exchanged through communication networks or carried out on multimedia devices. The digital watermarking is an efficient solution for these and other information security problems [2, 15, 22, 26]. It provides means for embedding a message or signature (watermark) proving the ownership in an image, video or audio signal without destroying its perceptual value [8, 21, 32]. The nature of the watermark is generally application dependent. For the covert communication, the watermark is a secret message in a form of a text. For broadcast monitoring applications, the watermark is habitually a pseudo random sequence of bits. For fingerprinting applications, the watermark is a unique binary sequence associated with the host signal to distinguish it from other fingerprints in the database. For monitoring or tracing back illegally produced copies of data, the transactional watermark is typically an alpha numerical identifier for each distributor. For copy control applications, the watermark is a unique binary sequence for all devices to prevent illegal reproduction. For copyright owner identification and proof of ownership applications, the watermark is usually a logo in the form of a binary image. In this paper, without loss of generality, we use a binary image as a watermark in view of the dominance of copy right protection applications.

Audio watermarking is more challenging than image and video watermarking. This is because of the fact that the human auditory system (HAS) is more sensitive than the human visual system (HVS) [20]. Digital audio watermarking techniques can be classified into two main classes, time- and transform-domain techniques. The latter found to be more advantageous than the former in terms of robustness and imperceptibility [16, 30], however, they suffer from higher computational complexities due to the required forward and inverse transforms. The approaches that are widely adopted by the transform-domain watermarking techniques are the spread spectrum introduced by Cox et al. in [11] and the quantization index modulation (QIM) proposed by Chen and Wornell in [4–7]. The QIM-Based techniques are computationally more efficient and provide a good trade-off between robustness and imperceptibility with high data embedding capacity. The QIM has many realizations [7] and the most interesting one is the well-known dither modulation. Due to its advantages, QIM has received a wide attention from researchers in the field of watermarking. Wu et al. proposed in [29] a QIM-based watermarking technique in the discrete wavelet transform (DWT) domain and a baseline mechanism for its self-synchronization. In [9], the authors presented a QIM-based audio watermarking technique for quantizing a weighted group amplitude of the lowest DWT approximation band, where the coefficients of the approximation band are then altered in a way that transparency is optimized. Chen et al. proposed in [3] a QIM-based audio watermarking technique that quantizes the absolute amplitude of the lowest 8-level DWT frequency band, where the signal coefficients are then modified in a manner that maximizes transparency. Hu et al. proposed in [14] a QIM-based audio watermarking technique by quantizing the standard deviations of the discrete cosine transform (DCT) bands. The main idea of this technique is to exploit the HAS in order to devise an adaptive quantization step, which results in a non-uniform quantization. In [13], the authors proposed a QIM-based audio watermarking technique, which embeds the watermark bits by quantizing the low frequency band of the host signal. The main idea of this technique is to firstly identify the sample

positions, which offer a good transparency when quantized. The sample positions vary from a frame to another and thus are stored as a side information to be transmitted separately over a high-fidelity channel to the decoder. Secondly, embed the watermark in a host signal many times then perform numerous attacks on the watermarked signal followed by several watermark extractions to serve the calculations performed by the employed stochastic optimization. This is to find a quantization step providing a good robustness to the considered attacks.

The above QIM-based audio watermarking techniques have many drawbacks, namely the vulnerability to high-pass filtering-like attacks and lossy compression of the quantization of the considered global frame characteristic (e.g. the absolute amplitude in [3, 9] and the standard deviation in [14]). Moreover, the optimization in [3, 9] is performed to maximize the transparency without considering the robustness, and the 8-level DWT in [3] restricts the maximum achievable capacity, which renders the technique in [3] not suitable for applications requiring high embedding capacity. The adaptive quantization steps in [14] serve for higher capacity at the expense of robustness due to the extra error caused by recovering the quantization steps in the extraction procedure, and the sensitivity of the extracted quantization steps to additive noise. Although [13] considers both robustness and transparency, the technique is not blind as the side information is essential to perform the watermark extraction, and time consuming because of the use of stochastic optimization, which is known for its downsides compared to analytical optimization. Finally, the most important deficiency in the techniques reported in [3, 9, 13, 14] is that the used QIM realization is neither justified nor optimized for a given technique. An exhaustive literature review shows that other watermarking techniques such as those reported in [18, 19, 25, 28, 31, 33, 34] also have such a deficiency. However, it is highly desirable to find a QIM realization that is more appropriate or optimal for a given watermarking technique.

In this paper, we propose a new blind audio watermarking technique based on the transform domain. It consists of (1) segmenting the audio signal into frames, (2) transforming each frame using an orthogonal transform, (3) embedding one watermark bit per transformed frame, and (4) applying the inverse transformation on each of the resulting watermarked frame. We adopt the single coefficient quantization approach for the watermark bit embedding process. In order to find a QIM realization that is optimal for the proposed watermarking technique, we (1) propose a new parametric QIM, which reduces for some values of the parameters to the QIM realizations used in [3, 9, 13, 14, 18, 19, 25, 28, 29, 31, 33, 34], (2) derive the theoretical expressions for the signal to watermark ratio (SWR) and the probability of error, and (3) propose an optimization technique based on the Lagrange multipliers method to find the optimal values for the parameters of the proposed parametric QIM, which is used to quantize a single coefficient per frame. The principle of the proposed optimization technique is to set the SWR to a fixed value SWR_0, which is according to the international federation of photographic industry (IFPI) higher than 20 dB [1], in order to guarantee the imperceptibility while minimizing the probability of error under an additive white Gaussian noise (AWGN) attack to maximize the robustness. Although any orthogonal transform can be used in the proposed technique, the DCT is chosen for its high energy compaction capability. As mentioned earlier, the transform-domain watermarking techniques suffer from higher computational complexities. Fortunately, this is not the case with the proposed technique for which we propose a very fast scheme to pass from time domain to transform domain and vice versa. The idea behind this scheme is to appropriately exploit the fact that the proposed watermarking technique embeds only one watermark bit per frame by adopting the single coefficient quantization approach. We also propose a procedure leading to a best interval from which an embedding position can be

selected to provide a good trade-off between the effects of high and low pass filtering attacks. This would be feasible due to the adopted single coefficient quantization approach, which offers a good interpretation for the high and low pass filtering attacks. The embedding positions and parameters of the obtained optimal parametric QIM can be used as a secret key in the proposed watermarking technique. Finally, we experimentally show the validity of the theoretical analysis developed in the paper and investigate the efficiency of the proposed watermarking technique in terms of robustness and imperceptibility, and compare it with those of the techniques reported in [3, 9, 14, 29].

The rest of the paper is organized as follows. Section 2 presents the proposed watermarking technique along with the proposed parametric QIM. Theoretical expressions for the SWR and the probability of error are derived in Section 3. The optimal parametric QIM is obtained in Section 4 using an optimization method based on the Lagrange multipliers. This section proposes also a closed-form expression approximating the derived analytical expression of the probability of error using the particle swarm optimization (PSO) [10, 12, 23]. Section 5 suggests a procedure to find the best interval for selecting watermark embedding positions. An algorithm for a fast implementation of the proposed watermarking technique is proposed in Section 6. Experimental results along with comparisons and discussions are presented in Section 7. Section 8 gives some conclusions and future directions.

2 Proposed watermarking technique

In this section, we propose an efficient blind audio watermarking technique based on the transform domain. We first briefly review the standard QIM and some of its realizations and then devise a parametric QIM to be used in the proposed watermark embedding and extraction algorithms.

2.1 Existing QIM realizations

In this subsection, we present some existing QIM realizations. The standard quantization operation on the original host sample or coefficient X is defined for the quantization step Δ as

$$Q(X,\Delta) = \text{round}\left(\frac{X}{\Delta}\right)\Delta \tag{1}$$

where the function round(·) denotes rounding a value to the nearest integer. The rounding is performed according to the watermark bit w to be embedded. For example, X/Δ is rounded to the nearest even integer when $w = 0$ and to the nearest odd integer when $w = 1$.

In the case of the QIM introduced in [7], the watermarked signal $\hat{X}$ is obtained by a quantizer that depends on w as

$$\hat{X} = Q(X,\Delta,w) \tag{2}$$

The QIM can have various realizations. For instance, the QIM used in [3, 9, 29] has the form

$$\hat{X} = \text{round}\left(\frac{X}{\Delta}\right)\Delta + w\frac{\Delta}{2} + \frac{\Delta}{4} \tag{3}$$

whereas that considered in [14, 31, 34] has the form

$$\hat{X} = \text{round}\left(\frac{X - w\frac{\Delta}{2}}{\Delta}\right)\Delta + w\frac{\Delta}{2} \tag{4}$$

The realisation of the QIM is adopted in [13, 18] as

$$\hat{X} = \text{round}\left(\frac{X - \frac{\Delta}{2} - w\frac{\Delta}{2}}{\Delta}\right)\Delta + w\frac{\Delta}{2} + \frac{\Delta}{2} \tag{5}$$

In [25], the realisation is

$$\hat{X} = \text{round}\left(\frac{X - \frac{\Delta}{2}}{\Delta}\right)\Delta + w\frac{\Delta}{2} \tag{6}$$

The form used in [33] is

$$\hat{X} = \text{round}\left(\frac{X - \frac{\Delta}{2}}{\Delta}\right)\Delta + w\frac{\Delta}{2} + \frac{\Delta}{2} \tag{7}$$

Another interesting realization of the QIM is the well-known dither-modulation introduced by Chen and Wornell in [7] and considered in [19], which can be formulated in the case of uniform scalar quantization as

$$\hat{X} = Q(X, \Delta, w) = \text{round}\left(\frac{X + d(w)}{\Delta}\right)\Delta - d(w) \tag{8}$$

with

$$d(1) = \begin{cases} d(0) + \frac{\Delta}{2}, & d(0) < 0 \\ d(0) - \frac{\Delta}{2}, & d(0) \geq 0 \end{cases} \tag{9}$$

where $d(0)$ can be chosen pseudorandomly with a uniform distribution over the interval $[-\Delta/2, \Delta/2]$.

2.2 Proposed parametric QIM

The diversity of the QIM realizations found in the literature suggests a sort of unification, which can be achieved by introducing a parametrization approach. Moreover, this approach can be exploited to find the optimal QIM realization for a given watermarking technique. Therefore, we define and propose a new parametric realization of the QIM as

$$\hat{X} = Q_{\alpha,\beta,\gamma}(X, \Delta, w) = \text{round}\left(\frac{X + \beta + \gamma w}{\Delta}\right)\Delta + w\frac{\Delta}{2} + \alpha \tag{10}$$

where α, β, and γ are parameters to be defined and w is the watermark bit to be embedded. The proposed watermark extraction process is given by

$$\hat{w} = Q_{\alpha}^{-1}\left(\hat{X}, \Delta\right) = \begin{cases} 1 & \text{if } \frac{\Delta}{4} \leq \text{mod}\left(\hat{X} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $\hat{w}$ is the extracted watermark bit.

It is clear that, for a suitable choice of the parameters α, β, and γ, the proposed parametric QIM defined in (10) reduces to any of the QIMs given by (3)–(7). For instance, for $\beta = \gamma = 0$ and $\alpha = \frac{\Delta}{4}$, the parametric QIM reduces to the QIM given by (3), which is used in [3, 9, 29], and for $\alpha = \beta = 0, \gamma = -\frac{\Delta}{2}$, reduces to the QIM given by (4), which is used in [14, 28, 31, 34].

2.3 Proposed watermark embedding

The objective of the proposed watermark embedding is to embed a binary image **I** in a digital audio signal as shown in Fig. 1. Hence, the image is firstly mapped into a vector **W** whose elements are the watermark bits w_m. The original digital audio signal is segmented into non-overlapping frames x^m each of length L, and an orthogonal linear transformation T is applied on each frame to obtain $X^{(m)}$ as

$$X_r^{(m)} = \sum_{l=0}^{L-1} x_r^{(m)} \phi_l^*(r), r = 0, 1, \ldots, L-1 \tag{12}$$

where $\phi_l(r)$, $0 \leq l$, $r \leq L-1$, is a set of linearly independent orthogonal basis constituting the kernel of the transform T and $(\cdot)^*$ denotes the complex conjugate operation.

We embed one bit per frame in the transform domain. Hence, the watermark bit w_m indexed by m is embedded according to the proposed parametric QIM given by (10) in a chosen coefficient $X_k^{(m)}$ of the transformed frame $X^{(m)}$ indexed by m. Thus, the watermarked coefficient $\hat{X}_k^{(m)}$ is obtained as

$$\hat{X}_k^{(m)} = Q_{\alpha,\beta,\gamma}\left(X_k^{(m)}, \Delta, w_m\right) = \text{round}\left(\frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta}\right)\Delta + w_m \frac{\Delta}{2} + \alpha \tag{13}$$

where α, $\beta + \gamma$ and Δ to be defined in a way to obtain the optimal robustness while ensuring imperceptibility (or inaudibility) in the time domain. Then, the inverse transformation T^{-1} is applied on each watermarked frame $\hat{X}^{(m)}$ in the transform domain to obtain the watermarked frame $\hat{x}^{(m)}$ in the time domain as

$$\hat{x}_l^{(m)} = \sum_{r=0}^{L-1} \hat{X}_r^{(m)} \phi_l(r), l = 0, 1, \ldots, L-1 \tag{14}$$

The resulting watermarked frames are all joined to construct the watermarked digital audio signal.

2.4 Proposed watermark extraction

The proposed watermark extraction process is illustrated by Fig. 2. The watermarked digital audio signal is firstly segmented into non-overlapping frames $\hat{x}^{(m)}$ each of length L, and the same orthogonal linear transformation T given by (12) used in the embedding process is

applied on each frame to obtain $\hat{X}^{(m)}$. The watermark bit $\hat{w}_m$ indexed by m is extracted from the coefficient $\hat{X}_k^{(m)}$ of the transformed frame $\hat{X}^{(m)}$. For this purpose, we apply the inverse parametric QIM given by (11) as

$$\hat{w}_m = Q_\alpha^{-1}\left(\hat{X}_k^{(m)}, \Delta\right) = \begin{cases} 1 & \text{if } \frac{\Delta}{4} \le \text{mod}\left(\hat{X}_k^{(m)} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Finally, the binary image $\hat{\mathbf{I}}$ is obtained after extracting all the bits of the watermark and arrange them in a vector $\hat{\mathbf{W}}$, which is then converted to a two-dimensional image.

3 Performance analysis

In this section, we derive the theoretical expressions for the signal to watermark ratio (SWR) and the probability of error (and the bit error rate (BER)) in order to use them for optimizing the proposed watermarking technique.

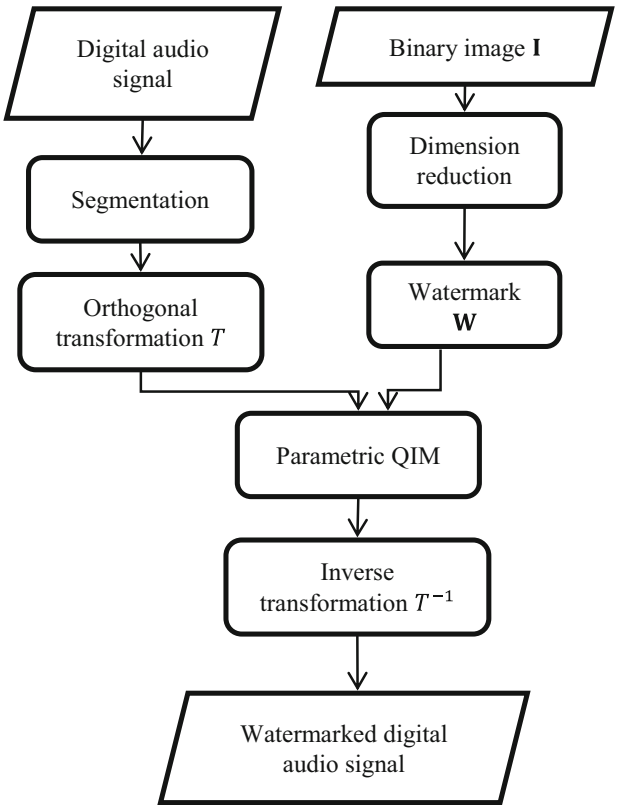

Fig. 1 Proposed watermark embedding algorithm

3.1 Signal to watermark ratio

The signal to watermark ratio is defined by

$$\mathrm{SWR} = \frac{P_S}{P_W} \tag{16}$$

where P_S is the signal power, which is the sum of the squared samples of the host audio signal divided by the signal length, and is given by

$$P_S = \frac{1}{L_w L} \sum_{m=1}^{L_w} \sum_{k=0}^{L-1} \left(x_k^{(m)} \right)^2$$

with L_w being the total number of the frames and P_W being the watermark power. The watermark power is the power of the noise added to the signal due to the embedding distortion and is given by

$$P_W = \left\langle \frac{1}{L} \sum_{l=0}^{L-1} \left(x_l^{(m)} - \hat{x}_l^{(m)} \right)^2 \right\rangle \tag{17}$$

where $x_l^{(m)}$ and $\hat{x}_l^{(m)}$ are the samples of the host and watermarked frames $x^{(m)}$ and $\hat{x}^{(m)}$, respectively, and $\langle \cdot \rangle$ denotes the statistical average operator. Since the frames $X^{(m)}$ and $\hat{X}^{(m)}$ are obtained by transforming the frames $x^{(m)}$ and $\hat{x}^{(m)}$, respectively, it is easy to show that P_W given by (17) can be expressed in terms of the difference

$$\varepsilon = X_k^{(m)} - \hat{X}_k^{(m)} \tag{18}$$

as

$$P_W = \frac{1}{L} \langle \varepsilon^2 \rangle = \frac{1}{L} \int_{\text{domain of } \varepsilon} \varepsilon^2 f_\varepsilon(\varepsilon) d\varepsilon \tag{19}$$

where $f_\varepsilon(\varepsilon)$ is the probability density function (PDF) of ε. The substitution of (13) in (18) yields

$$\varepsilon = X_k^{(m)} - \mathrm{round}\left(\frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta} \right) \Delta - w_m \frac{\Delta}{2} - \alpha \tag{20}$$

By adding and subtracting the quantity $\beta + \gamma w_m$, (20) can be rearranged as

$$\varepsilon = \left(X_k^{(m)} + \beta + \gamma w_m \right) - \mathrm{round}\left(\frac{X_k^{(m)} + \beta + \gamma w_m}{\Delta} \right) \Delta - w_m \frac{\Delta}{2} - \alpha - \beta - \gamma w_m \tag{21}$$

Thus, using the quantization given by (1), (21) can be expressed as

$$\varepsilon = z - \left(w_m \frac{\Delta}{2} + \alpha + \beta + \gamma w_m \right) \tag{22}$$

where

$$z = y - Q(y, \Delta) \tag{23}$$

and

$$y = X_k^{(m)} + \beta + \gamma w_m \tag{24}$$

The quantization step Δ in (23) must be chosen significantly smaller than the host sample y given by (24) in order to have a small quantization noise z. Therefore, according to Bennett's high rate model for quantization [17], the random variable z is uniformly distributed in the interval $\left[\frac{-\Delta}{2}, \frac{-\Delta}{2}\right]$, i.e., its PDF is given by

$$f_z(z) = \begin{cases} \frac{1}{\Delta} & z \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

Let us assume that the watermark binary image constitutes of p white pixels (p 1 s) and q black pixels (q 0 s). Therefore, the probability of the watermark bits can be expressed as

$$P(w_m) = \frac{1}{p+q}(q\delta(w_m) + p\delta(w_m - 1)) \tag{26}$$

where $\delta(l)$ is the Kronecker delta known as

$$\delta(l) = \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

It is clear that ε given by (22) is a function of z and w_m, which are two independent random variables. Therefore, using (22), (25) and (26) in (19), the watermark power can then be expressed as

$$P_W = \frac{1}{\Delta L} \sum_{w_m=0}^{1} \int_{\frac{-\Delta}{2}}^{\frac{\Delta}{2}} \left(z - w_m \frac{\Delta}{2} - \alpha - \beta - \gamma w_m\right)^2 \cdot \frac{(q\delta(w_m) + p\delta(w_m - 1))}{p+q} dz \tag{28}$$

After some mathematical manipulations, (28) reduces to

$$P_w = \frac{1}{L}\left(\frac{\Delta^2}{12} + (\alpha+\beta)^2 + \frac{q(2\gamma+\Delta)}{4(p+q)}((\Delta+2\gamma) + 4(\alpha+\beta))\right) \tag{29}$$

By substituting (29) in (16), an expression for the SWR is obtained as

$$\text{SWR} = \frac{P_S L}{\left(\frac{\Delta^2}{12} + (\alpha+\beta)^2 + \frac{q(2\gamma+\Delta)}{4(p+q)}((\Delta+2\gamma) + 4(\alpha+\beta))\right)} \tag{30}$$

3.2 Probability of error and the BER under an AWGN attack

In the case of transmission of the watermarked signal through an additive white Gaussian noise (AWGN) channel, the received noisy watermarked signal frame indexed by m is given by

$$\widehat{xn}^{(m)} = \hat{x}^{(m)} + n^{(m)} \tag{31}$$

where $n^{(m)}$ is the noise frame, which is assumed to be white Gaussian $\mathcal{N}(0, \sigma_n)$. By applying the linear discrete transformation T given by (12) on the frame $\widehat{xn}^{(m)}$ given by (31), we obtain

$$\widehat{XN}^{(m)} = \hat{X}^{(m)} + N^{(m)} \tag{32}$$

where $N^{(m)}$ is the transform of $n^{(m)}$ and it can be shown that it is also white Gaussian $\mathcal{N}(0, \sigma_n)$.

In the following, without loss of generality, we assume that the watermarked transformed frame $\hat{X}^{(m)}$ was obtained by embedding in the original host transformed frame $X^{(m)}$ the watermark bit $w_m = 0$ (the same analysis lead to the same results in the case of $w_m = 1$).

The extraction of the watermark bit from the received noisy frame $\widehat{XN}^{(m)}$ is performed according to (15) as

$$\begin{aligned}\hat{w}_m &= \begin{cases} 1 & \text{if } \frac{\Delta}{4} \le \text{mod}\left(\widehat{XN}_k^{(m)} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } \frac{\Delta}{4} \le \text{mod}\left(\hat{X}_k^{(m)} + N_k^{(m)} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } \frac{\Delta}{4} \le \text{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{33}$$

To obtain (33), we have used the fact that $\left(\hat{X}_k^{(m)} - \alpha\right)$ is a multiple of Δ, which can clearly be seen from (13). It is clear from (33) that a transmission error occurs in the case of

$$\frac{\Delta}{4} \le \text{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4} \tag{34}$$

which is satisfied for all values of $N_k^{(m)}$ in the range

$$r\Delta + \frac{\Delta}{4} \le N_k^{(m)} \le r\Delta + \frac{3\Delta}{4} \tag{35}$$

where r is any integer. Therefore, the probability of error is the sum over all possible values of r and is given by

$$P_e = P\left(\frac{\Delta}{4} \le \text{mod}\left(N_k^{(m)}, \Delta\right) < \frac{3\Delta}{4}\right) = \sum_{r=-\infty}^{\infty} P\left(r\,\Delta + \frac{\Delta}{4} \le N_k^{(m)} \le r\,\Delta + \frac{3\Delta}{4}\right) \tag{36}$$

Since $N^{(m)}$ is $\mathcal{N}(0, \sigma_n)$, the probability in (36) can be expressed as

$$P\left(r\,\Delta + \frac{\Delta}{4} \le N_k^{(m)} \le r\,\Delta + \frac{3\Delta}{4}\right) = \int_{r\,\Delta+\frac{\Delta}{4}}^{r\,\Delta+\frac{3\Delta}{4}} \frac{e^{-\frac{N_k^{(m)2}}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}}\, dN_k^{(m)} \tag{37}$$

After computing (37) and using it in (36), the probability of error becomes

$$P_e = \frac{1}{2} \sum_{r=-\infty}^{\infty} \left(\text{erf}\left(\frac{\sqrt{2}\Delta\,(4r+3)}{8\sigma_n}\right) - \text{erf}\left(\frac{\sqrt{2}\,\Delta(4r+1)}{8\sigma_n}\right)\right) \tag{38}$$

Finally, using the fact that $\sigma_n^2 = P_n$, where P_n is the noise power in the time domain, (38) can be expressed as

$$P_e = \frac{1}{2} \sum_{r=-\infty}^{\infty} \left(\text{erf}\left(\frac{\Delta\,(4r+3)}{\sqrt{32P_n}}\right) - \text{erf}\left(\frac{\Delta(4r+1)}{\sqrt{32P_n}}\right)\right) \tag{39}$$

The bit error rate is defined in terms of the probability of error as

$$BER \overset{\text{def}}{=} 100P_e \tag{40}$$

4 Optimization technique

The main objective of this section is to find the values of the parameters α, β, γ and Δ for which the parametric QIM defined in (10) becomes optimal for the proposed watermarking technique. The criterion for the proposed optimization technique is to guarantee the imperceptibility while maximizing the robustness under an AWGN attack. This is equivalent to fix the SWR to a value SWR_0, which is according to IFPI higher than 20 dB, while minimizing the probability of error under an AWGN attack.

It is first worth to notice that the probability of error derived in (39) depends only on the quantization step Δ and does not depend on the parameters α, β and γ, that is, $P_e = P_e(\Delta)$. However, the SWR derived in (30) depends on α, β, γ and Δ. Thus, in order to make the SWR constant, it is sufficient to fix the following function to a constant as

$$g(\alpha,\beta,\gamma,\Delta) = \frac{\Delta^2}{12} + (\alpha+\beta)^2 + \frac{q(2\gamma+\Delta)}{4(p+q)}((\Delta+2\gamma)+4(\alpha+\beta)) = \text{constant} \tag{41}$$

Therefore, the optimization problem can easily be solved using the method of Lagrange multipliers [27], where the Lagrangian can be defined as

$$L(\alpha,\beta,\gamma,\Delta) = g(\alpha,\beta,\gamma,\Delta) - \frac{1}{\kappa}P_e(\Delta) \tag{42}$$

with κ is the Lagrange multiplier. The optimal values of the parameters α, β and γ and the quantization step Δ can then be obtained by solving the following system of equations

$$\begin{cases} \overrightarrow{\nabla}_{\alpha,\beta,\gamma,\Delta} L(\alpha,\beta,\gamma,\Delta) = \overrightarrow{0} \\ SWR(\alpha,\beta,\gamma,\Delta) = \text{SWR}_0 \end{cases} \tag{43}$$

where

$$\overrightarrow{\nabla}_{\alpha,\beta,\gamma,\Delta} = \frac{\partial}{\partial\alpha}\overrightarrow{e_\alpha} + \frac{\partial}{\partial\beta}\overrightarrow{e_\beta} + \frac{\partial}{\partial\gamma}\overrightarrow{e_\gamma} + \frac{\partial}{\partial\Delta}\overrightarrow{e_\Delta} \tag{44}$$

is the gradient operator. Finally, the solution of (43) is obtained as

$$\begin{cases} \alpha \in \mathbb{R} \\ \beta = -\alpha \\ \Delta = \sqrt{\dfrac{12P_sL}{\text{SWR}_0}} \\ \gamma = -\dfrac{\Delta}{2} \end{cases} \tag{45}$$

It is worth to notice that the optimal quantization step Δ in (45) depends only on P_s, L and SWR_0. Consequently, the quantization step is the same for all the frames of the same audio signal, which means that the proposed quantization operation given by (13) is uniform. Using the optimal solution for the quantization step Δ given by (45) in (39), the probability of error reduces to

$$P_e = \frac{1}{2}\sum_{r=-\infty}^{\infty}\left(\operatorname{erf}\left(\sqrt{\frac{3L\times \mathrm{SNR}}{8\ \mathrm{SWR}_0}}(4r+3)\right)-\operatorname{erf}\left(\sqrt{\frac{3L\times \mathrm{SNR}}{8\ \mathrm{SWR}_0}}(4r+1)\right)\right) \tag{46}$$

where $\mathrm{SNR} = P_s/P_n$ is the signal to noise ratio. It is seen from (46) that the probability of error under an AWGN attack is independent of the considered discrete transform and the coefficient index k (the insertion position used to carry the watermark bit).

In order to provide a closed-form expression for the probability of error function obtained in (46) as an infinite summation, we propose an overall approximation of the form

$$\widetilde{P}_e = \operatorname{erfc}\left(\frac{1}{a-be^{-\frac{c}{x^d}}}\right) \tag{47}$$

which is parameterized by the parameters a, b, c, and d, where

$$x = \sqrt{\frac{3\ L\times \mathrm{SNR}}{8\ \mathrm{SWR}_0}} \tag{48}$$

It should be mentioned that we have tried many forms, but the one given by (47) is the best.

The maximum error between the probability of error function P_e obtained in (46) and its approximation $\widetilde{P}_e$ provided in (47) is given in terms of the parameters a, b, c and d as

$$E(a,b,c,d) = \max_{x>0}\left|P_e-\widetilde{P}_e\right| \tag{49}$$

We use the particle swarm optimization (PSO) to find the numerical values a_0, b_0, c_0, and d_0 of the parameters a, b, c and d, respectively, that minimize the maximum error in (49) and thus, ensure that $\widetilde{P}_e$ is a good approximation of P_e. The PSO is an iterative numerical optimization technique, which is able to find a solution that globally minimizes any objective function [10, 12, 23]. In our case, we use $E(a, b, c, d)$ as an objective function for PSO to find the solution (a_0, b_0, c_0, d_0) that satisfies

$$\forall (a,b,c,d)\in\mathbb{R}^4 : E(a_0,b_0,c_0,d_0)\le E(a,b,c,d) \tag{50}$$

Let $\{S_1, S_2, S_3, \ldots, S_N\}$ be a swarm of N particles, where the position and velocity of the k^{th} particle at the t^{th} iteration are given by $R_k(t) = [a_k(t), b_k(t), c_k(t), d_k(t)]$ and $v_k(t) = \left[v_k^{(a)}(t), v_k^{(b)}(t), v_k^{(c)}(t), v_k^{(d)}(t)\right]$, respectively, with $k = 1, 2, 3, \ldots, N$. The velocity and position of the k^{th} particle are calculated iteratively as

$$v_k(t+1) = \theta\cdot v_k(t) + \varsigma_1\, r_k^{(1)}(t)\left(\hat{R}_k(t)-R_k(t)\right) + \varsigma_2\, r_k^{(2)}(t)\,(G(t)-R_k(t)) \tag{51}$$

and

$$R_k(t+1) = R_k(t) + v_k(t+1) \tag{52}$$

respectively, where $\hat{R}_k(t) = \left[\hat{a}_k(t), \hat{b}_k(t), \hat{c}_k(t), \hat{d}_k(t)\right]$ is the personal best position in the position history of the k^{th} particle, i.e.,

$$\forall \tau\le t : E\left(\hat{a}_k(t), \hat{b}_k(t), \hat{c}_k(t), \hat{d}_k(t)\right)\le E(a_k(\tau), b_k(\tau), c_k(\tau), d_k(\tau)), \tag{53}$$

$G(t) = [a(t), b(t), c(t), d(t)]$ is the global best position in the position history of the swarm, i.e.,

$$\forall k \in \{1, 2, 3, \ldots, N\}: \quad E(a(t), b(t), c(t), d(t)) \leq E\left(\hat{a}_k(t), \hat{b}_k(t), \hat{c}_k(t), \hat{d}_k(t)\right), \tag{54}$$

$r_k^{(1)}(t)$ and $r_k^{(2)}(t)$ are random numbers uniformly distributed in the interval [0, 1], ς_1 and ς_2 are real-valued acceleration constants, which modulate the magnitude of the steps taken by the particle in the direction of its personal and global best positions, respectively, and θ is the inertia weight.

The implementation of PSO requires values for the initial conditions $R_k(0)$ and velocities $v_k(0)$, $k = 1, 2, 3, \ldots, N$, and parameters N, θ, ς_1, and ς_2. However, the choice of the required values has a large impact on the optimization performance. This problem has been the subject of many research and is still in debate [23]. In our implementation, we randomly and uniformly initialize the components of the particle positions and velocities from the intervals [0, 2] and [−2, 2], respectively, and use the values $\theta = 0.9$, $\varsigma_1 = \varsigma_2 = 0.5$, $N = 1000$. The maximization over x in (49) should theoretically be taken over the infinite interval $]0, \infty[$. In our case, we perform the maximization over the interval [0.001, 10], which is sufficiently large due to the fact that P_e in (46) is a monotonically decreasing function and its values are less than 10^{-10} for $x > 5$. We evaluate P_e for $|r| \leq 5000$ and confirm that larger values of $|r|$ lead to negligible terms having total contribution less than 10^{-7}. For the convergence criterion of PSO, we stop iterating (51) and (52) when the maximum speed of all particles is less than 10^{-8}, that is

$$\max_{k \in \{1,2,3,\ldots,N\}} \left(\sqrt{\left(v_k^{(a)}(t)\right)^2 + \left(v_k^{(b)}(t)\right)^2 + \left(v_k^{(c)}(t)\right)^2 + \left(v_k^{(d)}(t)\right)^2} \right) \leq 10^{-8} \tag{55}$$

Finally, the solution of the optimization problem is taken as the global best position $[a_0, b_0, c_0, d_0] = G(t_s) = [a(t_s), b(t_s), c(t_s), d(t_s)]$, where t_s is the iteration at which the convergence is attained, and the obtained numerical values are $a_0 = 2.1143$, $b_0 = 1.7831$, $c_0 = 0.4669$ and $d_0 = 2.0128$. By substituting this solution and (48) in (47), we obtain

$$\widetilde{P}_e = \operatorname{erfc}\left(\frac{1}{2.1143 - 1.7831 e^{-\frac{0.4669}{\sqrt{\frac{3L \times \mathrm{SNR}}{8\ \mathrm{SWR}_0}}^{2.0128}}}} \right) \tag{56}$$

The exact probability error P_e derived in (46) and its approximation $\widetilde{P}_e$ obtained in (56) are plotted in Fig. 3. In this figure, the evaluation of the exact probability function is done by taking the sum of the terms corresponding to $|r| \leq 5000$, which implies a plotting precision of order 10^{-7}. This figure shows clearly that $\widetilde{P}_e$, which has a closed-form expression, is a good overall approximation of P_e.

Using the optimal solutions obtained in (45) in the proposed parametric QIM defined in (10), a new one-parameter optimal QIM denoted by QIM^{α} can be defined as

$$\hat{X} = Q_\alpha(X, \Delta, w) = \operatorname{round}\left(\frac{X - w\frac{\Delta}{2} - \alpha}{\Delta} \right) \Delta + w\frac{\Delta}{2} + \alpha \tag{57}$$

and the corresponding watermark extraction process as

$$\hat{w} = Q_{\alpha}^{-1}\left(\hat{X}, \Delta\right) = \begin{cases} 1 & \text{if } \frac{\Delta}{4} \leq \text{mod}\left(\hat{X} - \alpha, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \tag{58}$$

As mentioned earlier, the parameter α can take any real value. However, it is appropriate to choose it arbitrarily from the interval $[-\Delta/2, \Delta/2]$ and then use it as a secret key. To make this key more effective and significantly enhance the security of the proposed watermarking technique, we use for different frames different values of the parameter α, i.e., α_m for the m^{th} frame. Since the proposed method consists of inserting one bit per frame, the key becomes a vector of length equals to the number of the watermark bits. The values α_m of this vector can be selected value by value by the user, or simply by using chaotic maps, which have the ability to generate chaotic vectors of any desired length from a given initial condition known usually as the seed or simply as the key. Therefore, the coefficient $X_k^{(m)}$ of the transformed audio frame is modified using (57) as

$$\hat{X}_k^{(m)} = \text{round}\left(\frac{X_k^{(m)} - w_m \frac{\Delta}{2} - \alpha_m}{\Delta}\right)\Delta + w_m \frac{\Delta}{2} + \alpha_m \tag{59}$$

and the watermark bit is extracted as

$$\hat{w}_m = \begin{cases} 1 & \text{if } \frac{\Delta}{4} \leq \text{mod}\left(\hat{X}_k^{(m)} - \alpha_m, \Delta\right) < \frac{3\Delta}{4} \\ 0 & \text{otherwise} \end{cases} \tag{60}$$

5 Proposed procedure for best watermark embedding positions

We have shown in Section 4 that in the optimal case, the probability of error under an AWGN attack is independent on the considered orthogonal linear transform and watermark insertion position. However, it is desirable to select the transform and embedding positions that offer robustness against other signal processing manipulations. In this section, we propose a procedure to find the best interval for selecting the watermark embedding positions.

It is clear that the re-quantization operation additively introduces a uniform noise in the time domain, and for large frame length L, it converges to a Gaussian noise in the transform domain, hence its probability of error is independent of the used transform and embedding position. The resampling and lossy compression are operations that can be regarded as additive noise and filtering. Therefore, the essential operations that should be considered for choosing a transform and an embedding position are the high and low pass filtering operations. Although the discrete Fourier transform (DFT) gives a good interpretation of filtering attacks, the DCT can be used to give a similar interpretation and is preferable for its energy compaction capability and real-valued nature. It is clear that if the embedding position is close to the DC coefficient, the watermark survives in the case of the low pass filtering and fail to survive in the case of the high pass filtering, and if the embedding position is near high frequencies, then the contrary occurs. Therefore, an embedding position is selected to provide a good trade-off between the high and low pass filtering effects. Fig. 4 shows, for different embedding positions, the empirical maximum BER corresponding to high and low pass filtering attacks with a cut-off frequency of 100 Hz and 11,025 Hz, respectively, where the sampling frequency

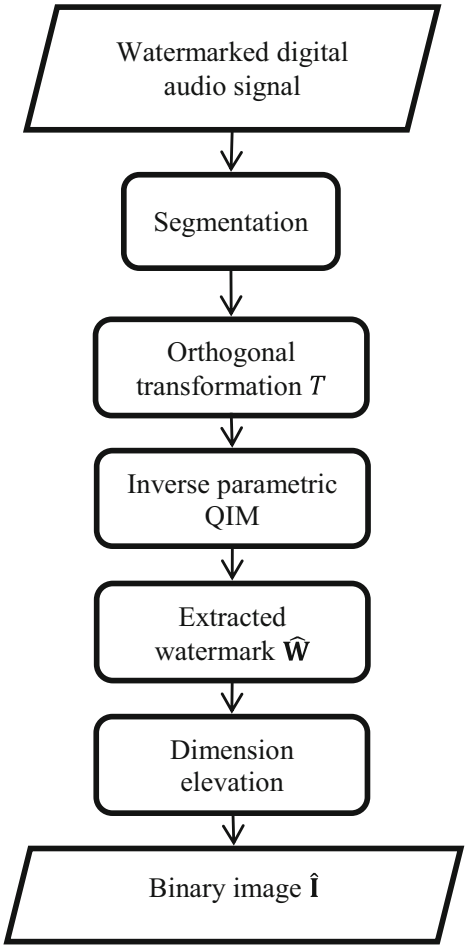

Fig. 2 Proposed watermark extraction algorithm

of the audio signals used is 44,100 Hz, and the frame length is 256. This empirical probability of error is obtained by taking the mean of about 100 experiments on different types of audio signals. The maximum BER (MBER) is calculated as

$$\mathrm{MBER} = \max(\mathrm{BERH}, \mathrm{BERL}) \tag{61}$$

where BERH and BERL denote the BER resulted from the high and low pass filtering, respectively.

The results shown in Fig. 4 allow us to claim that the region where $k \in [12, 82]$ gives a good trade-off between the two filtering attacks. This is because the maximum BER is flatly approaching zero in this region. Further experiments that we have carried out for this purpose allowed us to propose the region $k \in [0.05\ L, 0.32\ L]$ for the best selection of the insertion position k valid for any arbitrary frame length L, where the value of k must be integer. The value of k can be chosen randomly from the proposed region and is not necessarily the same for all the frames. Therefore, the insertion position k_m is used for the frame indexed by m and the vector of length equals to the number of the watermark bits containing all the insertion

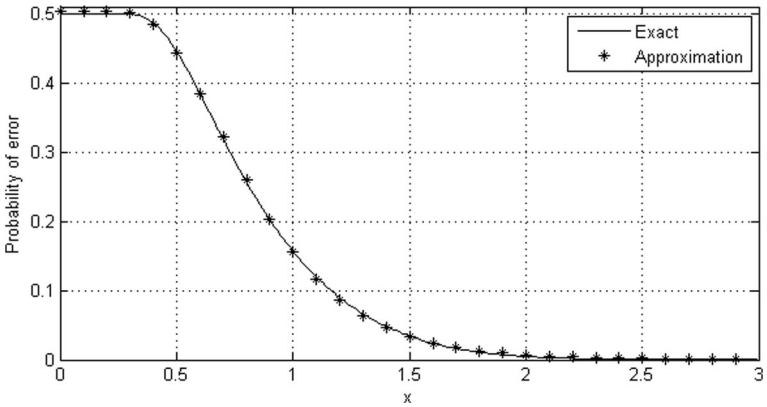

Fig. 3 Efficiency of the proposed probability approximation

positions can be coupled with the vector containing all the values of the parameter α_m to form a secret key for the proposed watermarking technique.

6 Fast implementation of the proposed watermarking technique

The computational complexity of a transform-based QIM watermarking technique is mainly due to the required forward and inverse transforms. Although this complexity can be reduced using fast algorithms to compute the forward and inverse transforms, it is still a great challenge for modern systems requiring faster watermark embedding and extraction algorithms. Fortunately, the computational complexity can significantly be reduced in the proposed watermarking technique and in any other technique that modifies only one coefficient per frame in the transform domain. This can easily be achieved using the fast implementation scheme proposed bellow.

The main idea behind the proposed fast implementation scheme to pass from the time domain to the transform domain and vice versa is an appropriate exploitation of the fact that only one watermark bit w_m is embedded in the transformed frame $X^{(m)}$, specifically in the coefficient $X_k^{(m)}$, which is the only coefficient need to be computed using the forward

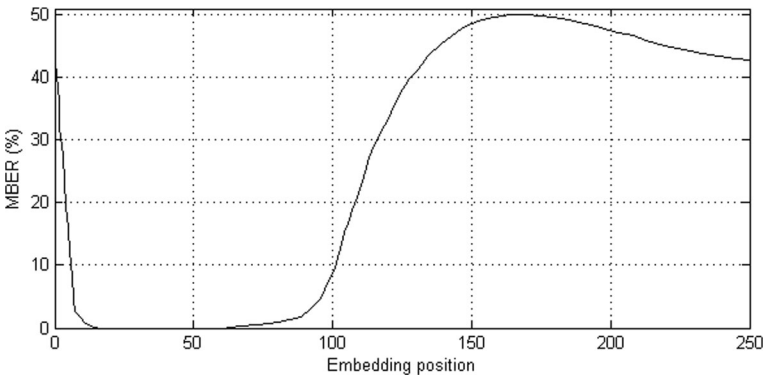

Fig. 4 Maximum BER resulted from low and high pass filtering attacks

transformation given by (12) for $r = k$, where k is the insertion position. It is clear that in this case at most L multiplications and $(L-1)$ additions are required. If the kernel $\phi_l(k)$ of the transform has some symmetry properties like in the case of the DFT or DCT, then the complexity can further be reduced. Specifically, in the case of the DCT of length that is an integral power of two, the required number of multiplications is $L/2$.

After embedding the bit w_m in the coefficient $X_k^{(m)}$ to get the coefficient $\hat{X}_k^{(m)}$ according to (13), the inverse transform given by (14) is then applied on the resulting watermarked frame $\hat{X}^{(m)}$ to obtain the watermarked frame $\hat{x}_l^{(m)}$, $l = 0, 1, \ldots, L-1$, in the time domain. Since the coefficients of $\hat{X}^{(m)}$ are all identical to those of $X^{(m)}$ except the one situated in the insertion position k, the following holds

$$\hat{X}_r^{(m)} = X_r^{(m)} + \left(\hat{X}_k^{(m)} - X_r^{(m)}\right)\delta(k-r), \quad r = 0, 1, \ldots, L-1 \tag{62}$$

By substituting (62) in (14), the watermarked frame in the time domain can simply be computed as

$$\hat{x}_l^{(m)} = x_l^{(m)} + \left(\hat{X}_k^{(m)} - X_k^{(m)}\right)\phi_l(k), \quad l = 0, 1, \ldots, L-1 \tag{63}$$

It is clear that (63) requires at most L multiplications and $(L+1)$ additions. For the case of the DCT, the kernel is given by

$$\phi_l(r) = \Lambda(r)\cos\left(\frac{\pi\,(2l+1)\,r}{2L}\right) \tag{64}$$

with

$$\Lambda(r) = \begin{cases} \dfrac{1}{\sqrt{L}} & \text{if } l = 0 \\ \sqrt{\dfrac{2}{L}} & \text{otherwise} \end{cases} \tag{65}$$

In this case, and for values of L that are integral powers of two, the following symmetry

$$\phi_{L-1-u}(k) = (-1)^k \phi_u(k), \quad u = 0, 1, \ldots, \frac{L}{2} \tag{66}$$

can be exploited in (63) to further reduce the number of the required multiplications to only $L/2$.

Therefore, using the proposed approach described above, the total computational complexity required by the forward and inverse transformations in the proposed watermark embedding algorithm (or in any other technique that modifies only one coefficient per frame in the DCT domain) is only $L \times q$ multiplications and $2L \times q$ additions, where q is the number of watermark bits. The total computational complexity required by the forward transform in the proposed watermark extraction algorithm is only $(L/2) \times q$ multiplications and $(L-1) \times q$ additions. This is because the inverse transformation is not needed in the extraction process. Thus, the proposed approach reduces the complexity significantly compared to that required by the direct use of fast algorithms for the forward and inverse DCTs. For instance, if $L = 256$, then the proposed approach requires $256 \times q$ multiplications and $512 \times q$ additions in the

embedding process, whereas the direct use of the fast algorithms reported in [24] for the forward and inverse DCTs requires $2048 \times q$ multiplications and $5634 \times q$ additions.

7 Experimental results

In this section, we evaluate the performance and robustness of the proposed audio watermarking technique and perform a comparison with the existing techniques reported in [3, 9, 14, 29] that are based on QIM. For this purpose, we apply these and proposed techniques for watermarking audio signals that are stored in 16-bit signed mono wave form audio format files sampled at a frequency of 44.1 kHz. We present here only the experimental results obtained for three audio signals; namely music like signal, human speech like signal and mixture of music and human speech signal denoted by S1, S2 and S3, respectively. The evaluation in terms of bit error rate is carried out by considering the following known attacks:

- ***Re-sampling***: the watermarked audio signal is down-sampled from 44.1 kHz to 22.05, 11.025 or 5 kHz and the resulting signal is then up-sampled to 44.1 kHz.
- ***Common lossy audio compressions***: we consider the MPEG layer III compression, usually referred to as MP3, which is the most popular lossy compression in the music industry, and supported by many video file formats such as Audio Video Interleave (AVI), Matroska, MPEG-4 part 14 (MP4), MPEG-1 (MPG), Materiel eXchange Format (MXF), QuickTime, etc. We apply MP3 at several bit rates 192, 128, 96, 80 and 64 kbps. We also consider other common lossy audio compressions, namely OPUS, advanced audio coding (AAC) and Vorbis. OPUS is used at a typical bit rate of 96 kbps. For the case of AAC, we apply two different versions, the low complexity AAC (LC-AAC), which is designed to reduce the complexity at the cost of quality and thus we apply it at a bit rate of 96 kbps, and the high efficiency AAC (HE-AAC), which is optimized to compress audio at low bit rates and thus we apply it at a bit rate of 56 kbps. The metric for Vorbis audio compression is quality, therefore, we perform Vorbis compression at a representative quality of 50%. The watermarked audio signal is compressed and then decompressed. Subsequently, the watermark is extracted in order to evaluate the robustness.
- ***Low pass filtering***: a low pass filter of a cut-off frequency of 11.025, 8 or 6 kHz is used to filter the watermarked audio signal.
- ***High pass filtering***: a high pass filter of a cut-off frequency of 20, 50, 100 or 4000 Hz is employed to filter the watermarked audio signal.
- ***Additive white Gaussian noise***: a white Gaussian noise is added to the watermarked audio signal with an SNR of 20, 15 or 10 dB.
- ***Re-quantization***: each of the watermarked audio samples is re-quantized from 16 bits to 8 or 4 bits.
- ***Amplitude scaling***: the watermarked audio signal is scaled by a factor of 0.8, 0.9, 1.1 or 1.2.

It is well known that there is a trade-off between the watermark payload, transparency measured by the SWR and robustness [6, 31, 32]. Therefore, for a fair comparison, we test the robustness for different attacks by fixing the SWR and payload.

The theoretical SWR that can be calculated using (30) is fixed to a desired level of 30 dB, which satisfies the IFPI recommendation and gives a good compromise between robustness

and imperceptibility. As discussed earlier, the theoretical SWR can be calculated from (30) using $\alpha = \frac{\Delta}{4}, \beta = -\frac{\Delta}{4}$ and $\gamma = 0$ for the techniques reported in [3, 9, 29], and $\beta = -\alpha$ and $\gamma = -\frac{\Delta}{2}$ for the proposed technique. This desired theoretical SWR and the corresponding experimental SWRs obtained by different techniques are given in Table 1 for different audio signals. It is seen from this table that the theoretical and experimental values of SWR are similar. Furthermore, to confirm the validity of (30), we carried out many experiments with different quantization step sizes and two distinct frame lengths (i.e., $L = 256$ and $L = 512$). The theoretical and experimental results are compared in Fig. 5, which clearly shows the agreement of the theoretical and experimental SWR and confirms the reliability of the expression derived in (30) for the SWR. It should be noted that the desired theoretical SWR cannot be fixed for the technique reported in [14], since it uses a psychoacoustic model as a measure of transparency. Therefore, for a fair comparison, we use for [14] the obtained experimental SWRs that are less than 30 dB.

The watermark payload (embedding capacity) is the number of watermark bits embedded per unit time (e.g. seconds) in the host digital audio signal. According to the IFPI [29], the payload should be greater than 20 bps (bits per second). Since the proposed technique embeds one bit per frame of length L, its watermark payload is obtained as

$$\mathcal{P} = \frac{f_s}{L} \tag{67}$$

where f_s is the sampling frequency of the host audio signal. The proposed technique can be applied for various desired watermark payloads. Since the payloads of the techniques in [3, 9, 29] are 172, 172 and 86 bps, respectively, we fix the payload given by (67) of the proposed technique to 172 and then to 86. The proposed

Table 1 Watermark payload and SWR of different watermarking techniques

Technique	Payload Bits/s	Domain	Signal	SWR (dB)	
				desired	Obtained
Proposed172	172	256-DCT	S1	30	29.79
			S2	30	29.80
			S3	30	29.92
Chen [9]		7 level DWT	S1	30	30.42
			S2	30	30.29
			3	30	30.29
Wu [29]		8 level DWT	S1	30	29.92
			S2	30	29.97
			S3	30	29.88
Proposed86	86	512-DCT	S1	30	29.97
			S2	30	29.72
			S3	30	29.98
Chen [3]		8 level DWT	S1	30	30.28
			S2	30	30.27
			S3	30	30.10
Hu [14]	85	4096-DCT	S1	/	28.58
			S2	/	25.74
			S3	/	23.50

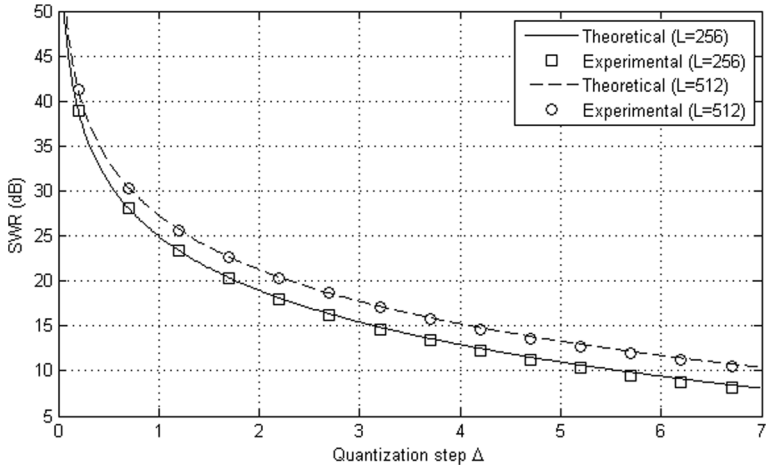

Fig. 5 Theotical and experimental SWR

technique versions corresponding to $\mathcal{P} = 172$ and $\mathcal{P} = 86$ are denoted by Proposed172 and Proposed86, respectively. The embedding positions in these two versions are selected to be the sample numbers 47 and 94, respectively, which belong to the best region defined in Section 4. In [14], the authors proposed four variants of their technique, which differ in payload. Therefore, we select the variant with 85 bps of [14] and compare it to Proposed86 and the technique reported in [3].

The experimental results are compared in terms of the BER in Tables 2, 3, 4, 5, 6, 7, 8, and 9. Tables 2 and 3 show the BER obtained by different techniques in the cases of the resampling and low-pass filtering attacks. It is clear from these tables that all the techniques are equally robust to the two attacks. Table 4 shows clearly that the proposed

Table 2 Re-sampling BER

Re-sampling frequency (Hz)	Signal	22,050	11,025	5000
Proposed172	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Chen [9]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Wu [29]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Proposed86	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Chen [3]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Hu [14]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0

Table 3 Low-pass filtering BER

Cut-off frequency (Hz)	Signal	11,025	8000	6000
Proposed172	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Chen [9]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Wu [29]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Proposed86	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Chen [3]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0
Hu [14]	S1	0	0	0
BER (%)	S2	0	0	0
	S3	0	0	0

technique versions are more robust against MPEG layer III compression attack than the corresponding existing techniques reported in [3, 9, 14, 29]. Particularly, in the case of low bit rate compression, the robustness of the proposed technique becomes significantly remarkable. The same observation can be seen from Table 5, which illustrates the robustness against common lossy audio compressions. The reason behind the superiority of the proposed technique is the embedding region selection procedure introduced in Section 5, which ensures that the watermark embedding is in a significant region of the audio signal, and the optimization procedure developed in Section 4, which ensures

Table 4 MPEG layer III compression BER

Bit rate (kbps)	Signal	192	128	96	80	64
proposed172	S1	0	0	0.5859	2.0508	4.8584
BER (%)	S2	0	0	0	0	0.3418
	S3	0	0	0.1465	0.5127	2.3193
Chen [9]	S1	0	1.1963	8.0078	17.6758	28.4668
BER (%)	S2	0	0	0.8301	2.2949	7.7637
	S3	0	0.0977	2.4414	7.5195	15.7227
Wu [29]	S1	0	1.0986	6.7139	15.0879	26.3916
BER (%)	S2	0	0	0.6592	2.2461	6.4941
	S3	0	0.0488	2.0020	6.1523	13.2080
Proposed86	S1	0	0	0	0.1465	0.8301
BER (%)	S2	0	0	0	0	0
	S3	0	0	0	0.1465	0.7080
Chen [3]	S1	0	0	0.8057	3.4668	9.4971
BER (%)	S2	0	0	0.4395	1.2695	3.4912
	S3	0	0	0.5127	2.0752	6.0059
Hu [14]	S1	0	0	0.1953	1.2451	4.0283
BER (%)	S2	0	0	0.0244	0.3662	0.6104
	S3	0	0	0	0.2197	0.8545

Table 5 BER of common audio compressions at typical bitrate/quality

Compression type	Signal	MP3 (96 kbps)	OPUS (96 kbps)	LC-AAC (96 kbps)	HE-AAC (56 kbps)	Vorbis (Quality 50/100)
Proposed172	S1	0.5859	1.3428	0.2197	10.8398	3.1982
BER (%)	S2	0	0.3906	2.6123	5.7617	1.6357
	S3	0.1465	0.5615	0.4395	9.4482	1.5381
Chen [9]	S1	8.0078	41.7725	0.6104	35.0098	16.1621
BER (%)	S2	0.8301	54.5654	5.9814	45.2637	18.0420
	S3	2.4414	39.5996	1.8066	36.7676	14.0381
Wu [29]	S1	6.7139	50.7080	0.4639	32.5684	15.1123
BER (%)	S2	0.6592	54.2236	5.5664	44.9951	16.7236
	S3	2.0020	53.4180	7.9102	35.9863	13.0615
Proposed86	S1	0	0.1953	0.0977	3.4668	0.5371
BER (%)	S2	0	0.0488	0.9766	2.7100	0.7080
	S3	0	0.2197	0.1953	4.9316	0.4883
Chen [3]	S1	0.8057	44.5068	0.7080	22.8516	8.9111
BER (%)	S2	0.4395	52.8076	2.6367	40.4785	11.5723
	S3	0.5127	44.8242	0.4395	30.1270	6.2500
Hu [14]	S1	0.1953	0.2441	0.7813	8.3496	0.5859
BER (%)	S2	0.0244	0.3662	1.0010	8.5449	0.7080
	S3	0	0.6592	0.7080	7.3486	0.3418

robustness to the additive noise introduced by compression. The fragility of the methods reported in [3, 9, 29] against lossy compression attacks is due to the embedding in the lowest approximation band of the DWT that is fragile to lossy compression. The method in [14] performs better than the DWT-based techniques as it only embeds a small fraction of the watermark bits in the low frequency of the DCT coefficients, and it is worse than the proposed technique because of the use of the adaptive quantization step that is fragile to compression noise. Table 6 demonstrates that the proposed technique is strongly

Table 6 High-pass filtering BER

Cut-off frequency (Hz)	Signal	20	50	100	4000
Proposed172	S1	0	0	0	0.9766
BER (%)	S2	0	0	0	0.2930
	S3	0	0		0.5615
Chen [9]	S1	49.8779	48.8037	50.1465	50.2197
BER (%)	S2	50.5859	51.2451	50.1221	50.5371
	S3	49.4873	50.8057	50.1221	49.7314
Wu [29]	S1	49.4629	49.3408	49.7559	49.4629
BER (%)	S2	48.7061	50.7568	50.3174	50.6104
	S3	50.0000	49.8535	50.1953	49.7559
Proposed86	S1	0	0	0	0.0244
BER (%)	S2	0	0	0	0
	S3	0	0	0	0
Chen [3]	S1	49.0967	50.3662	50.0977	49.7070
BER (%)	S2	51.0986	49.6094	51.4648	51.0742
	S3	49.1211	50.2197	50.7568	49.9756
Hu [14]	S1	0	0	0	50.4832
BER (%)	S2	0	0	0	49.9753
	S3	0	0	0	50.4437

Table 7 Additive white Gaussian noise BER

SNR (dB)	Signal	20	15	10
Proposed172	S1	0	1.4648	16.8213
BER (%)	S2	0	1.3916	17.0898
	S3	0	1.3672	16.0645
Chen [9]	S1	0.1709	7.1777	33.1787
BER (%)	S2	0.2197	8.4473	33.0322
	S3	0.2441	7.9102	32.7637
Wu [29]	S1	0.1221	6.6650	30.1758
BER (%)	S2	0.0732	6.5674	29.0283
	S3	0.1709	6.0303	29.5410
Proposed86	S1	0	0.0977	5.4199
BER (%)	S2	0	0	5.2246
	S3	0	0.0488	4.5654
Chen [3]	S1	0	1.3184	16.0889
BER (%)	S2	0	1.2939	16.1289
	S3	0	1.2695	16.9189
Hu [14]	S1	0.7324	5.1758	17.9932
BER (%)	S2	2.0996	9.6191	22.9248
	S3	0.1465	1.4648	8.8867

robust against the high-pass filtering attack even at exaggerated high cut-off frequencies, e.g., 4 kHz, and the existing techniques are significantly fragile to this attack even at low cut-off frequencies, e.g., 20 Hz, except the technique in [14], which is fragile only in the case of exaggerated high cut-off frequencies. The reason behind the strong robustness of the proposed technique in this attack is the proposed approach for the best watermark embedding positions introduced in Section 4, which is specifically designed to resist to high and low pass filtering. The weakness of the techniques in [3, 9, 29] is mainly due to the embedding in low frequencies of the host audio signal, which is very sensitive to the high-pass filtering. Table 7 shows that the proposed technique versions are more robust

Table 8 Re-quantization BER

16 bits to	Signal	8 bits	4 bits
Proposed172	S1	0	0.3447
BER (%)	S2	0	0.3906
	S3	0	0.1465
Chen [9]	S1	0	14.0869
BER (%)	S2	0	6.5430
	S3	0	1.7090
Wu [29]	S1	0	11.0840
BER (%)	S2	0	5.0537
	S3	0	1.2939
Proposed86	S1	0	0.0732
BER (%)	S2	0	0.0732
	S3	0	0
Chen [3]	S1	0	2.0508
BER (%)	S2	0	4.1992
	S3	0	0.3174
Hu [14]	S1	0	6.3721
BER (%)	S2	0	9.7412
	S3	0	0.8057

Table 9 Amplitude scaling BER

Scaling factor	Signal	0.8	0.9	1.1	1.2
Proposed172	S1	2.2949	0.2441	0.2441	2.2949
BER (%)	S2	3.7354	0.0488	0.0488	3.7354
	S3	3.3203	0.2197	0.2197	3.3203
Chen [9]	S1	44.3604	38.5986	38.5986	51.2939
BER (%)	S2	47.6563	40.7227	40.7227	51.1963
	S3	39.5508	34.5215	34.5215	45.3613
Wu [29]	S1	44.3848	37.6221	37.6221	51.1475
BER (%)	S2	46.9727	39.6484	39.6484	51.2939
	S3	38.9893	33.5938	33.5938	44.5313
Proposed86	S1	0.4883	0	0	0.4883
BER (%)	S2	1.5137	0	0	1.5137
	S3	1.4404	0.0244	0.0244	1.4404
Chen [3]	S1	25.7813	16.9434	16.9434	32.2998
BER (%)	S2	45.1172	37.3535	37.3535	45.1172
	S3	30.0293	15.3564	15.3564	38.9648
Hu [14]	S1	0	0	0	0
BER (%)	S2	0	0	0	0
	S3	0	0	0	0

than the corresponding existing techniques of [3, 9, 14, 29] in the case of AWGN attack. This can also be seen from Fig. 6, which illustrates the BER as a function of SNR resulting from an AWGN attack for the case of the signal S3. These results confirm the effectiveness of the proposed optimization technique introduced in Section 4 to guarantee the imperceptibility while maximizing the robustness under an AWGN attack. Moreover, in order to confirm the usefulness of the theoretical analysis performed in Section 3, we show in Fig. 7 the theoretical BER obtained using (56) and the BER obtained experimentally by the proposed technique versions Proposed172 and Proposed86. Specifically, this figure demonstrates the correctness of (56) derived for the theoretical BER. Table 8

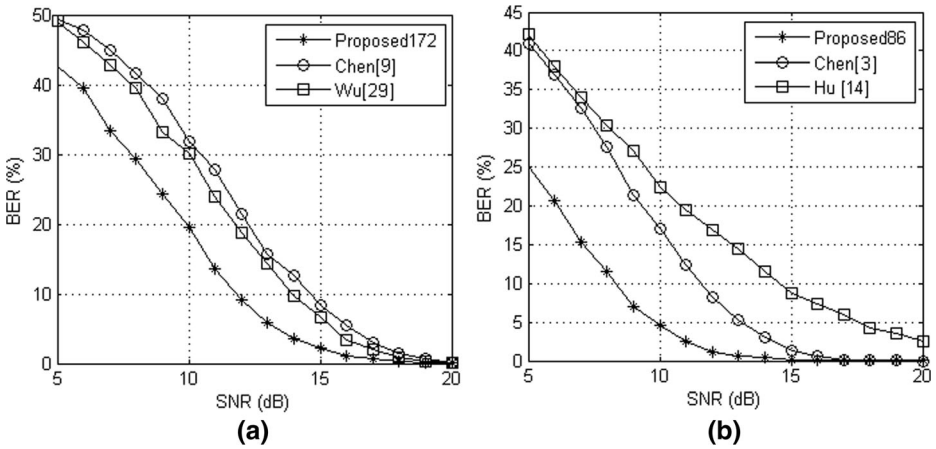

Fig. 6 BER of S3 under an AWGN attack, (a) techniques with a capacity of 172 bps, (b) techniques with a capacity of 86 bps

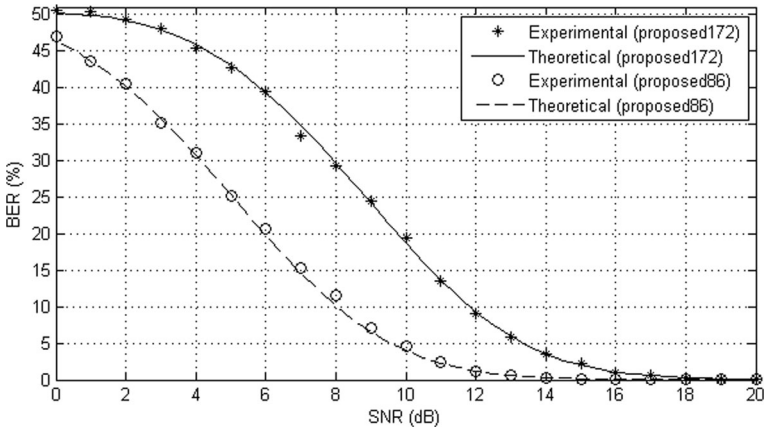

Fig. 7 Theoretical and experimental BER of the Proposed172 and Proposed86 versions under an AWGN attack

shows that the proposed technique versions are more robust to re-quantization attack than the corresponding techniques reported in [3, 9, 14, 29]. The re-quantization attack introduces a uniform noise in the time domain, which according to the central limit theorem converges to a Gaussian noise in the transform domain for sufficiantly large frame length. This is confirmed by the experimantal results shown in Fig. 8. Therfore, re-quantization in time domain is essentialy an AWGN in transform domain for which the robustness of the proposed technique is optimized, see Section 4. Table 9 confirms that the proposed technique outperforms the techniques reported in [3, 9, 14, 29] in the case of the amplitude scaling attack, except that the performance of the Proposed86 is almost equal to that of the technique in [14] even though the SWR obtained by [14] is less than the one obtained by the proposed technique. Moreover, the strong robustness of the technique of [14] is achieved in this attack due to the use of an adaptive quantization step invariant under scaling. However, this adaptive scheme has the drawback of low

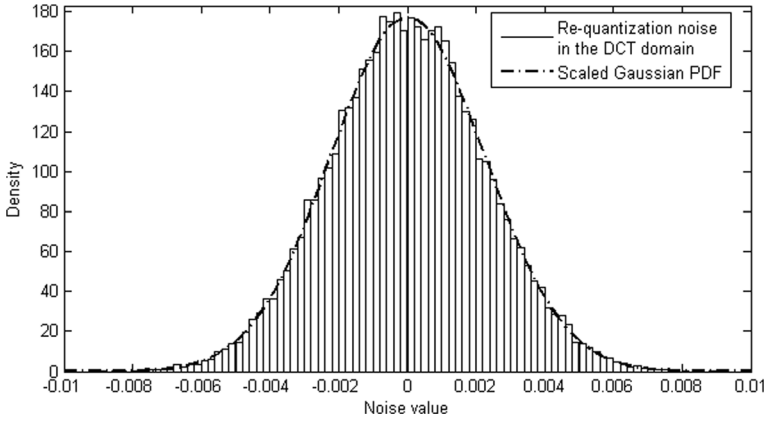

Fig. 8 Histogram of the re-quantization noise in the DCT domain fitted by a scaled Gaussian PDF for the frame size 256

robustness against noise attacks and substantially increases the computational complexity of the audio watermarking technique of [14] compared to that of the proposed technique.

8 Conclusion

In this paper, a new blind audio watermarking technique has been proposed in the transform domain by introducing a parametric QIM. We have shown that most of the existing QIM realizations are special cases of the proposed parametric QIM. In order to find the optimal values for the parameters of the proposed parametric QIM that ensure the imperceptibility while maximizing the robustness of the proposed watermarking technique under an AWGN attack, we have derived theoretical expressions for the signal to watermark ratio and probability of error and used them in an optimization technique based on the method of Lagrange multipliers. This optimization technique has led to a parametric QIM that is optimal for the proposed watermarking technique. Furthermore, in order to provide a good trade-off between the effects of high and low pass filtering attacks for the proposed watermarking technique, we have devised an efficient procedure to find the interval for the best selection of the watermark embedding positions in the DCT domain. In addition, to make the proposed watermarking technique more attractive, we have developed a very efficient algorithm for its fast implementation. Another interesting property of the proposed technique is the fact that its watermarking secret key can be constituted of the parameters of the optimal parametric QIM coupled with the embedding positions. All the experiments that we have carried out confirm the usefulness of the theoretical analysis presented in the paper and clearly show that the proposed technique outperforms the existing QIM-based audio watermarking techniques in most known attacks, specifically in AWGN, re-quantization, high pass filtering and common lossy compressions, and has performance similar to that of the existing QIM-based techniques in other attacks. Moreover, the proposed technique substantially reduces the computational complexity compared to the existing techniques. The proposed method does not provide excellent results in the case of amplitude scaling attack. Therefore, it is worth to investigate the maximization of the performance in this attack.

References

1. Al-Haj A (2014) A dual transform audio watermarking algorithm. Multimed Tools Appl 73:1897–1912. https://doi.org/10.1007/s11042-013-1645-z
2. Cheddad A, Condell J, Curran K, Mc Kevitt P (2010) Digital image steganography: survey and analysis of current methods. Signal Process 90:727–752. https://doi.org/10.1016/j.sigpro.2009.08.010
3. Chen ST, Huang HN (2016) Optimization-based audio watermarking with integrated quantization embedding. Multimed Tools Appl 75:4735–4751. https://doi.org/10.1007/s11042-015-2500-1
4. Chen B, Wornell GW (1998) Digital watermarking and information embedding using dither modulation. Multimed Signal Process 1998 I.E. Second Work 273–278. https://doi.org/10.1109/MMSP.1998.738946
5. Chen B, Wornell GW (1999) Provably robust digital watermarking. In: Tescher AG, Vasudev B, Bove, Jr. VM, Derryberry B (eds) Proceeding SPIE Multimed. Syst. Appl. II. pp 43–54
6. Chen B, Wornell GW (2000) Preprocessed and postprocessed quantization index modulation methods for digital watermarking. In: Wong PW, Delp III EJ (eds) SPIE - Secur. Watermarking Multimed. Contents II. pp 48–59
7. Chen B, Wornell GW (2001) Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inf Theory 47:1423–1443. https://doi.org/10.1109/18.923725

8. Chen OTC, Wu WC (2008) Highly robust, secure, and perceptual-quality echo hiding scheme. IEEE Trans Audio Speech Lang Process 16:629–638. https://doi.org/10.1109/TASL.2007.913022
9. Chen S-T, Wu G-D, Huang H-N (2010) Wavelet-domain audio watermarking scheme using optimisation-based quantisation. IET Signal Process 4:720. https://doi.org/10.1049/iet-spr.2009.0187
10. Clerc M, Kennedy J (2002) The particle swarm-explosion, stability, and convergence in a multidimensional complex space. IEEE Trans Evol Comput 6:58–73. https://doi.org/10.1109/4235.985692
11. Cox IJ, Kilian J, Leighton FT, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6:1673–1687. https://doi.org/10.1109/83.650120
12. Eberhart RC, Shi Y (2001) Particle swarm optimization: developments, applications and resources. Proc 2001 Congr Evol Comput (IEEE Cat No01TH8546) 1:81–86. https://doi.org/10.1109/CEC.2001.934374
13. Hemis M, Boudraa B, Merazi-meksen T (2015) Intelligent audio watermarking algorithm using multi-objective particle swarm optimization. In: 2015 4th Int. Conf. Electr. Eng. IEEE, pp 0–4
14. Hu HT, Hsu LY (2015) Robust, transparent and high-capacity audio watermarking in DCT domain. Signal Process 109:226–235. https://doi.org/10.1016/j.sigpro.2014.11.011
15. Hua G, Huang J, Shi YQ et al (2016) Twenty years of digital audio watermarking—a comprehensive review. Signal Process 128:222–242. https://doi.org/10.1016/j.sigpro.2016.04.005
16. Huang H, Chen S, Hsu C (2015) Wavelet-domain audio watermarking using optimal modification on low-frequency amplitude. IET Signal Process 9:166–176. https://doi.org/10.1049/iet-spr.2013.0399
17. Kalantari NK, Ahadi SM (2010) A logarithmic quantization index modulation for perceptually better data hiding. IEEE Trans Image Process 19:1504–1517. https://doi.org/10.1109/TIP.2010.2042646
18. Lei BY, Soon IY, Li Z (2011) Blind and robust audio watermarking scheme based on SVDDCT. Signal Process 91:1973–1984. https://doi.org/10.1016/j.sigpro.2011.03.001
19. Li Q, Cox IJ (2007) Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking. IEEE Trans Inf Forensics Secur 2:127–138. https://doi.org/10.1109/TIFS.2007.897266
20. Li Z, Sun Q, Lian Y (2006) Design and analysis of a scalable watermarking scheme for the scalable audio coder. IEEE Trans Signal Process 54:3064–3077. https://doi.org/10.1109/TSP.2006.875393
21. Lie WN, Chang LC (2006) Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. IEEE Trans Multimed 8:46–59. https://doi.org/10.1109/TMM.2005.861292
22. Liu N, Amin P, Ambalavanan A, Subbalakshmi KP (2006) An overview of digital watermarking. In: Multimed. Secur. Technol. Digit. Rights Manag. Elsevier, pp 167–195
23. Marini F, Walczak B (2015) Particle swarm optimization (PSO). A tutorial. Chemom Intell Lab Syst 149: 153–165. https://doi.org/10.1016/j.chemolab.2015.08.020
24. Puschel M (2003) Cooley-Tukey FFT like algorithms for the DCT. In: 2003 I.E. Int. Conf. Acoust. Speech, Signal Process. 2003. Proceedings. (ICASSP '03). IEEE, p II-501-4
25. Renza D, Ballesteros DM, Ortiz HD (2016) Text hiding in images based on QIM and OVSF. IEEE Lat Am Trans 14:1206–1212. https://doi.org/10.1109/TLA.2016.7459600
26. Tim J, Juergen S (2005) Digital watermarking and its impact on intellectual property limitation for the digital age. J Electron Commer Organ 3:72–82. https://doi.org/10.4018/jeco.2005010105
27. Wah B, Wang T (1999) Efficient and adaptive lagrange-multiplier optimization. J Glob Optim:1–25. https://doi.org/10.1023/A.1008203422124
28. Wang YG, Zhu G (2015) An improved AQIM watermarking method with minimum-distortion angle quantization and amplitude projection strategy. Inf Sci (Ny) 316:40–53. https://doi.org/10.1016/j.ins.2015.04.029
29. Wu S, Huang J, Huang D, Shi YQ (2005) Efficiently self-synchronized audio watermarking for assured audio data transmission. IEEE Trans Broadcast 51:69–76. https://doi.org/10.1109/TBC.2004.838265
30. Xiang S, Kim HJ, Huang J (2008) Audio watermarking robust against time-scale modification and MP3 compression. Signal Process 88:2372–2387. https://doi.org/10.1016/j.sigpro.2008.03.019
31. Xiang Y, Peng D, Natgunanathan I, Zhou W (2011) Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking. IEEE Trans Multimed 13:2–13. https://doi.org/10.1109/TMM.2010.2080668
32. Xiang Y, Natgunanathan I, Peng D et al (2012) A dual-channel time-spread echo method for audio watermarking. IEEE Trans Inf Forensics Secur 7:383–392. https://doi.org/10.1109/TIFS.2011.2173678
33. Yubao Bai Y, Sen Bai S, Guibin Zhu G, et al (2010) A blind audio watermarking algorithm based on FFT coefficients quantization. In: 2010 Int. Conf. Artif. Intell. Educ. IEEE, pp 529–533
34. Zareian M, Tohidypour HR (2013) Robust quantisation index modulation-based approach for image watermarking. Image Process IET 7:432–441. https://doi.org/10.1049/iet-ipr.2013.0048

Younes Terchi received the Bachelor degree in electronics (communication engineering), from University Ferhat Abbas Setif 1, Setif, Algeria, in 2010, the master degree in electronics (telecommunication systems and networks), from University Ferhat Abbas Setif 1, Setif, Algeria, in 2012. He is currently pursuing the Ph.D. degree in the area of signal processing at the Department of Electronics, University Ferhat Abbas Setif 1, Setif, Algeria. His research interests include multimedia security, digital watermarking, speech and image processing.

Saad Bouguezel received the Engineering degree in electronics (communication engineering) from University Ferhat Abbas Setif 1, Setif, Algeria, in 1993, the Masters' degree in electronics (industrial control) from Batna University, Batna, Algeria, in 1998, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 2004. From November 2004 to June 2006, he was a Postdoctoral Fellow in the Center for Signal Processing and Communications, Department of Electrical and Computer Engineering, Concordia University, where he was a Research and Teaching Assistant from March 2002 to October 2004. From 1998 to February 2002, he was a Lecturer within the Department of Electrical Engineering, Laghouat University, Laghouat, Algeria. In 2006, he joined the Department of Electronics at University Ferhat Abbas Setif 1, Algeria, where he is currently a Professor. His research interests include discrete transforms and their fast algorithms, techniques for signal and image processing, compression, encryption and watermarking.