PAPER

Logarithmic Adaptive Quantization Projection for Audio Watermarking

Xuemin ZHAO[†a)], **Yuhong GUO**[†], **Jian LIU**[†], **Yonghong YAN**[†], ***Nonmembers*, *and*** **Qiang FU**[†], ***Member***

SUMMARY In this paper, a logarithmic adaptive quantization projection (LAQP) algorithm for digital watermarking is proposed. Conventional quantization index modulation uses a fixed quantization step in the watermarking embedding procedure, which leads to poor fidelity. Moreover, the conventional methods are sensitive to value-metric scaling attack. The LAQP method combines the quantization projection scheme with a perceptual model. In comparison to some conventional quantization methods with a perceptual model, the LAQP only needs to calculate the perceptual model in the embedding procedure, avoiding the decoding errors introduced by the difference of the perceptual model used in the embedding and decoding procedure. Experimental results show that the proposed watermarking scheme keeps a better fidelity and is robust against the common signal processing attack. More importantly, the proposed scheme is invariant to value-metric scaling attack.

key words: *audio watermarking, quantization index modulation (QIM), psychoacoustic model*

1. Introduction

The digital watermarking reported provides a new method to enforce the copyright protection of multimedia data, which can be widely used in digital rights management, broadcast monitoring, and tracking illegal copies of the multimedia data. It involves a process of embedding a watermarking into a host signal with a perceptually transparent digital signature, carrying a message which the owner wants to transmit. It is now recognized that watermarking can be modeled as communication with side information [1]. In the communication channel, two noise sources exist. The first noise source is the interference of the host signal, which is entirely known to the transmitter. The second noise source is subsequent distortion between transmitter and receiver, which is unknown to both transmitter and receiver. According to the analysis by Costa [2], the first noise source which called side information need not interfere with the embedded watermark.

Based on Costa's work, the way to apply the practical implementation of this framework was first established by Cox *et al.* [1] and further developed by [3]–[5]. These methods can be grouped into a class of watermarking methods called host-interference rejecting methods, while the remaining watermarking methods can be grouped into host-interference non-rejecting methods, such as the spread spectrum methods [6]. The main advantage of host-interference rejecting methods is the higher capacity due to the absence of host signal's interfere. Among these watermarking techniques, the quantization index modulation (QIM) [3] method proposed by Chen and Wornell attracted more attention. It uses a structure lattice quantizers to quantize the host signal with a quantizer chosen by input message, which provides a computationally efficient watermarking algorithm with high data capacity.

The main weakness of lattice based QIM watermarking is its sensitivity to gain attack. Even small changes in the scale of watermarked signal can result in dramatic performance degradation [7]. The solutions proposed so far can be grouped into three categories. The first is to estimate the amplitude scaling factor by embedding an auxiliary pilot signal [8] or blind estimation methods [9]. These methods' performance depends on the accuracy of the scaling factor estimated. The second is the adoption of spherical codeword [10], together with a correlation decoder [11]. The problem with the spherical codes is that watermark embedding and recovery get very complicated, thus losing the simplicity of lattice based watermarking. The third approach is to define an embedding domain that is invariant to value-metric scaling, such as rational dither modulation (RDM) [7].

The conventional QIM scheme uses a fixed quantization step which does not exploit the perceptual character of the host signal to get better fidelity. To improve the fidelity of the QIM scheme, Perez-Gonzalez *et al.* developed an algorithm called quantization projection (QP) [12] which combines the QIM methods with a perceptual model. But it has a problem that the perceptual model should be calculated in both embedding and decoding procedures. The perceptual model used in the decoding procedure based on the distorted watermarked signal. There may be some differences in the perceptual model used in the embedding procedure, which will influence the system's performance.

In this paper, a logarithmic adaptive quantization projection (LAQP) method will be proposed. The watermark is adapting with the perceptual model which is only calculated in embedding procedure. At the same time, the watermarking algorithm is invariant to value-metric scaling attack. This paper is organized as follows. Sections 2 and 3 introduce the basic concepts of the QIM and QP methods, with some definitions described. Section 4 presents the algorithm of LAQP and its properties. A practical audio watermarking system will be described in Sect. 5. The ex-

Manuscript received September 8, 2011.
Manuscript revised January 3, 2012.

[†]The authors are with Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences, 21 Beisihuan XiLu, Beijing, China.
a) E-mail: zhaoxuemin@hccl.ioa.ac.cn

DOI: 10.1587/transinf.E95.D.1436

periment results and some discussions are given in Sect. 6, and conclusions are presented in Sect. 7.

2. Definition and Theoretical Framework

Watermarking with side information can be modeled by the communication system depicted in Fig. 1. The watermark message m is embedded into host signal x to produce the watermarked signal s. The difference $w = s - x$ is denoted the watermark signal. The watermarked signal then undergoes a number of distortions that are modeled as an unknown noise source n. The watermark detector receives a distorted watermarked signal $y = s + n$ and extracts the estimated message $\hat{m}$.

In 1999, Chen and Wornell introduced a class of data-hiding codes known as QIM schemes. The watermarking is achieved through a set of predefined quantizers to quantize the host signal. The quantizer used by the encoder depends on the watermark message m. The key to QIM is the design of the codebooks of the quantizers used to embed the watermark. The simplest solution is to adopt a set of scalar and uniform quantizers namely the Dither Modulation (DM) scheme. Specifically, the two codebooks $\mathcal{U}_0$ and $\mathcal{U}_1$ used in DM associated respectively to $m = 0$ and $m = 1$ are defined as

$$\mathcal{U}_0 = \{k\Delta + d, k \in \mathbb{Z}\} \tag{1}$$

$$\mathcal{U}_1 = \left\{k\Delta + \frac{\Delta}{2} + d, k \in \mathbb{Z}\right\} \tag{2}$$

where Δ is the quantization step and $\mathbb{Z}$ represents integer, d is an arbitrary parameter, possibly depending on a secret key to improve security. In the following, we assume $d = \Delta/4$, since in this way a lower distortion is obtained. Watermark embedding is achieved by applying either the quantizer Q_0 associated to $\mathcal{U}_0$

$$Q_0(x) = \arg\min_{u_0 \in \mathcal{U}_0} |u_0 - x|, \tag{3}$$

where u_0 are the elements of $\mathcal{U}_0$, or the quantizer Q_1 associated to $\mathcal{U}_1$

$$Q_1(x) = \arg\min_{u_1 \in \mathcal{U}_1} |u_1 - x|. \tag{4}$$

The embedding function embeds message m into host signal x to form the watermark signal s by

$$s = \begin{cases} Q_0(x), & m = 0 \\ Q_1(x), & m = 1 \end{cases}. \tag{5}$$

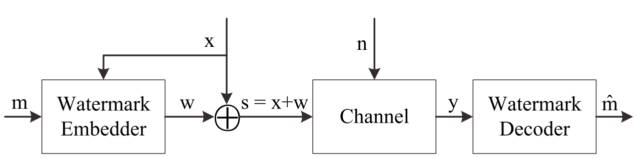

Fig. 1 The communication system model of watermarking with side information.

As to decoding, the distorted signal y is received. A minimum-distance decoder is adopted to estimate the message by

$$\hat{m} = \arg\min_{m \in \{0,1\}} \min_{u_m \in \mathcal{U}_m} |u_m - y|. \tag{6}$$

The maximum quantization error introduced by the embedding is $\Delta/2$. The embedding distortion also can be measured by squared-error distortion

$$D(s, x) = \|s - x\|^2 = \|w\|^2 \tag{7}$$

or its expectation $D_w = E[D(s, x)]$. According to analysis by [13], if the quantization error is uniformly distributed over $[-\Delta/2, \Delta/2]$, the embedding distortion is $D_w = \Delta^2/12$.

The distortion between watermarked signal s and received watermarked signal y is

$$D(y, s) = \|y - s\|^2 = \|n\|^2 \tag{8}$$

and its expectation can be defined as $D_n = E[D(y, s)]$. Here some quantities describing the relationship of the power of host signal, watermark and noise are defined. The Document-to-Watermark Ratio (DWR) is given by σ_x^2/D_w, σ_x^2 is the variance of host signal; the Watermark-to-Noise Ratio (WNR) is D_w/D_n. These quantities are expressed in decibels.

3. Conventional Quantization Projection Approach

In paper [12], Perez-Gonzalez *et al.* analyzed the performance of existing quantization algorithms and proposed the quantization projection (QP) method which couple the effectiveness of QIM scheme and spread spectrum methods. In the basic QP case, the projection function computes a weighted cross-correlation between the length-L watermarked signal $\boldsymbol{y}$ and projection vector $\boldsymbol{b}$; therefore, for a single transmitted bit, the projection r is such that

$$r = \sum_{i=1}^{L} \frac{y_i b_i}{a_i}. \tag{9}$$

In a second stage, a minimum-distance decoder is adopted, and the estimated message can be decoded by

$$\hat{m} = \arg\min_{m \in \{0,1\}} \min_{u_m \in \mathcal{U}_m} |u_m - r|. \tag{10}$$

In Eq. (9), the noise is assumed to be perceptually shaped, i.e., its variance is proportional to the perceptual mask a_i. If w_i denotes the i sample of the watermark and rewrite Eq. (9) as

$$r = r_x + r_w \tag{11}$$

where r_w is projected watermark

$$r_w = \sum_{i=1}^{L} \frac{w_i b_i}{a_i} \tag{12}$$

and with a similar definition for the projected host signal r_x

$$r_x = \sum_{i=1}^{L} \frac{x_i b_i}{a_i}. \tag{13}$$

The problem of the encoder is to select the watermark samples w_i so that (12) is satisfied. This problem resembles the so-called knapsack problem (an NP-complete one) which has infinitely many solutions. For the watermarking purpose, w_i is chosen to be proportional to a_i. It can be shown that under the perceptual constraints, this choice minimizes the probability of error [12]. Then

$$w_i = \rho a_i b_i \tag{14}$$

where ρ is a real number that can be determined by substituting (14) to (12). It can be gotten that $\rho = r_w/L$, and finally

$$w_i = \frac{r_w a_i b_i}{L}. \tag{15}$$

In Eq. (15), the projected watermark r_w can be gotten by employing the normal QIM embedding equation (5) to the projected host signal r_x.

4. Logarithmic Adaptive Quantization Projection

The perceptual model is used in both the embedding and decoding procedures of the QP methods. It implies that the perceptual weighted coefficients should be known to both the encoder and decoder to get the expected result. However, it is not practical in most applications. The perceptual model could be calculated in the embedding and decoding procedure separately. But the perceptual model used in the decoding procedure based on the distorted watermarked signal. There may be some differences in the perceptual model used in the embedding procedure. In fact, the perceptual model may be different due to the changes introduced in the embedding procedure. These problems will be avoided in our new proposed method, while the quantization step for each sample is still adaptive to the perceptual model, just like the conventional quantized projection method. But the quantization step for each sample does not need to be known or estimated during detection. At the same time, the watermark is embedded into the logarithm domain. The watermark will be invariant to the value-metric scaling when a balanced pseudo-random sequence is used, which will be presented later.

4.1 Proposed Method

The proposed Logarithmic Adaptive Quantization Projection (LAQP) method is based on the quantization of the original host signal in the logarithm domain. Firstly, the encoder calculates the correlation between $\boldsymbol{u} = (u_1, u_2 \cdots u_L) = (\ln(|x_1|), \ln(|x_2|) \cdots \ln(|x_L|))$ and projection vector $\boldsymbol{b}$

$$p_u = \boldsymbol{u} \cdot \boldsymbol{b} = \sum_{i=1}^{L} \ln(|x_i|) b_i. \tag{16}$$

The projected watermarked variable $p_u^Q = p_u + p_v$ can be defined as

$$p_u^Q = p_u + p_v = \begin{cases} Q_0(p_u), & m = 0 \\ Q_1(p_u), & m = 1 \end{cases} \tag{17}$$

where p_v is the projected watermark variable in the logarithm domain.

Let $\boldsymbol{u^Q} = (u_1^Q, u_2^Q \cdots u_L^Q) = (\ln(|s_1|), \ln(|s_2|) \cdots \ln(|s_L|))$ be the watermarked samples in the logarithm domain, $\boldsymbol{v} = (v_1, v_2 \cdots v_L) = \boldsymbol{u^Q} - \boldsymbol{u}$ the watermark samples in the logarithm domain. As the solution in the QP method, the watermark sample in the logarithm domain v_i is chosen to be perceptually weighted by the respective masks a_i, i.e., the watermark sample v_i is proportional to a_i, then

$$v_i = \rho a_i b_i \tag{18}$$

with ρ to be determined later. Because the projected watermark signal $p_v = \sum_{i=1}^{L} v_i b_i = \sum_{i=1}^{L} \rho a_i b_i^2$, when the b_i is the binary pseudo-random sequence, i.e. $b_i \in \{\pm 1\}$, it can be gotten that $p_v = \rho \sum_{i=1}^{L} a_i$. The ρ can be calculated by

$$\rho = \frac{p_v}{\sum_{i=1}^{L} a_i}. \tag{19}$$

Substitute (19) into (18), the watermark sample in the logarithm domain v_i is given by

$$v_i = p_v \frac{a_i}{\sum_{i=1}^{L} a_i} b_i. \tag{20}$$

Finally, the resulting watermark sample s_i is

$$\begin{aligned} s_i &= \operatorname{sign}(x_i) \cdot \exp(u_i^Q) \\ &= \operatorname{sign}(x_i) \cdot \exp(u_i + v_i) \\ &= x_i \cdot \exp(v_i) \\ &= x_i \cdot \exp\left(p_v \frac{a_i}{\sum_{i=1}^{L} a_i} b_i \right). \end{aligned} \tag{21}$$

In Eq. (21), the function $\exp(\cdot)$ is the exponential function.

At the decoding stage, let $\boldsymbol{u'} = (u'_1, u'_2 \cdots u'_L) = (\ln(|y_1|), \ln(|y_2|) \cdots \ln(|y_L|))$ be the received watermarked samples in the logarithm domain, and their projected feature be

$$p'_u = \boldsymbol{u'} \cdot \boldsymbol{b} = \sum_{i=1}^{L} \ln(|y_i|) b_i. \tag{22}$$

The estimated message can be decoded by

$$\hat{m} = \arg\min_{m \in \{0,1\}} \min_{u_m \in \mathcal{U}_m} |u_m - p'_u|. \tag{23}$$

Note that the perceptual model does not need to be calculated at the decoding stage. The estimated message $\hat{m}$ can be got from the received watermarked signal directly.

It can be seen that both the conventional and the proposed LAQP method use a projection variable to transmit a bit by the quantization scheme. For the conventional QP method, the projection variable to be quantized is gotten from dividing by the perceptual weighted coefficient. At the decoding stage, the variable should be calculated in the same manner as in the embedding procedure. For the proposed LAQP method, when the projection watermark variable is calculated, the embedding procedure allots a part of the projection watermark to each sample according to the perceptual weighted coefficients. The watermark is shaped by the perceptual model in this way. While the resulting projection watermarked variable can be directly gotten by correlation between the watermarked sample in logarithm domain and reference projecting vector.

4.2 The Analysis of LAQP

Some features of LAQP method will be discussed in this section, including the embedding distortion and the invariance to the value-metric scaling attack of LAQP method.

4.2.1 The Embedding Distortion of LAQP

The squared-error distortion $D(s,x) = \|s-x\|^2 = \|w\|^2$ is still used to measure the embedding distortion here. From the embedding equation (21), the distortion of each watermarked sample is

$$D(s_i,x_i) = \|s_i - x_i\|^2 = \|x_i(\exp(v_i)-1)\|^2. \tag{24}$$

If the $v \ll 1$ it is reasonable to approximate $\exp(v)-1 \approx v$, so whenever the condition is met (this is the case in most practical applications) the embedding distortion is

$$\begin{aligned} D(s_i,x_i) &= \|x_i(\exp(v)-1)\|^2 \\ &\approx \|x_i v_i\|^2 \\ &= \left\| x_i p_v \frac{a_i}{\sum_{i=1}^L a_i} \right\|^2. \end{aligned} \tag{25}$$

In Eq. (25), p_v is the projected watermark signal in the logarithm domain, which is gotten by quantization the projected host signal in the logarithm domain p_u. It can be considered as the watermark signal of the p_u by the use of the traditional QIM scheme. As the analysis of [13], when the quantization step is Δ, p_v is independent of p_u when the dither modulation [3] is used in QIM scheme, and it is uniformly distributed over $[-\Delta/2, \Delta/2]$. Because p_u is directly gotten from the host signal $\boldsymbol{x}$, p_v can be assumed independent of the host sample x_i. Based on the analysis above, the expectation of distortion D_{w_i} of each watermarked sample is

$$\begin{aligned} D_{w_i} &= E[D(s,x)] \\ &\approx \int_{-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2}\left(x_i p_v \frac{a_i}{\sum_{i=1}^L a_i}\right)^2 f_X(x_i) f_V(p_v) dx_i dp_v \\ &= \frac{a_i^2}{\left(\sum_{i=1}^L a_i\right)^2}\int_{-\infty}^{\infty} x_i^2 f_X(x_i)dx_i \int_{-\Delta/2}^{\Delta/2} p_v^2 f_V(p_v)dp_v \\ &= \frac{a_i^2}{\left(\sum_{i=1}^L a_i\right)^2}\sigma_{x_i}^2\frac{\Delta^2}{12} \end{aligned} \tag{26}$$

for any distribution of the host signal.

From Eq. (26), each dimension of embedding introduced distortion D_{w_i} to the host vector $\boldsymbol{x}$ is proportional to the host's power of each dimension $\sigma_{x_i}^2$, square of perceptual weighted coefficient a_i^2 and the square of the quantization step Δ^2, while $\left(\sum_{i=1}^L a_i\right)^2$ is a constant in each dimension. The embedding distortion is used for the perceptual model. The embedding distortion in each dimension should not exceed the masking level and be controlled by the perceptual model. A practical use of the embedding distortion for MPEG psychoacoustic model will be discussed later.

4.2.2 The Invariance to the Value-Metric Scaling Attack of LAQP

The conventional QIM and QP methods both have a disadvantage of sensitivity to the value-metric scaling attack. Even small changes in the scale of watermarked signal can result in dramatic increase in the bit error rate (BER). If the scaling factor ν is unknown to the decoder, using the minimum-distance decoder to estimate the message from the scaling version of $\boldsymbol{y}' = \nu\boldsymbol{y}$ directly may lead to bit error.

Embedding watermark in the logarithm domain has an advantage that the watermark is invariant to the value-metric scaling. Recall the decoding equation (23), let $\boldsymbol{y}' = \nu\boldsymbol{y}$ be the scaling version of $\boldsymbol{y}$, the projected received watermarked variable p_u' is

$$\begin{aligned} p_u' &= \sum_{i=1}^L \ln|y_i'|b_i \\ &= \sum_{i=1}^L \ln|\nu y_i|b_i \\ &= \sum_{i=1}^L (\ln|y_i| + \ln|\nu|)b_i \\ &= \sum_{i=1}^L \ln|y_i|b_i + \ln|\nu|\sum_{i=1}^L b_i. \end{aligned} \tag{27}$$

Compare Eqs. (22) and (27), the projected received watermarked feature p_u' does not change when the $\sum_{i=1}^L b_i = 0$. That implies that the number of $+1$ is equal to the number of -1 in projected binary pseudo-random sequence $\boldsymbol{b}$, i.e., the projected binary pseudo-random sequence must be balanced. Specifically, the Gold code sequences [14] can be used in the proposed method, the Gold code sequences are a quasi-orthogonal sequences widely used in the spread spectrum applications, which also satisfies the balanced condition. Here an embedding domain invariant to value-metric scaling attack is designed by utilizing the binary balanced pseudo-random sequence and embedding the watermark in logarithm domain.

5. An Audio Watermarking System Based on LAQP Method

In audio signal processing, the input original audio is segmented into overlapped frames. Then these samples are multiplied by a window function such as Hanning window and converted into frequency domain by FFT. After addition of the watermark, passing these modified frames to the inverse transform, then overlap-and-add between consecutive frames to generate the time-domain marked audio signal. If the transform coefficients to add the watermark are changed in the overlapped frame, the added watermark will become invalid because of the overlap-and-add structure and the non-linearity of QIM. And the overlap-and-add structure cannot be abandoned and modified the coefficients directly because of the edge effects across blocks.

The PQMF (Pseudo Quadrature Mirror Filter) filter banks are chosen to embed the non-linear QIM watermark [15]. The PQMF filter banks are cosine-modulated, critically-sampled, polyphase filter banks which are widely used in MPEG audio coding standards. The MPEG uses this filter banks to time-to-frequency mapping. The modified sub-band samples can be exactly recovered after the synthesis and analysis processing. The watermarked samples are smoothed and do not have discontinuous noise between the frames, because the procedure of synthesis can be regarded as continuous filtering by the synthesis filter. The sub-band samples have a certain frequency resolution. The psychoacoustic auditory model can be used to improve the watermark's fidelity.

5.1 The Procedures of Encoding and Decoding

The proposed system has two modules which correspond to the encoding and decoding procedure respectively. As Fig. 2 shows, the procedure of encoding watermark into audio signal is as follows.

- A watermark message is encoded into information bits m by error correction codes.
- The input PCM samples are decomposed into 32 equal width sub-bands samples by applying 32-bands PQMF.
- At the same time, the psychoacoustic analysis stage is performed with the same PCM samples to determine the perceptual weighted coefficient a_i for each sub-band.
- Then the length of L sub-bands samples are selected to obtain the host vector $\boldsymbol{x}$, to calculate the projected host signal in the logarithm domain p_u.
- Use Eq. (17) to quantize p_u with pre-defined step Δ to get the projected watermark signal p_v. The watermarked samples can be got from p_v with corresponding a_i and b_i in its dimension by applying Eq. (21).
- Finally, the watermarked samples are sent to PQMF synthesis filter banks to rebuild the time-domain signal.

The main difference between the encoding and decoding procedure is that the psychoacoustic analysis does not need to be performed in the decoding stage. The procedure of decoding is also shown in Fig. 2.

- At the decoding stage, similar to the encoding procedure, the received watermarked signals are decomposed into 32 equal width sub-bands samples by applying the same 32-bands PQMF analysis filter.
- Calculate the projected received watermarked feature p'_u from the decomposed sub-band samples. Then use Eq. (23) with p'_u to get the information bits $\hat{m}$.
- At the same time, the psychoacoustic analysis stage is performed with the same PCM samples to determine the perceptual weighted coefficient a_i for each sub-band.
- Then the length of L sub-bands samples are selected to obtain the host vector $\boldsymbol{x}$, to calculate the projected host signal in the logarithm domain p_u.
- Finally the estimated watermark message is recovered from information bits $\hat{m}$ by the error correction decoder.

5.2 The Use of Psychoacoustic Model

The proposed system uses the psychoacoustic model defined in the ISO-MPEG Audio Psychoacoustic Model [16]. The

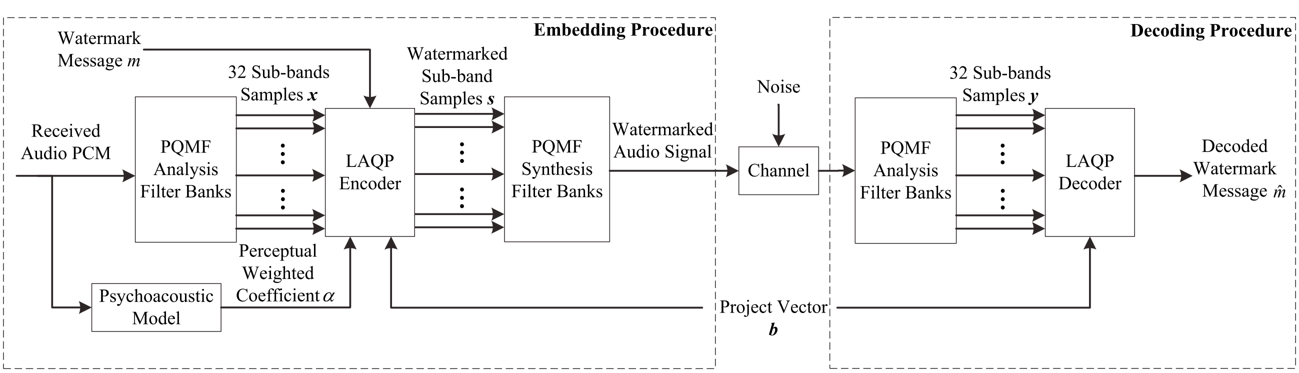

Fig. 2 The procedure of encoding and decoding.

detailed procedure to obtain the psychoacoustic model is described in [16]. In MPEG compression, the psychoacoustic model computes the signal to mask ratios (SMR) for each sub-band. The procedure of bit allocation is based on SMR. The resulting perceptual bit allocation result R_i^{opt} for each sub-band i is [17]

$$R_i^{opt} = R + \frac{\ln 10}{20 \ln 2}\left(\mathrm{SMR}(i) - \sum_{b=0}^{N-1} \mathrm{SMR}(b)\right) \tag{28}$$

where R is the average bits allocated for 32 subbands. If a sub-band is characterized by a high SMR, it will get more bits in bit allocation. This means putting smaller quantization noise or watermark in subband which is more audible. For the subband with lower SMR, the distortion of watermark in this subband will be slighter because it is less audible. According to the quantization theory [17], the quantization noise ε^2 in each subband is

$$\varepsilon^2 = \frac{1}{3}\frac{\sigma_x^2}{2^{2R^{opt}}} \tag{29}$$

where the σ_x^2 is the variance of the signal x to be watermarked.

According to the analysis above, the embedding distortion D_{w_i} of each subband is given by Eq. (26). Let $D_{w_i} = \varepsilon_i^2$ in each subband. The embedded watermark will be shaped by the same perceptual model used in MPEG compression. Substitute (28) into (29), the quantization noise ε_i^2 is

$$\varepsilon_i^2 = \frac{1}{3}\frac{\sigma_x^2}{2^{2R}} \cdot 10^{-\frac{\mathrm{SMR}(i) - \sum_{b=0}^{N-1}\mathrm{SMR}(b)}{10}}. \tag{30}$$

In Eq. (30), R and $\sum_{b=0}^{N-1}\mathrm{SMR}(b)$ are constant in all subbands. The quantization noise ε_i^2 is decided by the host's power of each subband σ_x^2 and each subband's SMR. Comparing Eqs. (26) and (30), and let $D_{w_i} = \varepsilon_i^2$, we will find that

$$\frac{a_i}{\sum_{i=1}^{L} a_i} = \frac{1}{2^{R-1} \cdot \Delta} \cdot 10^{-\frac{\mathrm{SMR}(i) - \sum_{b=0}^{N-1}\mathrm{SMR}(b)}{20}} \tag{31}$$

and a_i is proportional to $10^{-\frac{\mathrm{SMR}(i)}{20}}$. Other terms in the equation are equal in all the subbands. Assuming

$$a_i = K \cdot 10^{-\frac{\mathrm{SMR}(i)}{20}} \tag{32}$$

and substitute it to Eq. (26), it can be deduced that $D_{w_i} = \varepsilon_i^2$ whenever $2^{R-1} \cdot \Delta = 1$. In fact, the factor K in Eq. (32) can be any non-zero value, as it will be eliminated in the distortion equation (26).

As the SMR is used for controlling the subband's bit allocation in the MPEG compression, the perceptual weighted coefficient a_i is used for allocating the watermark in each subband. The higher a_i expects more watermark allocated in one sub-band, which corresponds to the lower SMR in MPEG compression, it will be allocated less bits for this subband and expect more quantization noise. The watermark's embedding distortion will have the same perceptual characteristics as the quantization noise caused by MPEG compression in this way.

6. Experimental Results and Discussions

The major measures to evaluate a watermarking system are usually fidelity and robustness, while these two requirements are conflicting. And there has to be a tradeoff between them. To further evaluate the watermarking system's performance, the audio quality test and robustness test were illustrated for proposed watermarking scheme. We applied various attacks provided in Stirmark Benchmark for Audio (SMBA) [18] and compare the experimental result of the proposed method with other watermarking methods. In our experiments, six types of audio materials were prepared for the experiments. Each type has three audio segments with a duration of several minutes. We chose about 10–20 seconds audio in each type to carry on listening test. All of the materials prepared are used in the robustness tests. The audio signals used in the experiments are sampled at 48 kHz with a 16 bits/sample. In our experiment, the DWR were restricted to the case of DWR = 25 dB.

6.1 Audio Quality Test

To further evaluate the watermarked audio quality, an subjective audio quality test was carried out. The subjective test is based on the standard for subjective evaluations of small impairments of high quality perceptual audio codec's which is specified in the ITUR BS.1116 [19]. The output of the listening tests are based on the so-called subjective difference grade (SDG) shown in the right column of Table 1. Transparency is assumed if the SDG value is 0 whereas a value of −4.0 is very annoying.

At the same time, an objective evaluation was carried on through an algorithm for objective measurement called perceived audio quality (PEAQ) [20]. The output of this algorithm is the objective difference grade (ODG), which describes the audibility of the introduced distortions like SDG used in subjective listening test, which is also shown in the right column of Table 1.

To test and verify the effectiveness of the psychoacoustic model in the watermarking system. We compared the results of our system and QIM system without the psychoacoustic model. The DWR of these two system are adjusted to the same level for comparison purposes. 27 listeners took part in the subjective listening test. All of the listeners have experience in music or audio engineering. The data thus obtained are subjected to a t-test to evaluate the listener's expertise. 16 listeners passed the test and their data were retained for find statistics analysis. The results of the listening

Table 1 ITU-R five-grade impairment scale.

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

Table 2 The SDG score of the subjective listening tests.

Material Type	QIM	LAQP
Classical	−0.26	−0.10
Voice	−2.11	−0.75
Violin	−0.75	−0.71
Opera	−1.06	−1.05
Flute	−1.15	−1.01
Piano	−1.29	−0.97

Table 3 The ODG score of the PEAQ tests.

Material Type	QIM	LAQP
Classical	−0.93	−0.84
Voice	−1.23	−0.44
Violin	−0.69	−0.59
Opera	−1.04	−0.73
Flute	−0.79	−0.76
Piano	−1.19	−0.86

tests and PEAQ test are shown in Tables 2 and 3 respectively.

In Tables 2 and 3, the QIM represents the conventional QIM system with the fixed quantization step, the LAQP represents the our system. Both the SDG and ODG scores of our LAQP system are higher than those with conventional QIM for all types of materials. The results show that the use of psychoacoustic model improves the audio's fidelity after watermark embedding.

6.2 Robustness Test

To evaluate the robustness of the proposed watermarking algorithm, several types of attack evaluations were conducted including amplitude scaling, adding white Gaussian noise, and MPEG compression. The QIM scheme and RDM scheme [7] with the same DWR participated in the robustness test for comparison, the order of RDM in our experiment is 300. After processing, the Bit Error Rates (BER) performance of watermark was calculated and shown in our test result. The BER defined as

$$\mathrm{BER} = \frac{\text{number of error bits}}{\text{number of total bits}} \times 100\%. \tag{33}$$

The detailed robustness test procedure is as follows.

• Value-Metric Scaling

The robustness to value-metric scaling for the proposed scheme, RDM scheme and the QIM scheme are shown in Fig. 3. The scaling factor was changed from 0.2 to 1.2. From the experimental results we observe that the BER has very small change in different scaling factor in the proposed scheme. The BER in RDM scheme also has very small change, while BER in QIM scheme results in a dramatic increase with changes in the scale of watermarked signal. The results prove that proposed algorithm is less sensitive to value-metric scaling. Both our proposed scheme and RDM scheme are designed to be invariant to value-metric scaling, but we can observe that the small BER change in the experiment. In fact, the PQMF filter banks are not perfect reconstruction filter banks. The noise in reconstruction will have

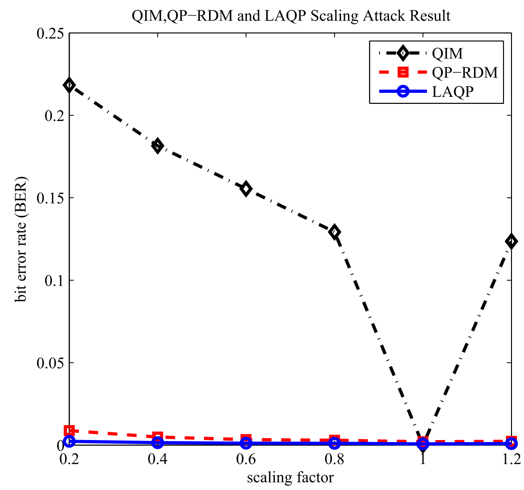

Fig. 3 The robustness test result of value-metric scaling attack.

Table 4 The BER after MPEG Compression (%).

Material Type	Bit Rate (kbps)		
	128	192	256
Classical	1.04	0.57	0.55
Voice	1.17	0.57	0.55
Violin	0.12	0.02	0.02
Opera	1.45	0.75	0.72
Flute	0.19	0.17	0.14
Piano	0.39	0.30	0.28

a small influence in the experimental results. Moreover, the samples may overflow the restriction in wave files and be clipped when the scale factor are higher than 1.0.

• Noise Addition

The White Gaussian Noise (WGN) was added to the watermarked audio signal. Then the watermark signal was detected from the noisy watermarked signal. In Fig. 4 (a) BER for some WNR (dB scale) is plotted. The proposed scheme with L = 128, the QIM scheme and RDM scheme were tested for comparison. The proposed scheme outperformed both the conventional QIM scheme and RDM scheme, especially when the WNR is low. The experimental results show that the LAQP scheme can improve the robustness performance of noise addition, which is frequently encountered in practical applications. It is important to note that the proposed scheme also has a better fidelity and value-metric invariance. The proposed method will have more widely use in the practical applications. And the performance can be improved when the spread length L is increased, as depicted in Fig. 4 (b). The length L should be determined depending on the practical applications.

• MPEG Compression

The MP3 compression is the most popular technique of audio compression. It is widely used in multimedia applications to reduce bit-rates and increase efficiency by a lossy coding. Table 4 show the experimental results of applying

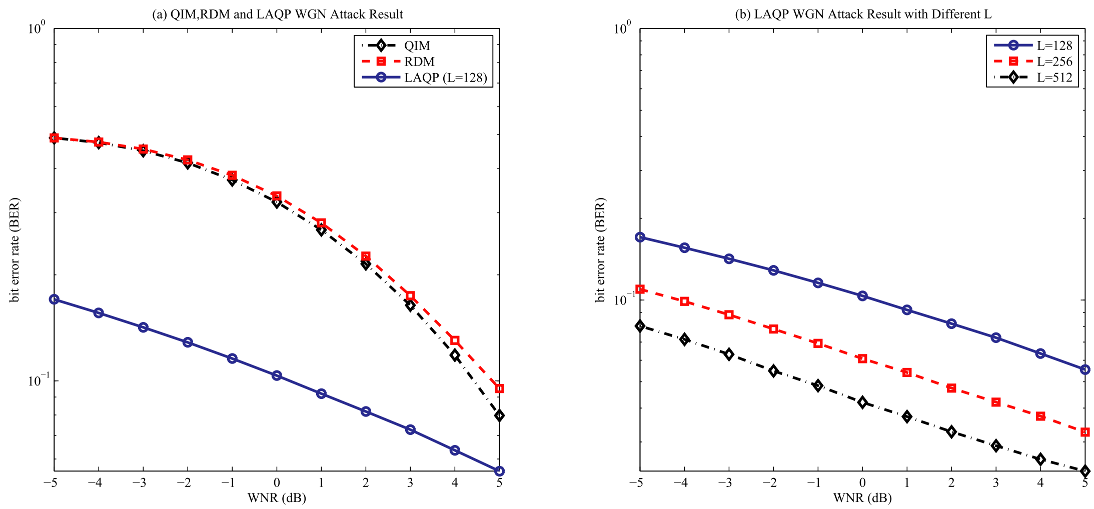

Fig. 4 The robustness test result of WGN attack.

Table 5 The BER after SMBA attack and comparison with other schemes.

Attack Name	Parameters	ODG of Attacked File	BER (%)			
			Proposed	[21]	[22]	[23]
AddBrumm	1 to 6k, 1 to 7k	−2.38 to −3.84	0 to 1	0 to 1	0	0 to 0.5
AddDynNoise	1 to 2	−0.10 to −2.67	0 to 2	2 to 7	20	0 to 8
AddFFTNoise	2048, 400	−0.43 to −0.58	0 to 2	0 to 2	11	0.5 to 1.5
AddNoise	1 to 20	−0.20 to −0.22	0 to 1	0 to 4	0 to 20	0 to 0.5
AddSinus	1 to 5k, 1 to 7k	−2.65 to −3.73	0	0	0	0
Amplify	10 to 200	−1.19 to −1.90	0 to 1	0	0 to 1	0
Compressor	2.1	−0.96 to −3.46	0 to 3	-	20	-
Echo	1 to 10	−0.15 to −1.35	10 to 18	0 to 3	63	0 to 0.5
FFT_HLPassQuick	2048, 1 to 10k, 18 to 22k	−1.36 to −3.89	0 to 3	0 to 2	5	1 to 3.2
FFT_Invert	2048	−0.06 to −0.19	0 to 1	0	2	1.5 to 2
FFT_RealReverse	2, 2048	−0.05 to −0.20	0 to 1	11 to 24	-	-
FFT_Stat1	2, 2048	−1.37 to −2.68	11 to 14	14 to 23	8	-
Invert	-	−0.18 to −0.20	0	0	-	0
LSBZero	-	−0.20 to −0.21	0	0	0	0
Normalize	-	−1.20 to −1.43	0	-	0	-
RC_HighPass	1 to 10k	−1.56 to −2.11	0 to 1	0 to 1	-	0 to 0.5
RC_LowPass	18 to 22k	−0.03 to −1.09	0 to 3	0 to 4	-	0
Resampling	22050	−1.48 to −1.58	20 to 21	38 to 47	0	5
Smooth	-	−2.19 to −3.46	11 to 15	15 to 31	11	-
Stat1	-	−0.35 to −2.34	11 to 18	21 to 44	8	-
VoiceRemove	-	−1.46 to −3.87	23 to 33	-	75	-

MP3 compression at different bit rates to the watermarked audio, the L = 512 in the test. The coding and decoding of MPEG compression have been performed by using a software implementation of ISO/MPEG-2 Audio Layer III coder. Although there are some bit errors in the test, the detection results were acceptable (all of the BER lower than 1% when bit rate are 192 kbps and 256 kbps, the BER lower than 1.5% when the bit rate is 128 kbps). The experimental results show that the proposed watermarking scheme can resist the common MPEG compression attack.

6.3 Audio Signal Attack with Stirmark Benchmark for Audio

To compare proposed method with other watermarking methods, we applied various attacks provided in Stirmark Benchmark for Audio (SMBA) [18] to the audio materials mentioned above. For comparison purpose, the DWR (which is equivalent to SNR in some papers) is adjusted to 30 dB, the audio signals are resampled to 44.1 kHz and the spread length L = 128 in these experiments. The SMBA software is used to attack the watermarked audio files and then the BER performance of watermark is calculated from the attacked files. The parameters of attacks are defined according to the SMBA web site [18]. We referred to some parameters used in paper [21]. For example, in AddBrumm, 1 to 6k shows the strength and 1 to 7k shows the frequency. It is illustrated that any value in the range 1 to 6k for the strength and 1 to 7k for frequency could be used with the change in BER lower than 1%. The experimental results of SMBA attacks are shown in Table 5. The table shows the range of ODG and BER for attacked files. The ODG

Table 6 The comparison of different watermarking algorithms.

Algorithm	DWR/SNR (dB)	ODG	Payload (bps)
[21]	33	−0.5	5501
[22]	-	-	689
[23]	30.5	−0.6	2996
Proposed	30	−0.2	689

is calculated between the watermarked signal and attacked watermarked signal.

In Table 6, we compare the performance of the proposed method and other watermarking schemes. In paper [22], the perceptual distortion is measured by mean opinion score (MOS). Our methods provide the same payload as the method in [22]. From the experimental results in Table 5, the proposed method has a better robustness in most of attacks. The methods in [21] and [23] provide a higher capacity than the proposed method, but we noted that the proposed method has a higher ODG score even the DWR/SNR in proposed method is lower than compared methods. That means the proposed method has a better fidelity. In our method, the perceptual model is used to improve the fidelity. In fact, when the spread length L is decreased, the capacity in the proposed method increases to the comparable level with methods [21] and [23]. As the experimental results in Sect. 6.2 illustrated, the selection of spread length L will introduce a trade off between the capacity and robustness. The length L should be determined according to the practical applications. The capacity of 689 bps is large enough in most applications. The experimental results show that the proposed method can provide a remarkable capacity while achieving a better performance in fidelity and robustness test.

7. Conclusion

In this paper a logarithmic adaptive quantization projection (LAQP) algorithm for digital watermarking system was proposed. For any watermarking system fidelity and robustness are always the goal to be optimized. The traditional QIM does not provide enough fidelity and robustness for practical application. More seriously, the sensitivity of value-metric scaling attack is the main weakness in the QIM methods for watermarking. In the proposed LAQP scheme, the perceptual model is used to improve the watermark's fidelity. The LAQP scheme used the binary balanced pseudo-random sequence and embedded the watermark into logarithm domain to construct an embedding domain invariant to value-metric scaling attack. In order to verify the characteristics of the algorithm, an audio watermarking system used the LAQP scheme and MPEG psychoacoustic model is designed. The experimental results show that the proposed scheme has better performance in fidelity than the conventional quantization scheme and is robust against common signal processing attack, especially the value-metric scaling attack.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (No's. 10925419, 90920302, 10874203, 60875014, 61072124, 11074275, 11161140319).

References

[1] I.J. Cox, M.L. Miller, and A.L. McKellips, "Watermarking as communications with side information," Proc. IEEE, vol.87, no.7, pp.1127–1141, July 1999.

[2] M. Costa, "Writing on dirty paper," IEEE Trans. Inf. Theory, vol.29, no.3, pp.439–441, May 1983.

[3] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good method for digital watermarking and information embedding," IEEE Trans. Inf. Theory, vol.47, no.4, pp.1423–1443, May 2001.

[4] J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," IEEE Trans. Signal Process., vol.51, no.4, pp.1003–1019, April 2003.

[5] P. Moulin and R. Koetter, "Data-hiding codes," Proc. IEEE, vol.93, no.12, pp.2083–2126, Dec. 2005.

[6] D. Kirovski and H.S. Malvar, "Spread-spectrum watermarking of audio signals," IEEE Trans. Signal Process., vol.51, no.6, pp.1020–1033, April 2003.

[7] F. Perez-Gonzalez, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: A high-rate data-hiding method invariant to gain attacks," IEEE Trans. Signal Process., vol.53, no.10, pp.3960–3975, Oct. 2005.

[8] J. Eggers, R. Bauml, and B. Girod, "Estimation of amplitude modifications before scs watermark detection," Proc. SPIE Security Watermarking Multimedia Contents, vol.4675, p.387–398, Jan. 2002.

[9] K. Lee, D.S. Kim, T. Kim, and K.A. Moon, "EM estimation of scale factor for quantization-based audio watermarking," Proc. Int. Workshop Digital Watermarking, p.316–327, Oct. 2003.

[10] J. Conway and N. Sloane, Sphere Packings, Lattices, and Groups, Springer-Verlag, New York, 1988.

[11] M.L. Miller, G.J. Doerr, and I.J. Cox, "Applying informed coding and embedding to design a robust, high capacity, watermark," IEEE Trans. Image Process., vol.13, no.6, pp.792–807, June 2004.

[12] F. Perez-Gonzalez, F. Balado, and J.R.H. Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," IEEE Trans. Signal Process., vol.51, no.4, pp.960–980, April 2003.

[13] R.M. Gray and T.G. Stockham, "Dithered quantizers," IEEE Trans. Inf. Theory, vol.39, no.3, pp.805–812, May 1993.

[14] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," IEEE Trans. Inf. Theory, vol.13, no.4, pp.619–621, Oct. 1967.

[15] H. Nussbaumer, "Pseudo-QMF filter bank," IBM Tech. Disclosure Bull, vol.24, pp.3081–3087, Nov. 1981.

[16] ISO/IEC, "Coding of moving pictures and associated audio for digital storage up to about 1.5 mbits/s," ISO/IEC IS 11172, 1993.

[17] M. Bosi and R. Goldberg, Introduction to digital audio coding and standards, Kluwer Academic Publishers, Dordrecht, 2003.

[18] "Stirmark benchmark for audio," tech. rep., http://wwwiti.cs.uni-magdeburg.de/˜alang/smba.php

[19] ITU-R, "Method for the subjective assessment of small impairments in audio systems including multichannel sound systems," Tech. Rep. BS. 1116, Geneva, Switzerland, 1994.

[20] ITU-R, "Method for objective measurements of perceived audio quality (PEAQ)," Tech. Rep. BS. 1387, Geneva, Switzerland, 1998.

[21] M. Fallahpour and D. Megias, "DWT-based high capacity audio watermarking," IEICE Trans. Fundamentals, vol.E93-A, no.1, pp.331–335, Jan. 2010.

[22] J.J. Garcia-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Data hiding in audio signal using rational dither modulation," IEICE Electronics Express, vol.5, no.7, pp.217–222, 2008.
[23] M. Fallahpour and D. Megias, "High capacity audio watermarking using FFT amplitude interpolation," IEICE Electronics Express, vol.6, no.14, pp.1057–1063, 2009.

Xuemin Zhao received the B.E. degree from the Nankai University, Tianjin, China. Now he is an M.S. & Ph.D. candidate in Key Laboratory of Speech Acoustics and Content Understanding at Institute of Acoustics, Chinese Academy of Sciences. His research interests include audio watermarking and information security.

Yuhong Guo received the B.E. degree from the Nanjing University, Nanjing, China. Now he is an M.S. & Ph.D. candidate in Key Laboratory of Speech Acoustics and Content Understanding at Institute of Acoustics, Chinese Academy of Sciences. His research interests include audio watermarking and speech recognition.

Jian Liu received the B.E. degree at the Electronic Engineering Department of University of Science and Technology of China (USTC), in 1993, and the M.S. & Ph.D. degree in 1996 & 1999 from Institute of Acoustics, Chinese Academy of Sciences. From 1999 to 2002, he worked as a Senior Researcher in Intel China Research Center (ICRC), Beijing, China. After 2002 he returned to Institute of Acoustics, Chinese Academy of Sciences. Now he is an associate professor of Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences. His research interests include speech recognition, audio watermark and audio signal processing. He has published more than 20 papers and holds 10+ patents.

Yonghong Yan received the B.E. degree at the Electronic Engineering Department of Tsinghua University in 1990, and Ph.D. in Computer Science and Engineering from Oregon Graduate Institute of Science & Engineering in 1995. From 1995 to 1998, he worked in OGI as Assistant Professor, Associate Director and Associate Professor of the Center for Spoken Language Understanding, and from 1998 to 2001 he worked as the Principal Engineer of Intel Microprocessors Research lab, Director and Chief Scientist of Intel China Research Center. In 2002 he returned to China to work for Chinese Academy of Sciences. He is a professor and director of Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences. His research interests include large vocabulary speech recognition, speaker/language recognition and audio signal processing. He has published more than 100 papers and holds 40 patents.

Qiang Fu received the B.E. degree from the Xi'an Technological Uninversity, Xi'an, China, in 1994, the M.S. degree in electronic engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 1997, and the Ph.D. degree in electronic engineering from Xidian University, Xi'an, in 2000. In 2000, he was working at a Researcher in Motorola China Research Center (MSRC), Shanghai, China. From 2001 to 2002, he was working as a senior Research Associate in Center for Spoken Language Understanding (CSLU), OGI School of Science and Engineering at Oregon Health & Science University, Oregon, USA. From 2002 to 2004, he was working as Senior Postdoctoral Research Fellow in Department of Electric and Computer Engineering, University of Limerick, Ireland. He is currently an Associated Professor in Institute of Acoustics, Chinese Academy of Sciences, China. His research interests are in robust signal processing for speech recognition, including speech analysis, enhancement and separation, and channel equalization, etc. Dr. Qiang Fu is a member of IEEE Signal Processing Society.