Multimedia Systems (2014) 20:155–164
DOI 10.1007/s00530-013-0325-1

REGULAR PAPER

Secure logarithmic audio watermarking scheme based on the human auditory system

Mehdi Fallahpour · David Megías

Published online: 9 June 2013

Abstract This paper proposes a high capacity audio watermarking algorithm in the logarithm domain based on the absolute threshold of hearing of the human auditory system (HAS), which makes this scheme a novel technique. When considering the fact that the human ear requires more precise samples at low amplitudes (soft sounds), the use of the logarithm helps us design a logarithmic quantization algorithm. The key idea is to divide the selected frequency band into short frames and quantize the samples based on the HAS. Using frames and the HAS improves the robustness, since embedding a secret bit into a set of samples is more reliable than embedding it into a single sample. In addition, the quantization level is adjusted according to the HAS. Apart from remarkable capacity, transparency and robustness, this scheme provides three parameters (frequency band, scale factor and frame size) which facilitate the regulation of the watermarking properties. The experimental results show that the method has a high capacity (800–7,000 bits per second), without significant perceptual distortion (ODG >1) and provides robustness against common audio signal processing such as added noise, filtering and MPEG compression (MP3).

Keywords Audio watermarking · Multimedia security · Digital rights management

M. Fallahpour (✉)
School of Information Technology and Engineering (SITE), University of Ottawa, Ottawa, Canada
e-mail: Fallahpour@gmail.com

D. Megías
Estudis d'Informàtica, Multimèdia i Telecomunicació, Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Spain
e-mail: dmegias@uoc.edu

1 Introduction

Traditional data protection methods, such as encryption, are not enough for audio copyright enforcement. Digital watermarking is a popular technique for digital data protection and digital rights management [26, 27]. According to the International Federation of Phonographic Industry (IFPI) [3], audio watermarking should meet the following requirements: (a) imperceptibility: the watermarking scheme should not affect the perceptual quality of audio—in this paper, this is achieved using a psychoacoustic model to guarantee that the watermarking process does not distort the cover audio signal—(b) capacity: refers to the number of bits that can be embedded into the audio signal within a second and (c) robustness: the embedded watermark data should not be removed or eliminated by common audio signal processing operations and attacks, such as additive and multiplicative noise, MP3 compression, and filtering. All these requirements are often conflicting with each other, which makes the design of high capacity, transparent and robust audio watermarking schemes a challenging task.

Several research results exist for watermarking in the logarithm and cepstrum domains. Lee et al. [4] introduced a digital audio watermarking such that the watermark is embedded into cepstrum coefficients of the audio signal using techniques analogous to spread spectrum communications. Li and Yu [5] suggested a robust and transparent audio data embedding method in the cepstrum domain. BCH code-based robust audio data hiding in the cepstrum domain is presented in [7]. Hsieh et al. [6] suggested an audio watermarking technique based on the time energy features. Li et al. [8] proposed an audio watermarking scheme in the cepstrum domain based on the statistical mean manipulation. The embedded watermark is robust against MP3 compression and additive noise. Hu and Chen

[9] proposed cepstral watermarking that manipulates the statistic mean. To avoid sharp discontinuities in the frame boundaries caused by the watermarking process, a small transition area is deliberately placed between frames, leading to an improvement in perceived quality as well. Ko et al. [21] suggested a digital watermarking method based on the log scaling of frequency in the decoding process for robust detection. Yang et al. [10] is the first technique on applying log-polar mapping to audio watermarking. The log-polar mapping is only applied to the frequency index, not to the transform coefficients, which prevents the reconstruction distortion of inverse log-polar transform and reduces the computational cost.

Watermarking methods based on the human auditory system (HAS) have been suggested in different previous works, such as [1, 2, 11, 30]. Garcia [1] proposed an algorithm to estimate the masking threshold in the psychoacoustic model of the HAS. Tsai et al. [2] proposed an intelligent audio watermarking method based on the characteristics of the HAS and neural networks in the DCT domain. Also, in [11], the watermark is embedded into selected DCT coefficients of the host audio signal such that the signal to noise ratio is maintained at a level which is audibly annoying to the HAS. Lie and Chang [30] proposed an algorithm that maintains an energy relation between every three sample sections to represent the embedded bit information by scaling up or down corresponding amplitudes and conserving audio waveforms that are perceivable to human ears.

When considering the embedding domain, audio watermarking techniques can be classified into time domain and frequency domain methods. In [12, 13], which were proposed by the authors of this paper, the discrete/fast Fourier transform (DFT/FFT) domain is selected to embed watermarks for taking benefit of the translation-invariant property of the FFT coefficients to resist small distortions in the time domain. In fact, as compared to time domain schemes, transform-based methods provide better perceptual quality and robustness against common attacks at the price of increasing the computational burden.

This paper presents an audio watermarking algorithm in the logarithm domain based on the HAS. Changing the quantization level based on the HAS in the logarithm domain makes the algorithm a novel and useful idea. Based on the requirements, a frequency band, a frame size and a scaling factor are selected and each secret bit is embedded into a frame. In addition to very high capacity, imperceptible distortion and robustness against common attacks, which make this scheme outperform other works in the literature, the other main features of the proposed algorithm are as follows: (1) using logarithm coefficients enhances the robustness, (2) watermark extraction is blind without using the host signal, (3) adjusting the quantization level based on the HAS improves transparency and robustness making it possible to enhance them at a same time, which is a significant challenge of many techniques, (4) embedding a single secret bit into a frame with an adjustable frame size provides a convenient solution to obtain a trade-off between the properties of the watermarking system, and (5) an encryption technique enhances the security of the system in such a way that an attacker, even if he/she knows the watermarking method, cannot extract the raw secret bits since a key is required to decrypt them.

The remainder of this paper is organized as follows. A brief overview of the HAS is given in Sect. 2. Section 3 combines the above techniques to propose a new method for audio watermarking. Moreover, the detailed watermark embedding and extraction algorithms are explained in that section. The experimental results and comparison with other schemes are given in Sect. 4 and, finally, relevant conclusions are drawn in Sect. 5.

2 Human auditory system

Extensive work has been performed over the years in understanding the characteristics of the HAS and applying this knowledge to audio compression and audio watermarking. Figure 1 shows a typical absolute threshold curve, where the abscissae are the frequencies measured in hertz (Hz) and the ordinates are the absolute thresholds in decibels (dB). As it can be observed, human beings tend to be more sensitive towards frequencies in the range from 1 to 4 kHz, while the threshold increases steeply at very high and very low frequencies. Based on the HAS, the human ear sensitivity in higher frequencies is lower than in middle frequencies. Thus, it is clear that, by embedding data in the high frequency band, which is used in the proposed

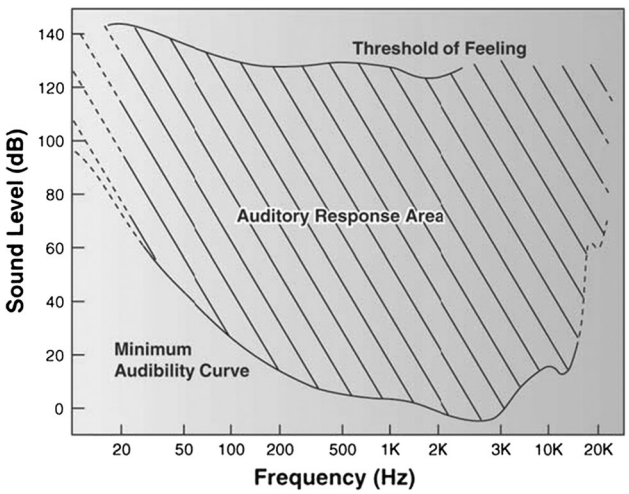

Fig. 1 Typical absolute threshold curve of the human auditory response

scheme, the distortion will be mostly inaudible and thus more transparency will be obtained.

The HAS can be modeled as a frequency analyzer containing a set of 25 band-pass filters, named critical bands, that cover the range 10 Hz–20 kHz. In the absence of other sounds, the perceived intensity of a single sound, called loudness, depends on this sound's pressure level (SPL), duration and frequency. The threshold for masking a sound is determined by the frequency and SPL [2].

3 Proposed method

In this method, we use the following technique to embed a bit stream (secret bits) into the logarithm coefficients. First, based on the desired capacity, transparency and robustness, the frequency band, frame size and scale factor should be selected. The selected band is then divided into short frames and each sample is quantized based on the HAS. Each single secret bit of the watermark stream is embedded into all samples of a frame, which makes the method more robust against attacks.

Based on the HAS, the human ear sensitivity is different in various frequencies, i.e. the absolute threshold of hearing (ATH) is different for different frequency bands. The embedding scheme takes advantage of changes of ATH in various frequency bands to adjust the quantization level.

3.1 Tuning

The proposed method provides three parameters to adjust three properties of the watermarking system. The frequency band, the scaling factor (α) and the frame size (d) are the three parameters of this method to adjust capacity, perceptual distortion and robustness.

Since most MP3 cut-off frequencies [25] are higher than 16 kHz, the high frequency band is set to 16 kHz. Then, to select the frequency band, only the low frequency band, f_l, should be adjusted. The default value for low frequency band is 9 kHz. Decreasing f_l implies increasing capacity and distortion.

Increasing the frame size, d, results in a better robustness, but capacity decreases. The default value for the frame size is $d = 5$. Finally, to achieve better transparency the scaling factor, α, should be increased. However, decreasing the scaling factor leads to better robustness.

Figure 2 shows the flowchart for the selection of the tuning parameters. In the initialization, f_l is 9 kHz, d is 5 and α is 10. This flowchart facilitates adjusting the parameters based on the requirements. However, adjusting the parameters based on some demands is very difficult and considering a trade-off between capacity, transparency and robustness is always necessary.

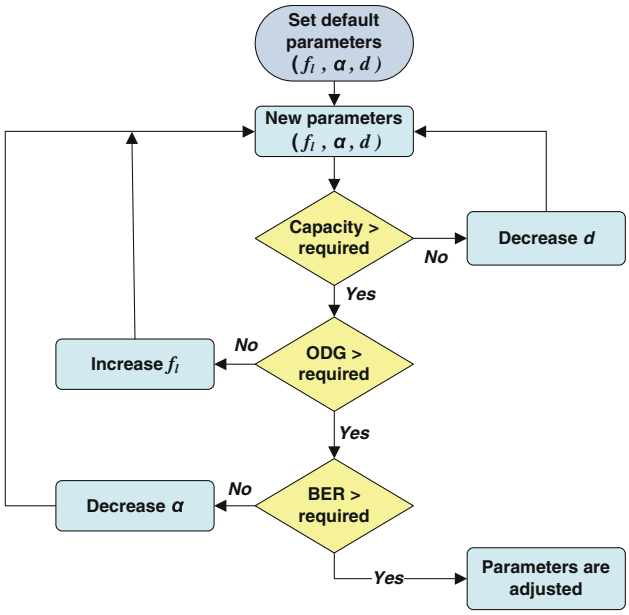

Fig. 2 Flowchart of tuning steps

3.2 Embedding the secret bits

The frequency band, the scaling factor (α) and the frame size (d) are the three required parameters in the embedding process which have to be adjusted according to the requirements. In this section, for simplicity, we do not consider the regulation of these parameters and just take them as fixed. The effects of these parameters are analyzed in Sect. 4.

In the embedding steps, first the FFT is calculated and then the logarithm is computed. The next step is embedding the secret bits and, finally, the inverse FFT is applied to generate the marked audio signal. The embedding steps are detailed below.

1. Compute the FFT of the original audio signal. We can use the whole file (for short clips, e.g. with less than 1 min) or blocks of a given length (e.g. 10 s) for longer files.
2. Calculate the logarithm coefficients of the FFT samples.
3. Divide the logarithm samples in the selected frequency band into frames of size d.
4. To improve the security, the secret bit stream, B, is encrypted by a key, C, to form the watermark signal W:

   $$W = E(C, B),$$

   where E is the encryption operation.
   For example, the embedded bit stream W may be computed as the exclusive-or (XOR) sum of the real watermark and a pseudo-random bit stream. Then, the seed C to produce the pseudo-random bit stream

would be required as part of the secret key both at the embedder and the detector [20].

5. The marked logarithm samples $\left\{c_j^{'}\right\}$ are obtained by using the following equation:

$$c_j' = \begin{cases} \lfloor c_j\delta_j \rfloor/\delta_j, & \text{if } w_l = 0, \\ \left(\lfloor c_j\delta_j \rfloor + 0.5\right)/\delta_j, & \text{if } w_l = 1. \end{cases}$$

where $l = \lfloor j/d \rfloor + 1$, w_l is the lth bit of the watermark, $\delta_j = \alpha/\mathrm{ATH}_j$, α is a scaling factor and $\lfloor x \rfloor$ denotes the nearest integer value to x towards negative infinity. ATH_j is the absolute threshold sound level for each sample which is calculated by:

$$\mathrm{ATH}(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 0.0010\left(\frac{f}{1000}\right)^4 \text{ (dB SPL)}.$$

Each secret bit is embedded into a suitable frame.

6. Finally use the inverse logarithm (exponential function) and inverse FFT to obtain the marked audio signal.

Figure 3 shows the flowchart for the embedding steps.

As it is evident, increasing the scale factor increases the accuracy of samples which results in better transparency (less distortion) but also less robustness against attacks. In addition, by enlarging the frequency band, the capacity and distortion increase and robustness decreases. Finally, increasing the frame size strengthens the robustness against attacks and reduces the capacity.

Note that the HAS model has been applied (in Step 5) using only its passive properties (without frequency masking). This choice is much more efficient from a computational point of view and makes it possible to use the proposed system in real-time applications. If real-time embedding is not a requirement, frequency masking could be considered in the scheme. However, the transparency results achieved with the scheme (as presented in Sect. 4) are remarkable even without using frequency masking. Thus, the application of frequency masking is left for future work.

3.3 Extracting the secret bits

The watermark extraction process is performed in the logarithm domain and the required parameters can be considered as side information. The scale factor, the frame size and the frequency band can be transmitted in a secure way to the decoder or they could be embedded using some fixed settings. For example, we could use default parameters to embed only the value of the adjusted parameters. Then, in the decoder, the adjusted parameters would be extracted using the default parameters and the secret bits would be obtained using the extracted adjusted parameters. Note that these parameters (frequency band, scaling factor and frame size) are also part of the secret key of the scheme (required both at the embedding and the detector side), together with the key C used for encryption. Hence, if the values of the tuning parameters are embedded at fixed (default) positions, they should be embedded as ciphertext for security reasons. Because the host audio signal is not required in the detection process, the detector is blind. The detection process can be summarized in the following steps:

1. Compute the FFT of the marked audio signal.
2. Calculate the logarithm of the FFT coefficients.
3. Divide the logarithm samples in the selected frequency band into frames of size d.
4. To detect a secret bit in a frame, each sample should be examined to check if it is a zero frame ("0" embedded) or a one frame ("1" embedded). Then, depending on the evaluation for all samples in the current frame, a secret bit can be detected. The extracted bit from each sample ($S_j^{'}$) is achieved using the following equation:

$$S_j' = \begin{cases} 0, & \text{if } 0.25 > \left|c_j'\delta_j - \mathrm{round}\left(c_j'\delta_j\right)\right|, \\ 1, & \text{if } 0.25 \le \left|c_j'\delta_j - \mathrm{round}\left(c_j'\delta_j\right)\right|. \end{cases}$$

After getting information about all samples in the frame, based on the number of samples which represent "0" or "1" (voting scheme), the secret bit ($w_l^{'}$) related to the frame can be extracted. If the number of samples identified as "0" is equal to or larger than half the frame size, the extracted bit is "0", otherwise it is "1".

5. To achieve the raw watermark stream we need to use the encryption key and the decryption algorithm.

In fact, an attacker should have access to the following information to extract the secret stream:

- Embedding algorithm;
- Encryption algorithm;
- Encryption/decryption key;
- Frequency band in the embedding procedure;
- Frame size in the embedding procedure;
- Scaling factor in the embedding procedure.

Thus, if it is not impossible, it is extremely difficult for an attacker to extract the secret information embedded into the audio signal.

4 Experimental results

To evaluate the performance of the proposed method and to consider the applicability of the scheme in a real scenario,

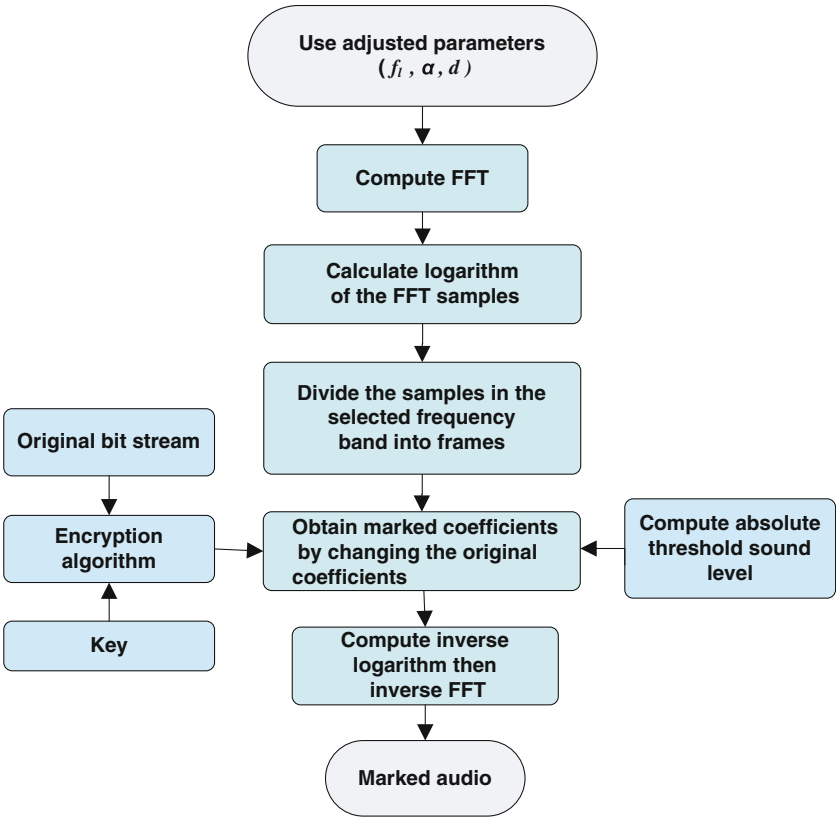

Fig. 3 Flowchart of embedding steps

all songs in the album *Rust* by No, Really [16] and the most popular tracks of different albums in different genres [28] have been used. All audio clips are sampled at 44.1 kHz with 16 bits per sample and two channels. Note that the presented results are just for one channel: the left one. In other words, we have converted the stereo signals to mono and used only the left channel.

Audio watermarking applications require a trade-off between the desired properties, namely, capacity, robustness and transparency. The following scenarios can be assumed to obtain different results regarding these three properties:

(1) No robustness: in this case, very high capacity and transparency can be achieved;
(2) Semi-robustness: robustness against MP3 compression and common attacks is demanded. In this case, more distortion should be accepted, as compared to Scenario 1;
(3) Robustness against many attacks with a wide range of changes is desirable. This is more difficult and complicated than the previous scenarios, since we need robustness against various attacks. Thus, according to the trade-off between capacity, transparency and robustness, a sacrifice in capacity and transparency is required.

The significant advantage of this scheme is providing superior results for all these three conditions.

The objective difference grade (ODG) has been used in this paper to evaluate the transparency of the proposed algorithm. The ODG is one of the output values of the ITU-R BS.1387 PEAQ [17] standard, where ODG $= 0$ means no degradation and ODG $= -4$ means a very annoying distortion. Values of ODG between -1 and 0 are required for transparent watermarking. The OPERA software [29] based on the ITU-R BS.1387 standard has been used to compute this objective measure of quality.

Table 1 shows the perceptual distortion, payload and BER under the MP3 compression attack with different bit rates. Note that different values for parameters are used to achieve a different trade-off between capacity, transparency and robustness, as usual for all watermarking systems. For example, for "Beginning of the end", a frame size $d = 1$ and a wide frequency band, the results show high capacity and robustness against MP3-128. On the other hand, using a frame size $d = 5$ and a narrower frequency band, less capacity and better robustness is achieved. Also, increasing the scaling factor results in more accuracy and better transparency, whereas decreasing it leads to better robustness.

In this scheme, we have three parameters and audio watermarking schemes have three main properties. Thus, we have three inputs and three outputs for a nonlinear system which works based on the HAS. Finding explicit equations to adjust the requirements is extremely difficult

Table 1 Results of three real song signals (robust against Table 2 attacks)

Audio file	Time (m:sec)	Scaling factor	Frame size	Frequency band (kHz)	SNR (dB)	MP3 attack		ODG of marked	Payload (bps)
						Rate	BER		
Beginning of the end	3:16	10	3	10–16	35.8	128	0.05	−0.37	2,017
		10	3	10–16	35.8	96	0.09	−0.37	2,017
		10	1	10–16	35.8	128	0.03	−0.37	6,051
		10	1	10–16	35.8	96	0.09	−0.37	6,051
		10	5	10–16	35.8	80	0.12	−0.37	1,210
		8	1	12–16	29.8	128	0.01	−0.49	4,050
		8	5	12–16	29.8	64	0.09	−0.49	810
		8	5	12–16	29.8	80	0.03	−0.49	810
Breathing on another planet	3:13	10	1	12–16	26.7	96	0.03	−0.38	4,050
		10	1	9–16	25.4	128	0.06	−0.74	7,050
		10	5	12–16	26.4	96	0.01	−0.43	810
		9	5	12–16	21.9	80	0.06	−0.30	810
Thousand yard stare	3:57	10	5	12–16	27.7	96	0.01	−0.21	810
		10	5	10–16	27.6	96	0.08	−0.27	1,210
Floodplain	3:13	10	5	11–16	32.8	128	0.02	−0.36	1,010
		8	5	10–16	27.1	128	0.02	−0.40	1,210
Do you know where your children	2:31	10	5	12–16	25.3	80	0.05	−0.87	810
		10	5	12–16	25.3	64	0.09	−0.87	810
Rust	2:33	9	5	12–16	27.1	80	0.06	−0.20	810
		9	5	10–16	26.8	96	0.08	−0.74	1,210
Molten	2:09	10	5	10–16	32.9	112	0.05	−0.46	1,210
		9	5	12–16	30.4	80	0.05	−0.48	810
Citizen, go back to sleep	1:57	10	5	13–16	36.2	64	0.09	−0.57	610
Go	1:51	9	5	10–16	23.9	96	0.05	−0.72	1,210
		10	5	9–16	27.8	96	0.08	−0.89	1,410
Stop payment	2:09	10	5	12–16	32.8	64	0.09	−0.19	810
		10	1	12–16	32.8	80	0.11	−0.19	4,050
Face the day	4:37	20	5	12–16	30.7	96	0.11	−0.91	810
		23	5	10–16	30.1	128	0.6	−0.88	1,210
Faded war	3:34	10	5	10–16	32.2	96	0.08	−0.30	1,210
		9	5	10–16	27.2	96	0.07	−0.79	1,210
The Easton Ellises—dance it, …	3:46	11	5	10–16	25.6	96	0.09	−0.80	1,210
		10	5	12–16	30.7	80	0.10	−0.95	810
Lucky one	3:34	10	5	12–16	32.8	128	0.01	−0.59	810
		9	5	12–16	29.6	96	0.06	−0.86	810
I want you (pop rock remix)	3:24	11	5	10–16	28.9	128	0.02	−0.97	1,210
		11	5	10–16	28.9	96	0.12	−0.97	1,210
Bionic	3:58	10	5	12–16	31.2	128	0.01	−0.69	810
		9	5	12–16	28.6	96	0.09	−0.89	810
Wonder doll	3:28	10	5	10–16	30.8	96	0.11	−0.88	1,210
		11	5	12–16	31.5	128	0.05	−0.66	810
Average	3:07	10.26	4.31	11–16	29.86	128	0.03	−0.58	1,680
						112	0.05		
						96	0.08		
						80	0.07		
						64	0.09		

and sometimes impossible. We may use different loops and conditions to obtain better results.

As mentioned in the Sect. 3.1, we have general tuning rules which can help us to reach the requirements or to get close to them very quickly. The frame size has more effect on robustness, whereas the scaling factor and frequency band have more effect on transparency and capacity. In other words, by increasing the frame size better robustness is achieved. In addition, increasing the frequency band leads to better capacity. Finally, by increasing the scaling factor better transparency can be achieved.

Note that these parameters allow to regulate the ODG between 0 (not perceptible) and −1 (not annoying), with about 800–7,000 bits per second (bps) of capacity and allowing robustness against MP3-128, which are extremely better than typical requirements.

The default parameter values (frequency band 12–16 kHz, frame size equal to 5 and scaling factor equal to 10) have been selected for "Stop payment" and "Breathing on another planet" audio test files. The ODG for "Breathing on another planet" is −0.43 and for "Stop payment" it is −0.19.

Table 2 illustrates the effect of several common attacks, provided by the Stirmark Benchmark for Audio (SMBA) v1.0 [14], on ODG and BER for the two selected audio test files. The parameters were selected for each signal, then the embedding method was applied, the SMBA software was used to attack the marked files and, finally, the detection method was applied for the attacked files. The ODG in Table 2 is calculated between the marked and the attacked-marked files. The parameters of the attacks are selected according to the definitions provided in the SMBA web site [18].

For example, in AddBrumm, 1–4 k shows the strength and 1–4.5 k shows the frequency. This row reports that any value in the range 1–4 k for the strength and 1–4.5 k for the frequency can be used without any significant change in BER. In fact, this table provides the average results for the test signals based on the BER and, in the case with the same BER, based on the limitation of the parameters. It can be seen that the proposed scheme produces excellent robustness against all these attacks (BER close to zero) even if the attacks significantly distort the audio files (even for ODG lower than −3).

Table 3 shows how considering the HAS improves the properties of the watermarking system. In fact, in the proposed method, the quantization level is adjusted by the ATH which results in better properties of the method. For example for "Breathing on another planet", when the frequency band is 9–16 kHz, the BER rate for both,

Table 2 Robustness test results

Attack name	Stop payment			Breathing on another planet		
	ODG of attacked file	Parameters	BER	ODG of attacked file	Parameters	BER
AddBrumm	−3.3	1–4 k, 1–4.5 k	0.00	−3.1	1–3 k, 1–4 k	0.00
	−3.1	1–4 k, 12–16 k	0.28	−2.8	1–3 k, 12–16 k	0.23
AddDynNoise	−2.5	1–10	0.05	−2.5	1–12	0.01
AddNoise	−1.7	1–200	0.01	−0.9	1–200	0.01
AddSinus	−2.8	1–5 k, 1–5 k	0.00	−2.2	1–5 k, 1–5 k	0.00
	−2.1	1–5 k, 12–16 k	0.00	−1.7	1–5 k, 12–16 k	0.00
Amplify	−0.1	60–130	0.00	−0.1	60–190	0.01
BassBoost	−3.6	1–50, 1–50	0.00	−3.6	1–70, 1–65	0.01
Echo	−3.2	1–5	0.12	−3.3	1–5	0.10
FFT_RealReverse	−2.8	2	0.00	−3.2	2	0.00
FFT_Stat1	−0.2	2	0.00	−0.1	2	0.00
Invert	−2.8	–	0.00	−3.2	–	0.00
LSBZero	−0.1	–	0.00	−0.1	–	0.00
RC_HighPass	−2.3	0–18 k	0.00	−3.7	0–18.5 k	0.00
RC_LowPass	−3.3	7 k–20 k	0.00	−1.7	8 k–20 k	0.00
Stat1	−2.6	–	0.03	−0.7	–	0.02
MP3	−0.0	256	0.00	−0.0	256	0.00
	−0.2	128	0.00	−0.2	128	0.00
	−0.5	96	0.03	−0.4	96	0.01
	−0.6	80	0.05	−0.5	80	0.06
	−1.0	64	0.09	−0.8	64	0.13

Table 3 Adaptive vs. constant quantization

Audio file	Quantization	Scaling factor	Frame size	Frequency band (kHz)	MP3 attack		ODG of marked	Payload (bps)
					Rate	BER		
Beginning of the end	Adaptive	10	1	10–16	128	0.03	−0.37	6,050
	Constant	20	1	10–16	128	0.03	−0.95	6,050
	Adaptive	8	1	12–16	128	0.01	−0.49	4,030
	Constant	20	1	12–16	128	0.03	−0.46	4,030
Breathing on another planet	Adaptive	10	1	12–16	96	0.03	−0.38	4,030
	Constant	20	1	12–16	96	0.12	−0.65	4,030
	Adaptive	10	1	9–16	128	0.06	−0.74	7,050
	Constant	30	1	9–16	128	0.04	−2.86	7,050
Stop payment	Adaptive	10	5	12–16	64	0.09	−0.19	810
	Constant	30	5	12–16	64	0.07	−1.37	810
	Adaptive	10	1	12–16	80	0.11	−0.19	4,050
	Constant	30	1	12–16	80	0.10	−1.37	4,050

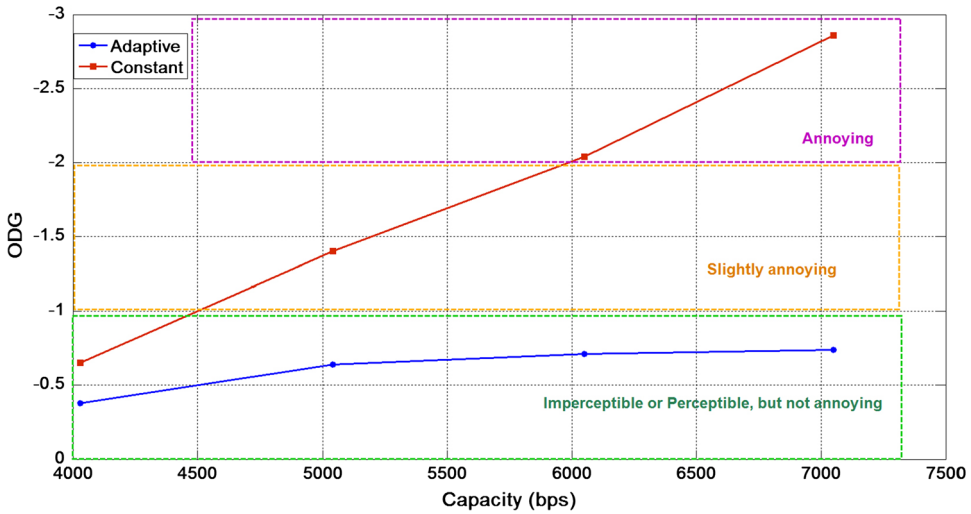

Fig. 4 Difference between constant versus adaptive quantization for "Breathing on another planet"

considering the HAS and constant quantization (without any HAS model), is about 0.05. However, the distortion caused by watermarking for adaptive quantization is almost imperceptible whereas it is absolutely annoying when constant quantization is used.

To reduce the computational time and memory usage, songs can be divided into small clips, e.g. 10 s each. Then, the synchronization method described in [19] and the embedding algorithm described in this paper was applied for each clip separately.

Figure 4 shows the difference between adaptive and constant quantization. As this plot illustrates, using adaptive scaling quantization, the transparency can be improved and kept in a perceptible but not annoying area, which is the typical requirement for a watermarking system. However, using constant quantization, the embedding method can destroy the cover audio signal and the ODG will be in the annoying area when capacity is increased beyond some threshold.

The method proposed in this paper has been compared with several recent audio watermarking strategies. Almost all the audio data hiding schemes which produce very high capacity are fragile against signal processing attacks. Because of this, it is not possible to establish a comparison of the proposed scheme with other audio watermarking schemes which are similar to it as capacity is concerned. Hence, we have chosen a few recent and relevant audio watermarking schemes in the literature. In Table 4, we compare the performance of the proposed watermarking algorithm and several recent audio watermarking strategies robust against the MP3 attack. Speech applications and codecs are considered in [14]. The distortion introduced to the marked signal is slightly *ann*oying, capacity is very low and robustness is achieved against compression attacks.

Table 4 Comparison of different watermarking algorithms

References	Capacity (bps)	Imperceptibility in SNR (dB)	Imperceptibility (ODG)
[14]	8	Not reported	$-3 < \text{ODG} < -1$
[15]	64	30–45	$-1 < \text{ODG}$
[12]	3 k	30.55	-0.6
[13]	1.5–8.5 k	35–45	$-0.8 < \text{ODG} < -0.1$
[22]	4–512	Not reported	$-1 < \text{ODG}$
[23]	7–30	Not reported	Not reported
[24]	80	Not reported	-1.04
[30]	15	Not reported	Not reported
[31]	41–165	Not reported	$-1.14 < \text{ODG} < -0.88$
Proposed	800–7 k	21–36	$-1 < \text{ODG} < -0.1$

Recently, [15] introduces a very fast scheme which uses the Fourier transform. The embedding bit-rate is low, 64 bits per second, but the scheme is very robust against several attacks. Lie et al. [30] consider the HAS to present a method in the time domain, but the embedding capacity is quite low. Baras et al. [31] present a transparent technique, but, in some cases, the distortion is slightly annoying. The provided capacity in [31] is about a hundred bits. Fallahpour et al. [12, 13], which were also proposed by the authors of this paper, have a remarkable performance in the different properties, but the scheme proposed in this paper can manage the needed properties better, since there are three useful adjustable parameters. In particular, the results of this paper make it possible to improve the transparency results with respect to [12, 13] due to the explicit use of the HAS and adaptive quantization. This comparison shows the superiority in both capacity and imperceptibility of the suggested method for the same robustness with respect to other robust schemes. This is particularly relevant, since the proposed scheme can embed much more information and, at the same time, introduces less distortion in the marked file. In short, the proposed scheme achieves higher capacity if we compare it with methods with similar robustness and imperceptibility, and more robustness and imperceptibility if we compare it to methods with similar capacity.

5 Conclusions

This paper suggests an audio watermarking algorithm in the logarithm domain based on the HAS. The human ear requires more precise samples at low amplitudes (soft sounds) and taking advantage of the logarithm it is possible to design a logarithmic quantization algorithm to exploit this property. Adjusting the quantization level based on the HAS in the logarithm domain results in a very high capacity, imperceptible distortion and robustness. The most notable features of the proposed algorithm are as follows: (1) blind watermark extraction, (2) adaptive quantization based on the HAS that improves transparency and robustness, making it possible to enhance them simultaneously, which is a main challenge of many techniques; and (3) embedding a single secret bit into all samples of a frame, with an adjustable frame size, delivers a suitable solution to obtain a convenient trade-off between the properties of the watermarking system.

The experimental results show that the scheme provides high capacity (800–7,000 bps), without significant perceptual distortion (ODG is greater than -1) whilst achieving robustness against common audio signal processing, such as added noise, filtering and MPEG compression (MP3).

Acknowledgments This work was partly funded by the Spanish Government through projects TSI2007-65406-C03-03 "E-AEGIS", TIN2011-27076-C03-02 "CO-PRIVACY" and CONSOLIDER INGENIO 2010 CSD2007-0004 "ARES".

References

1. Garcia, R.: Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory. In AES 107th Convention, pp. 123–131 (1999)
2. Tsai, H.H., Cheng, J.S., Yu, P.T.: Audio watermarking based on HAS and neural networks in DCT domain. EURASIP J. Appl. Signal Process **3**, 252–263 (2003)
3. Katzenbeisser, S., Petitcolas, F.A.P.: Information hiding techniques for steganography and digital watermarking. Artech. House, Boston (2000)
4. Lee, S.K., Ho, Y.S.: Digital audio watermarking in the cepstrum domain. IEEE Trans. Consum. Electron. **46**(3), 744–750 (2000)
5. Li, X., Yu, H.H.: Transparent and robust audio data hiding in cepstrum domain. In: IEEE International Conference on Multimedia and Expo, vol. 1, pp. 397–400 (2000)
6. Hsieh, C.-T., Sou, P.-Y.: Blind cepstrum domain audio watermarking based on time energy features. In: 14th International Conference on Digital signal processing, vol. 2, pp. 705–708 (2002)
7. Liu, S.C., Lin, S.D.: BCH code based robust audio watermarking in the cepstrum domain. J. Inform. Sci. Eng. **22**, 535–543 (2006)
8. Li, S., Cui, L., Choi, J., Cui, X.: An audio copyright protection schemes based on SMM in cepstrum domain. In: International Workshops on Structural, Syntactic, and Statistical Pattern Recognition (SSPR and SPR'06), LNCS, vol. 4109, pp. 923–927 (2006)
9. Hu, H.T., Chen, W.H.: A dual cepstrum-based watermarking scheme with self-synchronization. Signal Process. **92**(4), 1109–1116 (2012)
10. Yang, R., Kang, X., Huang, J.: Robust Audio Watermarking Based on Log-Polar Frequency Index. 7th International Workshop on Digital Watermarking, IWDW 2008, Volume 5450 of Lecture Notes in Computer Science, pp. 124–138, Springer (2008)
11. Dutta, M.K., Gupta, P., Pathak, V.K.: A perceptible watermarking algorithm for audio signals. Multime'd. Tools Appl. pp. 1–23 Feb 2012
12. Fallahpour, M., Megías, D.: High capacity audio watermarking using FFT amplitude interpolation. IEICE Electron. Express **6**(14), 1057–1063 (2009)

13. Fallahpour, M., Megías, D.: Robust high-capacity audio watermarking based on FFT amplitude modification. IEICE Trans. Inf. Syst. **E93-D**(01), 87–93 (2010)
14. Nishimura, A.: Audio data hiding that is robust with respect to aerial transmission and speech codecs. Int. J. Innov. Comput. Inf. Control **6**(3(B)), 1389–1400 (2010)
15. Kang, X., Yang, R., Huang, J.: Geometric invariant audio watermarking based on an LCM feature. IEEE Trans. Multime'd **13**(2), 181–190 (2011)
16. No, Really,"Rust". http://www.jamendo.com/en/album/7365
17. Thiede, T., Treurniet, W.C., Bitto, R., Schmidmer, C., Sporer, T., Beerens, J.G., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K., Feiten, B.: PEAQ—The ITU standard for objective measurement of perceived audio quality. J. AES **48**(1/2), 3–29 (2000)
18. Stirmark Benchmark for Audio. http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php
19. Wang, X.Y., Zhao, H.: A novel synchronization invariant audio watermarking scheme based on DWT and DCT. IEEE Trans. Signal Process. **54**(12), 4835–4840 (2006)
20. Megías, D., Herrera-Joancomartí, J., Minguillón, J.: Total disclosure of the embedding and detection algorithms for a secure digital watermarking scheme for audio. 7th International Conference on Information and Communication Security, ICICS 2005. Volume 3783 of Lecture notes in computer science, pp. 427–440, Springer (2005)
21. Ko, B.S., Nishimura, R., Suzuki, Y.: Log-scaling watermark detection in digital audio watermarking. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04), **3**, pp. 81–84 (2004)
22. Unoki, M., Hamada, D.: Method of digital-audio watermarking based on cochlear delay characteristics. Int. J. Innov. Comput. Inf. Control **6**(3(B)), 1325–1346 (2010)
23. Kondo, K., Nakagawa, K.: A digital watermark for stereo audio signals using variable inter-channel delay in high-frequency bands and its evaluation. Int. J. Innov. Comput. Inf. Control **6**(3(B)), 1209–1220 (2010)
24. Gulbis, M., Muller, E., Steinebach, M.: Content-based audio authentication watermarking. Int. J. Innov. Comput. Inf. Control **5**(7), 1883–1892 (2009)
25. Burnett, I.S., Pereira, F., Van de Walle, R., Koenen, R.: The MPEG-21 book, Wiley (2006)
26. Xu, C.S., Feng, D.D.: Robust and efficient content-based digital audio watermarking. Multimedia Syst. **8.5**, 353–368 (2002)
27. Peinado, M., Petitcolas, F.A.P., Kirovski, D.: Digital rights management for digital cinema. Multimedia Syst. **9.3**, 228–238 (2003)
28. http://www.jamendo.com/en/
29. http://www.opticom.de/products/opera-demoversion.html
30. Lie, W.N., Chang, L.C.: Robust and high-quality time-domain audio watermarking subject to psychoacoustic masking. The 2001 IEEE International Symposium on Circuits and Systems, 2001. ISCAS 2001, vol. 2, IEEE (2001)
31. Cléo, B., Moreau, N., Dymarski, P.: Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system. IEEE Transactions on Audio, Speech, and Language Processing, 14.5, pp. 1772–1782 (2006)