International Journal of Innovative Computing, Information and Control
Volume **8**, Number **4**, April **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 2477–2489

HIGH CAPACITY ROBUST AUDIO WATERMARKING SCHEME BASED ON FFT AND LINEAR REGRESSION

Mehdi Fallahpour and David Megias

Estudis d'Informàtica, Multimèdia i Telecomunicació
Internet Interdisciplinary Institute
Universitat Oberta de Catalunya
Rambla del Poblenou, 156, Barcelona 08018, Spain
{MFallahpour; DMegias}@uoc.edu

Received December 2010; revised April 2011

Abstract. *This paper proposes a novel high capacity audio watermarking algorithm to embed data and extract them in a bit-exact manner by changing some of the magnitudes of the FFT spectrum. The key idea is to divide the FFT spectrum into short frames and change the magnitudes of the selected FFT samples using linear regression and the average of the samples of each frame. Using the average of FFT magnitudes leads to improved robustness, since this variable is more invariant against manipulations compared with the magnitudes of single samples. In addition, linear regression helps to minimize the alterations of FFT samples, which results in better transparency. Apart from very remarkable capacity, transparency and robustness, this scheme provides three parameters which facilitate the regulation of these properties. The experimental results show that the method has a high capacity (0.5 to 2.3 kbps), without significant perceptual distortion (ODG is about −1) and provides robustness against common audio signal processing such as echo, added noise, filtering and MPEG compression (MP3).*
Keywords: Audio watermarking, Multimedia security

1. **Introduction.** The growth of the Internet, sudden production of low-cost and reliable storage devices, digital media production and editing technologies have led to widespread forgeries of digital documents and unauthorized sharing of digital data. As a result, the music industry alone claims multi-billion illegal music downloads on the Internet every year. Thus, it is vital to develop robust technologies to protect copyrighted digital media from illegal sharing and tampering.

Considering the embedding domain, audio watermarking techniques can be classified into time domain and frequency domain methods. In frequency domain watermarking [1-9,14-16], after taking one of the usual transforms such as the Discrete/Fast Fourier Transform (DFT/FFT) [4-6,17], the Modified Discrete Cosine Transform (MDCT) or the Wavelet Transform (WT) from the signal [7,9,14-16], the hidden bits are embedded into the resulting transform coefficients.

In addition to wavelet transform, [14] uses optimization for quantization, [15] takes advantage of the patchwork embedding method, and normalized energy is used in [16].

To design an audio watermarking system, different properties of the audio signal can be exploited. [10] uses properties of the human cochlear system, and [11] embeds data using delay values of the high-frequency channel signals. Using sub-band amplitude modulation, [12] results in a low bit-rate robust scheme. Finally, [13] introduces a content-based authentication watermarking technique.

In frequency domain schemes, the Fourier transform (FT) is very popular. Among different Fourier transform, the Fast Fourier transform (FFT) is often used due to its

reduced computational burden and it has been the chosen transform for the proposed scheme. This transform is also used by different authors, such as in [17], which proposes a multi-bit spread-spectrum audio watermarking scheme based on a geometric invariant log coordinate mapping (LCM) feature. The watermark is embedded in the LCM feature, but it is actually embedded in the Fourier coefficients which are mapped to the feature via LCM. Consequently, the embedding is actually performed in the FT domain.

In [4-6], which were proposed by the authors of this paper, the FFT domain is also selected to embed watermarks to take advantage of the translation-invariant property of the FFT coefficients to resist small distortions in the time domain. In fact, using methods based on transforms provides better perceptual quality and robustness against common attacks at the price of increasing the computational complexity with respect to time domain approaches.

In the algorithm suggested in this paper, we select a frequency range – or frequency band – of the FFT spectrum for embedding the secret bits. The selected band is divided into short frames and a single secret bit is embedded into each frame. The linear regression value of the FFT magnitudes in each frame must be computed and depending on the corresponding secret bit to be embedded, all samples in each frame are changed. If the secret bit is "0", all FFT magnitudes should be changed by the corresponding linear regression value. If the secret bit is "1", we divide the FFT samples into two groups and we change the magnitude of the first group by adding a quantity consisting of a scaling factor, α, which multiplies the average of all samples, to the linear regression of the corresponding sample. For the second group, we subtract a quantity resulting of the same scaling factor, α, which multiplies the average of all samples, from the linear regression of the corresponding sample. These changes in embedding "0" or "1" are performed in such a way that the average and the regression line of the FFT frame's magnitudes are independent of the embedded secret bit.

As mentioned above, the FFT is used to design a scheme in many watermarking systems. The novelty of the proposed method is the use of linear regression and the average of the magnitudes of a frame which is very useful to increase transparency and robustness against attacks, whereas embedding a secret bit into a single FFT sample is usually very fragile. In addition, the use of FFT magnitudes results in better robustness against attacks compared to the use of the real or the imaginary parts only.

The experimental results show that this method achieves a high capacity (from about 0.5 to 2.3 kbps), provides robustness against common signal processing attacks (even for strong disturbances) and entails very low perceptual distortion.

The rest of the paper is organized as follows. In Section 2, the proposed scheme is presented. Section 3 introduces the parameters and properties of the scheme. In Section 4, the experimental results are shown. Finally, Section 5 summarizes the most relevant conclusions of this work and outlines some ideas for future research.

2. **Proposed Scheme.**

In this scheme, the following method is used to embed a bit stream (secret bits) into the selected FFT magnitudes. Firstly, according to the desired capacity, transparency and robustness properties, three tuning parameters (frequency band, frame size and scaling factor) must be selected.

The selected band is divided into short frames and the linear regression of the FFT magnitudes of each frame is calculated. Each single secret bit of the stream is embedded into a frame. The average of the linear regression of FFT magnitudes of a frame plays a key role in the embedding and extracting processes. In the embedding process, all FFT samples in a frame are changed with a value related to the linear regression value and the average of the magnitude values depending on the secret bit. In the extracting process,

the secret bits are detected by using the linear regression values and the average of each frame in the decoder. Using the average of FFT magnitudes leads to improved robustness, since the average of the frame is more invariant to signal processing operations than FFT magnitudes themselves. Furthermore, using linear regression minimizes the changes in the FFT values, which results in better transparency. In addition, we have chosen the FFT domain to embed the hidden data in order to exploit the translation-invariant property of the FFT transform which allows small distortions in the time domain to be resisted. Compared to other schemes, such as quantization or odd/even modulation, to keep the relationship of neighboring FFT coefficients is a more realistic approach under several distortions.

In the last few years, an extensive work has been performed in understanding the characteristics of the human auditory system (HAS) and applying this knowledge to audio compression and audio watermarking. Human beings tend to be more sensitive towards frequencies in the range from 1 to 4 kHz. Taking the HAS into account, the human auditory sensitivity in higher frequencies is smaller than in lower frequencies. It is thus clear that, by embedding data in the middle and high frequency bands, as suggested in the proposed scheme, the distortion will be mostly inaudible and thus more transparency can be achieved.

2.1. **Linear regression.** Linear regression analyzes the relationship between two variables, X and Y. For each subject (or experimental unit), both X and Y are known. The best straight line through the data must be found. In some situations, the slope and/or the intercept of such straight line have a scientific or physical meaning. In other cases, the linear regression line is just used as a standard curve to find new values of X from Y, or vice versa. In general, the goal of linear regression is to find the line that best predicts Y from X. Linear regression does this by finding the line that minimizes the sum of the squares of the vertical distances of the data points from the line.

The idea behind the proposed scheme is to use the linear regression of a set of samples to embed a secret bit into a frame. As Figure 1 shows, to embed a zero, we just change the FFT magnitudes to the linear regression values of the FFT magnitudes. On the other hand, to embed a one, we modify the regression values and use them as marked FFT samples, as shown in Figure 1(c). To change the linear regression values, the following constraints must be observed:

1. The regression line after embedding must be the same as the regression line without changes. Since, in the extraction process, ones and zeroes are extracted depending on the distances from the regression line, the line must be kept invariant such that the distance from it can be measured after embedding.
2. The average of the modified FFT samples should not be changed since this value is used as a threshold in the decoder.

It can be easily shown that a sufficient condition for these constraints to hold is that the variations in the FFT magnitudes are symmetric and the sum of these variations is zero. Figure 1(c) shows symmetric changing of the FFT magnitudes such that the sum of all variations is zero, which satisfies the demands. In other words, the regression line after embedding is the same as the line without changes. In addition, the average of the FFT magnitudes is the same when we have symmetric changes if the sum of those variations is zero.

2.2. **Embedding the secret bits.** There are three adjustable (tuning) parameters of the suggested embedding scheme, namely the frequency band chosen for embedding the secret bits, a scaling factor (α) and the frame size (d). These parameters need to be

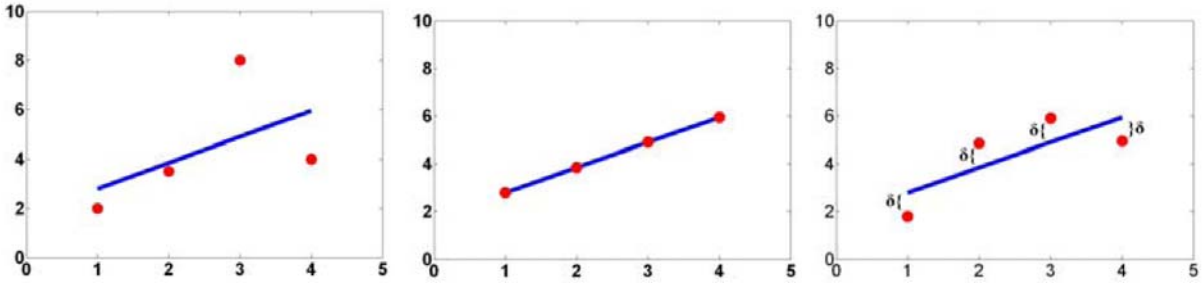


FIGURE 1. (a) Original FFT samples; (b) Marked FFT samples after embedding "0"; (c) Marked FFT samples after embedding "1"

adjusted according to the requirements of the watermarking scheme in terms of capacity, imperceptibility and robustness. In this section, for simplicity, we do not consider the regulation of these parameters and just use them as fixed. The effect of varying these parameters is analyzed in Section 3.

The embedding steps are described below:

1. Calculate the FFT of the audio signal. We can use the whole file (for short clips, e.g., with less than one minute) or blocks of a given length (e.g., 10 seconds) for longer files.
2. Divide the FFT samples in the selected frequency band into frames of size d.
3. Compute the linear regression of the magnitudes of the FFT samples for each single frame, where r_j is the regression value of f_j.
4. Calculate the average of the regression of the magnitudes of the FFT samples in each frame by using Equation (1).

$$m_i = \frac{1}{d} \sum_{j=(i-1)d+1}^{id} r_j \tag{1}$$

where $\{r_j\}$ are the linear regressions of the FFT magnitudes of the selected frequency band, d is the frame size and m_i is the average of the i^{th} frame. Note that this equation could be simplified, since $\{r_j\}$ are points in a straight line and, thus, a closed equation for m_i can be easily obtained.

5. The marked FFT magnitudes $\{f_j^{'}\}$ are obtained by using Equations (2) and (3). If d is even Equation (2) should be used and, for odd values of d, Equation (3) is applied.

$$f_j^{'} = \begin{cases} r_j - \alpha m_i & \text{if } \bmod(j-1,2) = 1,\ \bmod(j,d) < \frac{d}{2},\ w_l = 1 \\ r_j + \beta m_i & \text{if } \bmod(j-1,2) = 0,\ \bmod(j,d) < \frac{d}{2},\ w_l = 1 \\ r_j + \beta m_i & \text{if } \bmod(j-1,2) = 1,\ \bmod(j,d) \geq \frac{d}{2},\ w_l = 1 \\ r_j - \alpha m_i & \text{if } \bmod(j-1,2) = 0,\ \bmod(j,d) \geq \frac{d}{2},\ w_l = 1 \\ r_j & \text{if } w_l = 0 \end{cases} \tag{2}$$

$$f_j^{'} = \begin{cases} r_j - \alpha m_i & \text{if } \bmod(j,2) = 1,\ w_l = 1 \\ r_j + \beta m_i & \text{if } \bmod(j,2) = 0,\ w_l = 1 \\ r_j & \text{if } \bmod(j,d) = (d+1)/2,\ w_l = 1 \\ r_j & \text{if } w_l = 0 \end{cases} \tag{3}$$

for

$$\beta = \begin{cases} \alpha & \text{if } d = 4n \text{ or } d = 4n+1 \\ \left(\frac{n+1}{n}\right)\alpha & \text{if } d = 4n+2 \text{ or } d = 4n+3 \end{cases}$$

where $i = \lfloor j/d \rfloor + 1$, w_l is the l^{th} bit of the secret stream, $0 < \alpha < 1$ is a scaling factor and mod denotes the residual function. Each secret bit is embedded into a suitable frame and thus, after embedding the bit, the index l is increased and the next secret bit is embedded into the next suitable frame.

6. In the previous embedding steps, the FFT phases are not altered. The marked audio signal in the time domain is obtained by applying the inverse FFT with the new magnitudes and the original FFT phases.

Figure 1 shows the FFT samples of a frame of size 4. Figure 1(a) depicts the original FFT samples before modification. Figure 1(b) shows embedding "0", where all samples are changed by the linear regression value of the corresponding sample. Figure 1(c) illustrates that, to embed "1", half of samples are changed to $r_j - \alpha m_i$ and the others to $r_j + \alpha m_i$ in a symmetric way.

2.3. **Extracting the secret bits.** The watermark extraction is performed by using the FFT transform and the tuning parameters, which can be considered as side information. The scaling 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 adjustable parameters. Then, in the decoder, the tuning parameters would be extracted by using the default parameters and, later on, the secret bits would be obtained using the extracted values of the adjustable parameters. Since 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. Calculate the FFT of the marked audio signal.
2. Divide the FFT samples in the selected frequency band into frames of size d.
3. Compute the linear regression of the FFT magnitudes for each single frame, where r'_j is the linear regression value of f'_j.
4. Compute the average of the regression values of the magnitudes of the marked FFT samples in each frame by Equation (4). Again, as in the embedding method, a closed equation can be obtained for this average.

$$m'_i = \frac{1}{d} \sum_{j=(i-1)d+1}^{id} r'_j \tag{4}$$

5. 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 extracted. If $|f'_j - r'_j| < \alpha m'_i/2$, then we consider f'_j as a sample in a zero frame ("0" embedded) otherwise it is considered as a one frame ("1" embedded).

After getting information about all samples, depending on the number of samples which represent "0" or "1" (voting scheme) a secret bit can be extracted. If the number of samples which present "0" is equal or larger than half the frame size, then the extracted secret bit is "0", otherwise it is "1".

To increase security, pseudo-random number generators (PRNG) can be used to change the secret bit stream to another stream which makes it more difficult for an attacker to extract the secret information. For example, the embedded bit stream can be constructed

as the XOR sum of the real watermark and a pseudo-random bit stream. The seed of the PRNG would be required as a secret key both at the sender and the detector [23].

3. **Parameters and Properties.** The watermarking process should not introduce any perceptible artifacts into the original contents (e.g., an audio signal). Ideally, there must be no perceptible difference between the watermarked and the original digital contents; i.e., the watermark data should be embedded imperceptibly into the audio media. Apart from imperceptibility, capacity and robustness are two other fundamental properties of audio watermarking schemes. Capacity is defined as the number of bits which are embedded in a second of the audio file. The watermark should be extractable after various intentional and unintentional attacks. These attacks may include additive noise, re-sampling, MP3 compression, low-pass filtering, and any other attack which may remove the watermark or confuse the watermark extraction system. Considering a trade-off between capacity, transparency and robustness is the main challenge for audio watermarking applications; i.e., in an ideal case, we would demand a very high capacity, transparent and robust scheme but, in practice, obtaining all these properties at the same time is extremely difficult or even impossible. Thus, depending on the priorities or the particular application, a trade-off between these properties must be attained.

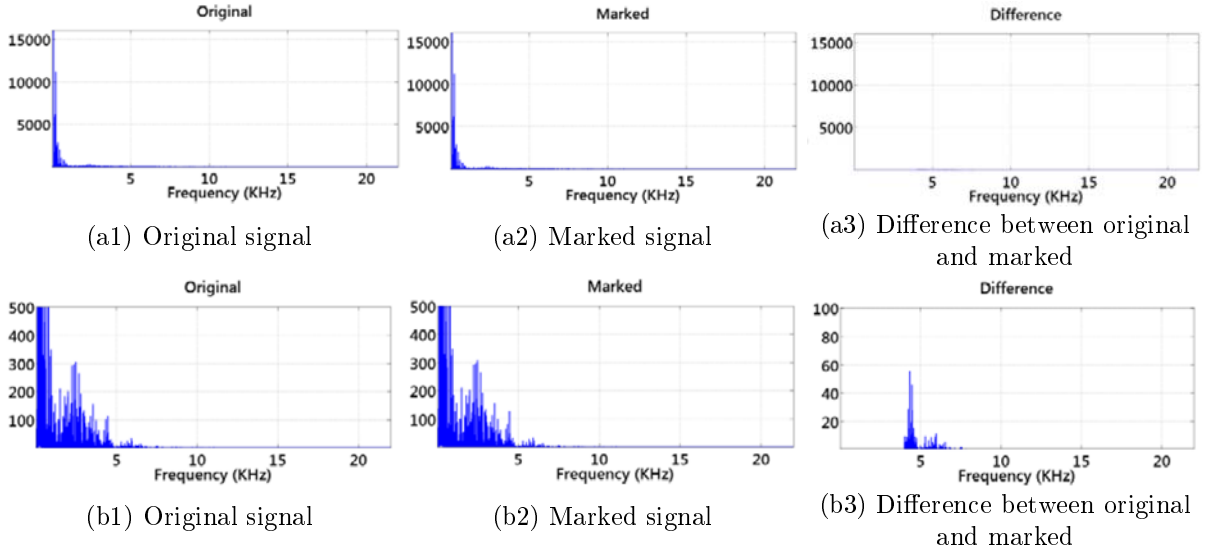
In the proposed scheme, there are three parameters which make it possible to regulate the properties of the watermarking system, namely:

1. Frequency band: the embedding area is called frequency band. The frequency band has a direct effect on capacity. In fact, each frame is useful for embedding a secret bit. Hence, the number of frames in the selected frequency band should be equal to demanded capacity. Furthermore, to attain better robustness, we can check the effect of a specific attack on the FFT spectrum. For example, Figures 2(b2) and 2(b3) show that 0–5 kHz is a very suitable area when robustness against resampling to 22 kHz is needed. However, this does not mean that only this band is useful for embedding, since other frequency areas can be used for embedding information too.
2. Frame size (d): the frequency band is divided into frames and each single frame is used for embedding a single secret bit. Thus, the larger the frame size, the smaller capacity results. On the other hand, increasing the frame size results in better robustness. To embed a secret bit into a frame, all samples of the frame are modified. For example, it is evident that changing four FFT samples to embed a bit leads to more robustness than changing just one or two samples, though imperceptibility will decrease.
3. Scaling factor (α): this parameter is defined to manage the robustness-imperceptibility trade-off of the scheme. In fact, by increasing the scaling factor we get better robustness and more distortion. The experimental results show the effect of α on transparency and robustness. However, we cannot assert that increasing α always results in better robustness and more distortion, since it depends on the particular audio file and its spectrum.

Usually, introducing general tuning rules which can be applied to all audio files is difficult, since different audio files have different spectral behavior. However, the ideas pointed out above are a good summary of the tuning guidelines for the different parameters.

To illustrate the transparency of the proposed scheme, Figure 2 depicts a plot of the frequency spectrum of the original, the marked signals and the difference between them. The original signal is "Beginning of the End" [19] and, to mark it, frequency samples between 4 to 8 kHz and $\alpha = 0.5$ were selected. Figure 2(a1) shows the magnitudes of the original signal, Figure 2(a2) those of the marked signal and Figure 2(a3) the difference between them in the original scale, without zoom. As Figure 2(a3) shows, the difference

is very low. Figure 2(b), zooms-in these plots, helping to visualize the value of the FFT magnitudes. Figure 2(b3) presents the tiny difference that occurs as a consequence of the embedding process.



(a1) Original signal (a2) Marked signal (a3) Difference between original and marked

(b1) Original signal (b2) Marked signal (b3) Difference between original and marked

FIGURE 2. Effect of embedding algorithm on FFT spectrum of marked signal

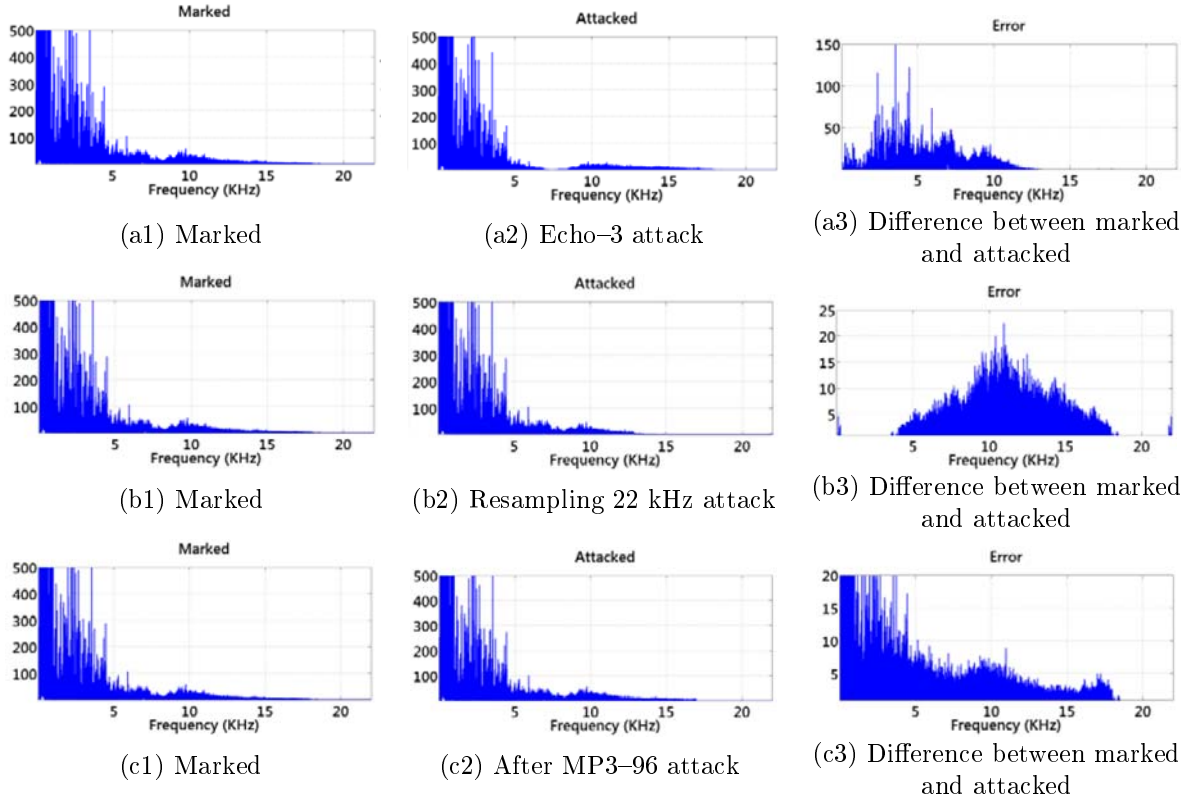
Apart from remarkable transparency, the proposed scheme is robust against several attacks. Figure 3 shows the effect of three different attacks (Echo–3, Resampling 44–22–44, MP3–96) on the frequency spectrum of the marked signal shown in Figure 2. In Figure 3, the first column shows the marked signal, the second one shows the attacked signal and the third one is the difference between the marked and the attacked signals. The third column shows a considerable distortion caused by the attacks. Despite the distortion introduced by the attacks on the marked signal, the secret information can still be extracted as shown in the next section.

4. **Experimental Results.** To evaluate the performance of the proposed method and to discuss the applicability of the scheme in a real scenario, the songs "Beginning of the End", "Breathing On Another Planet", "Do You Know Where Your Children Are" and "Floodplain" included in the album *Rust* by No, Really [11] have been selected. All audio clips are sampled at 44.1 kHz with 16 bits per sample and two channels. The experiments have been performed for each channel of the audio signals separately.

Considering a trade-off between capacity, transparency and robustness is the main challenge for audio watermarking applications. The following conditions were assumed to obtain different capacity, transparency and robustness scenarios:

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 compared to the previous scenario.
3. Robustness against many attacks with a wide range of changes is desirable. This is more complicated than the previous two scenarios, since we need robustness against most varied attacks. Thus, according to the trade-off between capacity, transparency and robustness, a sacrifice in capacity and transparency is required.

The Objective Difference Grade (ODG) has been used to evaluate the transparency of the proposed algorithm. The ODG is one of the output values of the ITU-R BS.1387



(a1) Marked (a2) Echo–3 attack (a3) Difference between marked and attacked

(b1) Marked (b2) Resampling 22 kHz attack (b3) Difference between marked and attacked

(c1) Marked (c2) After MP3–96 attack (c3) Difference between marked and attacked

FIGURE 3. Attacks' effect on the FFT spectrum of the marked signal

PEAQ [20] standard, where ODG = 0 means no degradation and ODG = −4 means a very annoying distortion. Additionally, the OPERA software [21] based on the referred standard (PEAQ Advanced) has been used to compute this objective measure of quality.

To test the resistance property against malicious attacks or signal modifications, after the embedding process of watermarking, some attacks have been applied. The robustness of our method has been evaluated using the bit error rate (BER) defined as follows:

$$BER = \frac{\text{number of error bits}}{\text{number of total bits}}$$

Table 1 shows the perceptual distortion, in SNR and ODG, payload, in bits per second (bps), and BER under the MP3 compression attack with different bit rates. Note that different values for the parameters are used to achieve different trade-off solutions between capacity, transparency and robustness, as usual for all watermarking systems. For example, for the "Beginning of the End" file, by decreasing the scaling factor (α), in rows 1 and 2, robustness against MP3–96 is decreased whilst improving transparency. Comparing rows 2 and 8 shows that changing the frequency band can increase the robustness in trade-off with capacity. Moreover, rows 8 and 9 show that, by increasing the scaling factor, we can get robustness against MP3–64.

As explained in Section 3, there are three adjustable parameters and audio watermarking schemes are evaluated in terms of three main properties. Thus, we have three inputs (parameters) and three outputs (properties) for a nonlinear system which works according to the human auditory system. Finding specific functions to adjust the requirements using the input parameters is extremely difficult and sometimes impossible. We can just use different loops and conditions to get better results in an adaptive way. In this scheme, we have general tuning guidelines which help to reach the requirements or to get close to

them very quickly. The frame size has more effect on the robustness and, for the experiments, we have used a frame size equal to 4 which produces convenient results. In other words, by increasing the frame size better robustness is achieved at the price of decreasing capacity.

TABLE 1. Results of 4 mono signals (robust against Table 2 attacks)

Audio File	*Time (m:sec)*	*Row number*	*Scaling factor*	*Frequency band (KHz)*	*SNR (dB)*	*MP3 Attack*		*ODG of marked*	*Payload (bps)*
						Rate	*BER*		
Beginning of the End	3:16	1	0.5	4–8	39	96	0.02	–0.83	1012
		2	0.25	4–8	41.8	96	0.07	–0.35	1012
		3	0.75	4–8	36.8	96	0.01	–1.43	1012
		4	0.5	6–15	47.2	128	0.01	–1.2	2262
		5	0.25	6–15	52.1	128	0.04	–0.51	2262
		6	0.75	6–15	44	128	0.01	–1.90	2262
		7	0.5	2–4	28	96	0.002	–0.35	512
		8	0.25	2–4	31	96	0.01	–0.19	512
		9	0.75	2–4	25	**64**	0.08	–0.90	512
Breathing On Another Planet	3:13	10	0.25	10–16	45	128	0.02	–0.94	1512
Do You Know Where Your ...	2:31	11	0.25	4–8	33	128	0.04	–0.34	1012
Floodplain	3:13	12	0.25	2–4	23	80	0.08	–0.12	512
		13	0.5	2–4	27	**64**	0.08	–0.27	512

TABLE 2. Robustness test results for two selected files

Attack name	*Beginning of the End*			*Floodplain*		
	ODG of attacked file	*Parameters*	*BER*	*ODG of attacked file*	*Parameters*	*BER*
AddBrumm	–2.5	1–5k, 1–5k	0.0–0.03	–3.1	1–8k, 1–8k	0.0–0.01
AddDynNoise	–3	1–2	0.04–0.08	–3.7	1–20	0.0–0.05
AddNoise	–0.5	1–70	0.0–0.07	–1.3	1–1000	0.0–0.03
AddSinus	–2.5	1–5k, 1–5k	0	–2.8	1–5k, 1–5k	0
Amplify	–0.2 – 0	20–160	0–0.02	–0.9	1–130	0.0–0.04
BassBoost	–3.2	0–25, 0–50	0–0.03	–3.6	0–45, 0–60	0.0–0.01
Echo	– 3.7	0–1.5k	0–0.15	–3.13	0–2.5k	0.0–0.02
FFT_RealReverse	–2.5	2	0.11	–3.3	2,4	0.02–0.04
FFT_Stat1	–2.7	2	0.13	–3.2	2,4	0.02–0.05
Invert	–2.6	–	0.02	–3.6	–	0.00
LSBZero	–0.2	–	0.02	–0.1	–	0.00
RC_HighPass	–0.1 to –3.4	0–18k	0–0.02	–2.4	0–18k	0.0–0.01
RC_LowPass	–0.5 to –3.4	1k–20k	0.04–0.0	–3.2	0.1k–20k	0.0–0.01
Smooth	–1.5	–	0.15	–2.7	–	0.01
Smooth2	–3.5	–	0.14	–3.3	–	0.01
Stat1	–0.2	–	0.02	–0.4	–	0.01
Stat2	–0.2	–	0.02	–1.6	–	0.01
MP3	–0.1	256	0.0	–0.0	256	0.0
	–0.2	128	0.0	–0.2	128	0.0
	–0.4	96	0.02	–0.5	96	0.005
	–0.6	80	0.06	–0.6	80	0.03
	–0.7	64	0.11	–1.0	64	0.08
Resampling	–1.2	44–22–44	0.02	–1.2	44–22–44	0.01
				–3.1	44–11–44	0.01

In addition, increasing the frequency band leads to better capacity. Finally, with the scaling factor, robustness and transparency can be regulated. Note that these parameters

allow to regulate the ODG between 0 (not perceptible) and –1 (not annoying), with about 1 to 2 kbps capacity allowing robustness against MP3–128, which are typical requirements.

To check the robustness of the proposed scheme against different attacks, we have used the Stirmark Benchmark for Audio (SMBA) v1.0 [22] which provides a large number of attacks. Table 2 illustrates the effect of several common attacks, provided by the SMBA, on ODG and BER for the two selected audio test files. The parameters were selected for each signal, and, then, the embedding method was applied. For the "Beginning of the End" file, the scaling factor has been chosen as 0.5 and the frequency band is 4–8 kHz. On the other hand, for "Floodplain", the scaling factor is 0.5 and the frequency band is 2–4 kHz. The SMBA software has been used to attack the marked files and, finally, the detection method has been applied for the attacked files. The ODG in Table II is calculated between the marked and the attacked-marked files. The parameters of the attacks are defined in the SMBA web site [22]. For example, in AddBrumm, 1–5k shows the strength and 1–5k shows the frequency. This row illustrates that any value in the range 1–5k for the strength and 1–5k for the frequency could be used without any significant change in BER. 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).

In order to reduce computation time and memory usage, audio files can be divided into small clips, e.g., 10 seconds each. Then, the synchronization method described in [24] can be applied for each clip separately. A MATLAB implementation of the scheme on a Computer with 2.4–GHz Intel Core 2 duo CPU has been used to embed and extract the secret information bits. The computation time for audio signals with less than 30 seconds is about 50% to 80% of the playing time, which proves that this scheme can be used for real-time applications. In a practical implementation of the extracting algorithm, the computation time can be less than 50% of the playing time.

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 fair comparison of the proposed scheme with some fragile audio watermarking schemes which are similar to it as capacity is concerned. Hence, we have chosen a few recent and relevant robust audio watermarking schemes in the literature. In Table 3, we compare the performance of the proposed watermarking algorithm and several recent audio watermarking strategies robust against the MP3 attack. In this table, low capacity schemes (with less than 1 kbps) are shown in the first few rows, whereas high capacity schemes (more than 1 kbps) are summarized in the last few rows[1]. The analyzed methods are the following:

- [1,2,8] have low capacity but are robust against common attacks. [1] proposes a very robust, low capacity and high distortion scheme.
- [3] evaluates distortion by using the mean opinion score (MOS), which is a subjective measurement, and achieves transparency between imperceptible and perceptible but not annoying, MOS = 4.7. The method described [3] and the scheme proposed in this paper lead to high capacity and low distortion but they are not as robust as the low-capacity method described in [1].
- [10] proposes a technique based on cochlear delay characteristics which is robust against MPEG compression and resampling.
- [11] describes a very robust scheme against resampling and compression but it has a very low capacity (7–30 bps). The quality of the marked signal is source-dependent;

[1]Please note the "k" denoting "kilo" in the capacity value of the last few rows.

i.e., for some audio signals the quality of the marked signal is good and for others is significantly low.

- Speech applications and codecs are considered in [12]. In that paper, the distortion introduced to the marked signal is slightly annoying, capacity is very low and robustness is achieved against compression attacks.
- Using content features in [13] results in 80 bps in term of capacity, -1 in terms transparency (ODG) and robustness against a few attacks such as compression.
- Recently, [17] 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.
- [4], which was also proposed by the authors of this paper, has 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. For example, in the proposed scheme, by using a frame size of 8, makes it possible to achieve robustness against MP3–64; whereas, in [4], low bit rate MP3 compression was not considered.
- The methods [6,9], also proposed by the authors of this paper, have high capacity but they are not too robust against attacks compared with the scheme proposed in this paper.

In addition to Table 3, which compares capacity and transparency of a large number of published papers, Table 4 compares the robustness of high capacity schemes [4,6,9] with the proposed method. In fact, the most valuable achievement of the proposed scheme is robustness against difficult attacks such as Echo with a wide range of parameters. For example, [9] is robust against Echo in the range [1, 5], whereas the proposed method is robust against Echo in the range [1, 1500]. Furthermore, the proposed scheme is also robust against MP3–64.

This comparison shows the superiority in both capacity and imperceptibility of the suggested method with respect to other schemes in the literature, and in robustness as schemes with similar capacity and imperceptibility are concerned. 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. Conclusion.

In this paper, a high-capacity watermarking algorithm for digital audio, which is robust against common audio signal processing attacks and the Stirmark Benchmark for audio, is presented. A scaling factor, the frame size and the selected frequency band are the three adjustable parameters of this method which regulate the capacity, the perceptual distortion and the robustness trade-off of the scheme accurately. Furthermore, the suggested scheme is blind, since it does not need the original signal for extracting the hidden bits. The experimental results show that this scheme has a high capacity (0.5 to 2.3 kbps) without significant perceptual distortion (ODG greater than –1) and provides robustness against common signal processing attacks such as echo, added noise, filtering or MPEG compression (MP3). In addition, the proposed method clearly overcomes the robustness results of recent methods which can be compared with it in terms of capacity.

An open problem which is not often considered is modern attacks, such as compression, channel fading, jitter and packet drop. Recently, with rapid production of digital media,

TABLE 3. Comparison of transparency and capacity

	Algorithm	***Capacity (bps)***	***Imperceptibility in SNR (dB)***	***Imperceptibility (ODG)***
Low capacity schemes (< 1 kbps)	[1]	2	42.8 to 44.4	–1.66 < ODG < –1.88
	[2]	4.3	29.5	Not reported
	[3]	689	Not reported	Not reported
	[8]	2.3	Not reported	Not reported
	[10]	4–512	Not reported	–1 < ODG
	[11]	7–30	Not reported	Not reported
	[12]	8	Not reported	–3 < ODG < –1
	[13]	80	Not reported	–1.04
	[17]	64	30–45	–1< ODG
High capacity schemes (≥ 1 kbps)	[4]	3k	30.55	–0.6
	[6]	1.5k–8.5k	35–45	–0.8 < ODG < –0.1
	[9]	11k	30	–0.7
	Proposed	512–2.3k	29 to 46	–0.1 < ODG < –1.5

TABLE 4. Comparison of robustness

Algorithm	***Echo***		***MP3***		***MP3***		***Resampling***	
	Parameter	***BER***	***Parameter***	***BER***	***Parameter***	***BER***	***Parameter***	***BER***
[4]	1–10	0.005	128	0.03–0.05	64	N.A.	44–22–44	0.04
[6]	N.A.	–	128	0	64	N.A.	N.A.	N.A.
[9]	1–5	0.01–0.03	128	0–0.02	64	N.A.	44–22–44	0.07–0.11
Proposed	**0–1.5k**	**0–0.15**	**128**	**0**	**64**	**0.11**	**44–22–44**	**0.01–0.02**

these modern attacks are becoming more important. These attacks are particularly relevant in various networks such as GSM and the Internet. For future work, we will consider this particular problem.

Acknowledgment. This work is partially supported by the Spanish Ministry of Science and Innovation and the FEDER funds under the grants TSI2007-65406-C03-03 E-AEGIS, CONSOLIDER-INGENIO 2010 CSD2007-00004 ARES and TIN2011-27076-C03-02 CO-PRIVACY.

REFERENCES

[1] S. Xiang, H. J. Kim and J. Huang, Audio watermarking robust against time-scale modification and MP3 compression, *Signal Processing*, vol.88, no.10, pp.2372-2387, 2008.

[2] M. Mansour and A. Tewfik, Data embedding in audio using time-scale modification, *IEEE Trans. Speech Audio Processing*, vol.13, no.3, pp.432-440, 2005.

[3] J. J. Garcia-Hernandez, M. Nakano-Miyatake and H. Perez-Meana, Data hiding in audio signal using rational dither modulation, *IEICE Electron. Express*, vol.5, no.7, pp.217-222, 2008.

[4] M. Fallahpour and D. Megías, High capacity audio watermarking using FFT amplitude interpolation, *IEICE Electron. Express*, vol.6, no.14, pp.1057-1063, 2009.

[5] M. Fallahpour and D. Megías, High capacity method for real-time audio data hiding using the FFT transform, *Advances in Information Security and Its Application*, pp.91-97, 2009.

[6] M. Fallahpour and D. Megías, Robust high-capacity audio watermarking based on FFT amplitude modification, *IEICE Trans. on Information and Systems*, vol.E93-D, no.01, pp.87-93, 2010.

[7] M. Fallahpour and D. Megías, DWT-based high capacity audio watermarking, *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol.E93-A, no.01, pp.331-335, 2010.

[8] W. Li and X. Xue, Content based localized robust audio watermarking robust against time scale modification, *IEEE Trans. Multimedia*, vol.8, no.1, pp.60-69, 2006.

[9] M. Fallahpour and D. Megías, High capacity audio watermarking using the high frequency band of the wavelet domain, *Multimedia Tools and Applications, Springer*, vol.52, pp.485-498, 2011.

[10] M. Unoki and D. Hamada, Method of digital-audio watermarking based on cochlear delay characteristics, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3(B), pp.1325-1346, 2010.

[11] K. Kondo and K. Nakagawa, A digital watermark for stereo audio signals using variable inter-channel delay in high-frequency bands and its evaluation, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3(B), pp.1209-1220, 2010.

[12] A. Nishimura, Audio data hiding that is robust with respect to aerial transmission and speech codecs, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3(B), pp.1389-1400, 2010.

[13] M. Gulbis, E. Muller and M. Steinebach, Content-based audio authentication watermarking, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1883-1892, 2009.

[14] S. T. Chen, G. D. Wu and H. N. Huang, Wavelet-domain audio watermarking scheme using optimisation-based quantisation, *IET Signal Processing*, vol.4, no.6, pp.720-727, 2010.

[15] N. K. Kalantari, M. A. Akhaee, M. Ahadi and H. Amindavar, Robust multiplicative patchwork method for audio watermarking, *IEEE Trans. on Audio, Speech, and Language Processing*, vol.17, no.6, pp.1133-1141, 2009.

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

[17] X. Kang, R. Yang and J. Huang, Geometric invariant audio watermarking based on an LCM feature, *IEEE Trans. on Multimedia*, vol.13, no.2, pp.181-190, 2011.

[18] *SQAM Sound Quality Assessment Material*, http://www-sipl.technion.ac.il/Info/Downloads_DataBases_Audio_Quality_Assessment_Readme_e.shtml, 2011.

[19] *No, Really, "Rust"*, http://www.jamendo.com/en/album/7365, 2011.

[20] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg and B. Feiten, PEAQ – The ITU standard for objective measurement of perceived audio quality, *Journal of the AES*, vol.48, no.1-2, pp.3-29, 2000.

[21] *Opticom Opera Software Site*, http://www.opticom.de/products/opera.html, 2011.

[22] *Stirmark Benchmark for Audio*, http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php, 2010.

[23] D. Megías, J. Herrera-Joancomartí and J. Minguillón, Total disclosure of the embedding and detection algorithms for a secure digital watermarking scheme for audio, *Proc. of the 7th International Conference on Information and Communication Security*, Beijing, China, pp.427-440, 2005.

[24] D. Megías, J. Serra-Ruiz and M. Fallahpour, Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification, *Signal Processing*, vol.90, no.12, pp.3078-3092, 2010.