

A Multi-Level Robust and Perceptually Transparent Blind Audio Watermarking Scheme Using Wavelets

Farooq HUSAIN⁽¹⁾, Omar FAROOQ⁽²⁾, Ekram KHAN⁽²⁾

⁽¹⁾ *Moradabad Institute of Technology*
Moradabad, India; e-mail: farooqhusain70@gmail.com

⁽²⁾ *Aligarh Muslim University*
Aligarh, India

(received March 18, 2014; accepted September 23, 2014)

In this paper, a robust and perceptually transparent single-level and multi-level blind audio watermarking scheme using wavelets is proposed. A randomly generated binary sequence is used as a watermark, and wavelet function coding is used to embed the watermark sequence in audio signals. Multi-level watermarking is used to enhance payload capacity and can be used for a different level of security. The robustness of the scheme is evaluated by applying different attacks such as filtering, sampling rate alteration, compression, noise addition, amplitude scaling, and cropping. The simulation results obtained show that the proposed watermarking scheme is resilient to various attacks except cropping. Perceptual transparency of watermark is measured by using Perceptual Evaluation of Audio Quality (PEAQ) basic model of ITU-R (PEAQ ITU-R BS.1387) on Speech Quality Assessing Material (SQAM) given by European Broadcasting Union (EBU). Average Objective Difference Grade (ODG) measured for this method is -0.067 and -0.080 for single-level and multi-level watermarked audio signals, respectively. In the proposed single-level digital audio watermarking scheme, the payload capacity is increased by 19.05% as compared to the single-level Chirp-Based Digital Audio Watermarking (CB-DAWM) scheme.

Keywords: digital audio watermarking, robustness, single-level watermarking, multi-level watermarking, payload capacity.

1. Introduction

Nowadays powerful personal computers, low-cost storage devices, such as flash drives, DVDs, etc., are easily available. Due to audio recording and editing software and broad-band internet, copying, editing, and distribution of digital media can be easily done by a person with little knowledge of computer and editing software. In such a scenario, it is very important to have means for protection and enforcement of intellectual property rights (IPRs) for multimedia contents. Digital watermarking has been proposed as a viable solution to improve multimedia security and to verify authenticity of the content while offering robustness against any attempt to alter it. Digital watermarking techniques have been used in copyright protection, content authentication, temper proofing, broadcast monitoring, and integrity of network (BARNI, BARTOLINI, 2004; BENDER *et al.*, 1996; COX *et al.*, 2002). Digital watermarking techniques can be applied to any multimedia data such as text, audio, image, and video.

Data protection methods such as steganography and cryptography are not useful in these applications as they make multimedia data imperceptible and useless. On the other hand, digital watermarking is now attracting attention for protection against unauthorized copying and distribution of multimedia data (audio, images, and video) (BARNI, BARTOLINI, 2004; COX *et al.*, 2002; LANGELAAR *et al.*, 2000; NEUBAUER, HERRE, 1998). There are three necessary requirements for any effective data hiding algorithm, such as imperceptibility (inaudibility in the case of audio and speech signals, and invisibility in the case of images and video signals), robustness against signal processing attacks, and data embedding capacity (payload). The relative importance is to be given to each of these requirements in the implementation of watermarking scheme depends on the desired application of the system. In practice, a fundamental trade-off between the three requirements in an effective watermarking method exists. However, a special attention must be given to imperceptibility (inaudibility in the

case of audio) because if the original quality of a multimedia signal cannot be preserved, then neither users nor owners will accept the watermarking technology for their applications (AL-HAJ, MOHAMMAD, 2010; XU *et al.*, 1999).

There is little research work available on digital audio watermarking as compared to watermarking of images and videos (ARNOLD *et al.*, 2003). This is due to the fact that audio signals are represented by much fewer samples per time-interval implying that a smaller number of bits of information (watermark data) can be embedded robustly in audio data as compared to the number of bits embedded in visual data.

Generally, a watermark in audio can be embedded into the time-domain (BASSIA, PITAS, 2001; COX *et al.*, 1997; KIROVSKI, MALVAR, 2003; KO *et al.*, 2005; LIE, CHANG, 2006; LILI *et al.*, 2007; OH *et al.*, 2001; XIONG, MING, 2006) or the frequency-domain (AL-HAJ, MOHAMMAD, 2010; VIERU *et al.*, 2005; WANG, ZHAO, 2006; XIE *et al.*, 2006). Some of the common time-domain watermark embedding methods are Least Significant Bits (LSB) alteration (XIONG, MING, 2006), echo addition (KO *et al.*, 2005; OH *et al.*, 2001), and spread spectrum (COX *et al.*, 1997; KIROVSKI, MALVAR, 2003; LILI *et al.*, 2007) methods. The time-domain methods embed the watermark directly into the time-domain samples. In the LSB method, watermark information bits are embedded into the LSBs of the audio signals. In the echo addition method, the watermark information bits are embedded in delayed attenuated versions of the original audio signals. In communication, the spread spectrum is used to hide a signal against an unintended listener and ensuring information privacy. The same concept is useful in digital watermarking of audio signals. In a spread spectrum watermarking, computational complexity and synchronization overhead may be unacceptably high (AL-HAJ, MOHAMMAD, 2010). Frequency-domain audio watermarking methods (AL-HAJ, MOHAMMAD, 2010; VIERU *et al.*, 2005; WANG, ZHAO, 2006; XIE *et al.*, 2006) employ human perceptual properties and frequency masking characteristics of HAS for watermarking. These techniques use different transformation tools such as Fast Fourier Transform (FFT) (XIE *et al.*, 2006), Discrete Cosine Transform (DCT) (WANG, ZHAO, 2006), and Discrete Wavelet Transform (DWT) (AL-HAJ, MOHAMMAD, 2010; QUAN, ZHANG, 2004; VIERU *et al.*, 2005; WANG, ZHAO, 2006), etc., to transform the audio signals to locate the appropriate embedding position. The time-domain methods are relatively easy to implement and computational complexity of these algorithms are lower as compared to frequency-domain algorithms. The chirp-based digital audio watermarking (CB-DAWM) schemes (BLACKLEDGE, FAROOQ, 2008; FAROOQ *et al.*, 2008a; 2008b) are also considered in the category of time-domain audio watermarking schemes.

In these watermarking schemes, a chirp-coded binary watermark is embedded into host audio signals and the watermark is extracted blindly. These watermarking schemes are robust against most of the audio signal processing attacks.

In this paper, a multi-level robust and imperceptible (inaudible) audio watermarking algorithm which uses the mother wavelet to embed a watermark is proposed. This is a blind (non-informed) watermarking scheme because the watermark extraction algorithm does not require the original host audio signal. This watermarking scheme is used for both single-level and multi-level watermark embedding. By using multi-level watermarking, an enhanced payload capacity and different level of security can be achieved. The scheme is found to be robust against different audio signal processing attacks such as low-pass filtering, upsampling, downsampling, resampling, amplitude scaling, AWGN, and MP3 compression for both single-level and multi-level watermarking. The proposed scheme shows limited robustness under high-pass and band-pass filtering operations for both single-level and multi-level schemes, however, it is not robust under cropping attacks. By using the proposed single-level Wavelet-Based Digital Audio Watermarking (WB-DAWM) scheme, the payload capacity is increased by 19.05% as compared to the single-level Chirp-Based Digital Audio Watermarking (CB-DAWM) scheme.

2. Proposed watermarking method

The high correlation property of wavelet function is exploited, and in phase and out of phase wavelet functions are embedded for '1' and '0', respectively. A wavelet is a small wave which oscillates and decays in the time-domain quickly. As compared to the sinusoidal basis function, wavelets are compact both in time and frequency. They have several families such as 'Haar', 'Meyer', 'Morlet', 'Daubechies', etc. which are fundamentally different from each other. A wavelet is defined by the wavelet function, i.e. mother wavelet $\psi(t)$ and the scaling function, i.e. father wavelet $\varphi(t)$. The main purpose of the mother wavelet is to provide a source function to generate the daughter wavelets by using the scaling function (father wavelet). By scaling and translation of these two orthogonal functions, a complete set of wavelet basis is obtained. The energies of wavelet and scaling functions are finite. The scaling function is primarily responsible for improving the coverage of the wavelet spectrum.

Ingrid Daubechies proposed a compactly supported orthogonal wavelet which is known as 'Daubechies' wavelet (DAUBECHIES, 1990). This wavelet has made discrete wavelet analysis practical. The names of the 'Daubechies' family wavelets are written dbN, where N is the order, and db is the 'Daubechies' wavelet. The number of vanishing moments of a wavelet analy-

sis represents the order of the wavelet. A wavelet has ‘m’ vanishing moments if and only if its scaling function can generate polynomials of degree smaller than or equal to m. A wavelet with a higher order will result in better signal approximations.

Figure 1 shows the time-domain plots for db5 with iterations 10, 13, and 14 used for coding watermark sequence in the proposed scheme. The central frequency of this mother wavelet using 10 iterations is 35.53 Hz, as shown in its power spectral density (PSD) plot in Fig. 2. The same mother wavelet db5 using 13 iterations and 1.6719 second duration has the central frequency of 4.441 Hz, while using 14 iterations and 3.3437 second duration has the central frequency of 2.221 Hz.

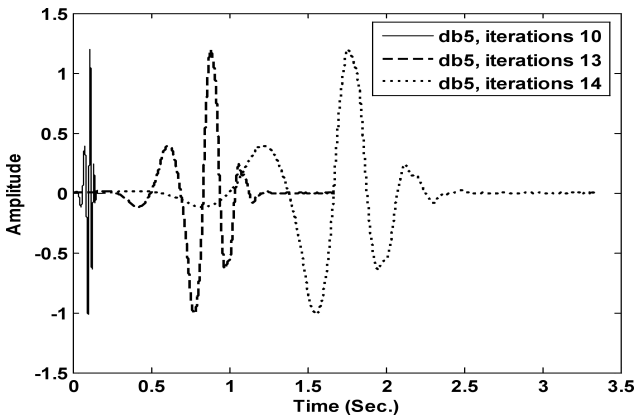


Fig. 1. Plot of mother wavelet db5 for iterations 10, 13, and 14.

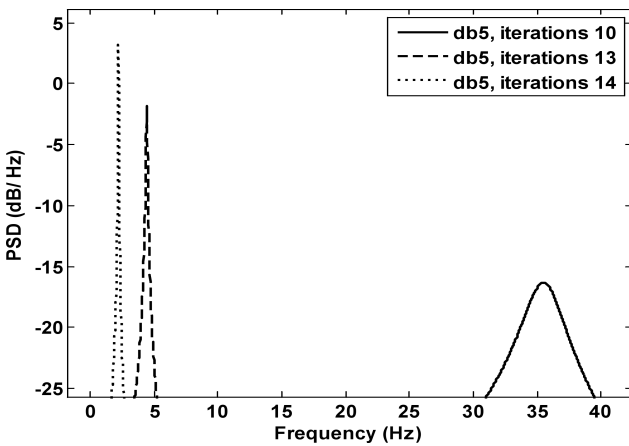


Fig. 2. PSD plots for mother wavelet db5 for iterations 10, 13, and 14.

2.1. Generation of wavelet-based watermark

In the proposed watermarking scheme, watermark data are generated using a random binary sequence which is phase-coded using a wavelet function of different orders and iterations. The purpose of the wavelet function coding is to diffuse each bit over a range of compact support. In order to differentiate between 0

and 1, the polarity of the generated wavelet function is reversed for 0. For example, a binary sequence 1 0 1 will be transformed into the signal $x(t)$ given by:

$$x(t) = \begin{cases} +\psi(t), & t \in (0, T), \\ -\psi(t), & t \in (T, 2T), \\ +\psi(t), & t \in (2T, 3T), \end{cases} \quad (1)$$

where T is the duration of wavelet function. The period over which the wavelet function is applied depends upon the length of the host signal and the length of the watermark binary sequence. The watermark signal (data) is obtained by concatenating different wavelet functions obtained after coding the watermark binary sequence.

2.2. Watermark embedding algorithm

To embed a watermark, a wavelet function is generated with its parameters such as the type of wavelet, order of wavelet function, and number of iterations used for approximation. To embed a ‘1’ the wavelet function is generated using its parameters, while for ‘0’ its phase is reversed. The binary phase-coded wavelet functions corresponding to the watermark sequence of N bits are concatenated and embedded into the host audio signal (x_h). The watermarked audio signal (x_w) is given by:

$$x_w = x_h + \alpha w_c, \quad (2)$$

where w_c is the wavelet coded signal and α is the watermark scaling factor. The scheme of watermark embedding process is shown in Fig. 3.

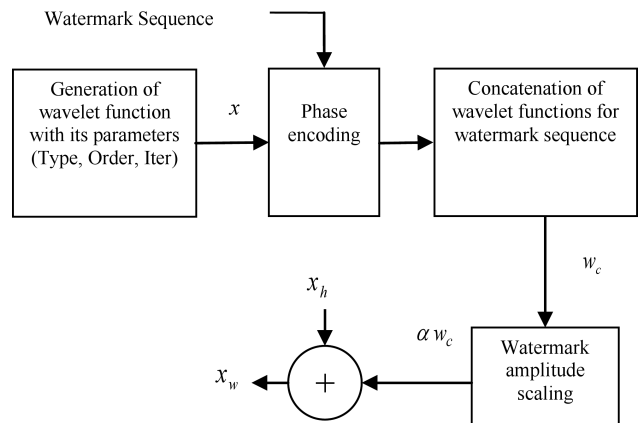


Fig. 3. Block diagram of watermark embedding process.

2.3. Multi-level watermarking

In order to increase the payload without a sacrifice in perceptual quality of speech, an additional watermark can be embedded. This can be achieved by adding a watermark to the previously watermarked signal but in a different frequency band. This can be achieved by using the same mother wavelet but with

a different number of iterations which occupy different frequency bands as shown in Fig. 2.

Thus, multi-level watermarking can be defined as a process of embedding multiple watermarks to the same host signal, where each watermark can be detected or extracted separately and securely with corresponding keys without the knowledge of other watermarks in the host signal (SHEPPARD *et al.*, 2001). Multi-level watermarking can be used to increase payload capacity and achieve different levels of robustness which offer different levels of security. Different applications use watermarks for different purposes. Each individual application has its own set of mutually conflicting requirements such as payload capacity, robustness, and perceptual quality. Multi-level watermarking can be performed using Eq. (2) successively. The i -th level watermarked signal (x_{wi}) can be obtained by the following equation:

$$x_{wi} = x_{w(i-1)} + \alpha_i w_{ci}, \quad 1 \leq i \leq L, \quad (3)$$

where $x_{w0} = x_h$ is the original host signal, α_i is the watermark scaling factor, and w_{ci} is the watermark to be embedded. The above equation can be repeated iteratively to finally get L level watermarked signal.

2.4. Watermark extraction algorithm

With the knowledge of the type of wavelet function and its parameters, the watermark can be extracted by measuring segment-by-segment correlation of the watermarked audio signal with the wavelet function. Watermark extraction process, shown in Fig. 4, uses a wavelet function x similar to the one used during the process of embedding. The watermarked audio signal (x_w) is segmented into N equal parts each of duration equal to the wavelet function, and each i -th segment of x_w is denoted by x_w^i . The watermark bits from x_w are extracted by calculating the cross-correlation coefficient (F_X) between the wavelet function x and x_w^i . The cross-correlation function (F_X) of x and x_w^i is determined by using the relationship given as:

$$F_X(i) = \sum_{k=0}^{L-1} x_w^i(k)x(k), \quad 1 \leq i \leq N, \quad (4)$$

where L is the number of samples.

The detected watermark bit w_d is obtained using a simple threshold logic by:

$$w_d = \begin{cases} 1 & \text{if } F_X(i) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The performance of the proposed embedding method is measured in terms of bit error rate (BER) of the detected watermark bits, as compared to the original watermark bits. The higher the BER, the poorer is the performance of the watermark algorithm.

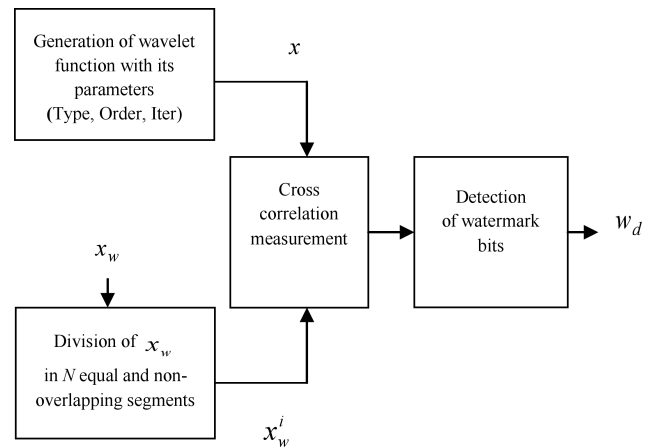


Fig. 4. Block diagram of watermark detection process.

3. Simulation results

Wavelet based multi-level watermarking up to $L = 3$ was implemented and compared with a similar chirp based watermarking. For these watermarking schemes, 14 audio files (6 speech files and 8 music files) selected from Speech Quality Assessment Material (SQAM) (SQAM, 2008) are used. These audio files are sampled at 44.1 kHz and have a resolution of 16 bits per sample. Daubechies (db) mother wavelet has been used for generating a watermark signal for the proposed scheme. The proposed Wavelet-Based Digital Audio Watermarking (WB-DAWM) scheme is compared with the Chirp-Based Digital Audio Watermarking (CB-DAWM) scheme.

In CB-DAWM Scheme (FAROOQ *et al.*, 2008b), a chirp signal is generated with its parameters such as its initial frequency (f_0), final frequency (f_1), and target time (t_1). The generated chirp signal is coded according to the watermark sequence to be embedded. To embed a '1' the chirp is generated using the above parameters, while for '0' its phase is reversed. Finally, these chirp signals corresponding to the watermark sequence of N bits are concatenated to form a signal of duration exactly equal to the audio signal. The watermarked audio signal x_w is generated in a similar manner as described in Eq. (2).

With the knowledge of the type of chirp (linear, quadratic, or logarithmic) and its parameters, the watermark can be extracted by measuring segment-by-segment correlation of the watermarked audio signal with the chirp signal. A chirp signal x similar to the one used during the process of embedding is generated. The watermarked audio signal x_w is segmented into N equal parts each of duration t_1 second and each i -th segment of x_w is denoted by x_w^i . The watermark bits from x_w are extracted by calculating the cross-correlation coefficient between the chirp signal x and x_w^i . The simulation parameters for the CB-DAWM scheme are given as follows:

- Type of the chirp used: logarithmic chirp;
- First-level watermark:
 $f_0 = 10$ Hz, $f_1 = 60$ Hz, $t_1 = 0.25$ sec;
- Second-level watermark:
 $f_0 = 10$ Hz, $f_1 = 30$ Hz, $t_1 = 1.0$ sec;
- Third-level watermark:
 $f_0 = 10$ Hz, $f_1 = 15$ Hz, $t_1 = 1.5$ sec.

The simulation parameters for the proposed WB-DAWM scheme are given as follows:

- Type of wavelet used: Daubechies (dbN) mother wavelet function;
- First-level watermark: db5 with iterations 10;
- Second-level watermark: db5 with iterations 13;
- Third-level watermark: db5 with iterations 14.

3.1. Performance metrics

The performance of an audio watermarking algorithm can be measured in terms of Signal-to-Watermark Ratio (SWR), Objective Difference Grade (ODG) using PEAQ, Subjective Listening Evaluation, and Bit Error Rate (BER). In our proposed digital audio watermarking scheme, two types of SWR are defined, namely, SWR_o and SWR_a as:

$$SWR_o = 10 \log_{10} \left\{ \frac{\sum_{i=0}^{N_s-1} x_h^2(i)}{\sum_{i=0}^{N_s-1} [x_h(i) - x_w(i)]^2} \right\}, \quad (6)$$

$$SWR_a = 10 \log_{10} \left\{ \frac{\sum_{i=0}^{N_s-1} x_h^2(i)}{\sum_{i=0}^{N_s-1} [x_h(i) - x_{aw}(i)]^2} \right\}, \quad (7)$$

where x_h , x_w and x_{aw} are host, watermarked, and attacked audio signals, respectively, and N_s is the number of samples in x_h , x_w and x_{aw} .

ODG Measurement using PEAQ Algorithm: the PEAQ algorithm is the ITU-R recommendation (ITU-R BS.1387-1) (PEAQ, 1998; KABAL, 2002) for perceptual evaluation of wide-band audio coders. Two versions of the PEAQ model are available: the basic and advanced ones. The PEAQ algorithm models the fundamental properties of the Human Auditory System (HAS) with physiological and psychoacoustic effects. This algorithm uses both original and watermarked audio signals to find differences between them. An ODG is evaluated using a total of eleven Model Output Variables (MOV) of the basic version of PEAQ model. The ODG values mimic the listening test ratings and have values from -4.0 for very

annoying quality to zero for imperceptible (inaudible) difference quality. The ODG values are interpreted as: 0 for EXCELLENT (imperceptible), -1 for GOOD (perceptible but not annoying), -2 for FAIR (slightly annoying), -3 for POOR (annoying), and -4 for BAD (very annoying). The ODG is calculated by the PEAQ algorithm specified in ITU-R BS.1387-1 and it corresponds to the Subjective Difference Grade (SDG) used in human based audio tests. They are computed with respect to the original reference audio signal. The resulting indexes are named ODG. The ODG for audio watermarking schemes is determined by subtracting the grade of the original host audio signal from the grade of the watermarked audio signal.

Subjective Listening Evaluation: Human listening tests are the only real subjective method for evaluating perceptual audio quality. Mean Opinion Score (MOS) grades are used in human listening tests for judging perceptual audio quality. The MOS is a five-point scale of quality which is associated with a set of standardized objective description; 5 for EXCELLENT (imperceptible), 4 for GOOD (perceptible but not annoying), 3 for FAIR (slightly annoying), 2 for POOR (annoying), and 1 for BAD (very annoying). MOS evaluations are well accepted and sometimes supplemented with measurement of intelligibility and acceptability. The subjective quality of audio watermarking schemes is measured by determining MOS through human listening tests.

Percentage Bit Error Rate is defined as:

$$BER = \frac{\text{No. of erroneously detected bits}}{\text{No. of embedded bits}} \times 100\%. \quad (8)$$

3.2. Performance evaluation and discussion

For Wavelet-Based Digital Audio Watermarking (WB-DAWM), Daubechies mother wavelet has been investigated with different orders using 10 iterations for approximating its value. Different results for a single-level WB-DAWM scheme using Daubechies (dbN) mother wavelet with its order (N) 1 to 7 are given in Table 1. Mother wavelet db5 using 10 iterations for approximating its value is chosen for coding watermark sequence because 75 bits are embedded per audio file and the embedded watermark is extracted without any error ($BER_{av} = 0$). This watermarking scheme is also imperceptible (inaudible) because the average Objective Difference Grade (ODG_{av}) is approximately zero (-0.067) and the average Signal-to-Watermark Ratio (SWR) is 30 dB. Here, ODG_{av} , SWR, and BER_{av} are the average values of Objective Difference Grade (ODG), Signal-to-Watermark Ratio (SWR), and Bit Error Rate (BER) in the extracted watermark.

Table 1. Results for single-level WB-DAWM scheme using Daubechies (dbN) mother wavelet.

Mother wavelet	Iterations	ODG _{av}	SWR [dB]	BER _{av}	Bits embedded
db1	10	-1.746	30	13.08	682
db2	10	-0.447	30	3.81	227
db3	10	-0.269	30	0.74	136
db4	10	-0.157	30	0.07	97
db5	10	-0.067	30	0	75
db6	10	-0.003	30	0	62
db7	10	0.039	30	0	52

Table 2. Results for WB-DAWM and CB-DAWM.

WB-DAWM					
Level	Bits embedded	Average ODG	MOS Score	SWR _a [dB]	Total bits embedded
I	75	-0.067	4.95	30.00	1050
II	9	-0.078	4.90	27.02	126
III	4	-0.080	4.88	25.33	56
Total No. of bits embedded (N_w)					1232
CB-DAWM					
I	63	-0.38	4.80	30.00	882
II	15	-0.46	4.75	27.04	210
III	10	-0.48	4.73	25.09	140
Total No. of bits embedded (N_c)					1232

Average performances (without attacks) for the wavelet-based and chirp-based audio watermarking schemes are given in Table 2. The proposed and chirp-based schemes are simulated as per parameters discussed in the previous section.

The ODG has been measured using the PEAQ algorithm which is an objective audio quality measure, and the MOS score (real subjective audio quality measure) is determined through human listening tests. It has been found from the results given in Table 2 that the proposed Single-Level and Multi-Level WB-DAWM schemes are imperceptible (inaudible) because average values of ODG and MOS for first, second and third levels of watermarked audio signals are in the imperceptible ranges (ODG is close to zero and MOS is close to 5). Our proposed Single-Level and Multi-Level WB-DAWM schemes are more imperceptible (inaudible) as compared to the corresponding CB-DAWM schemes because the measured ODG and MOS values are much closer to zero and 5, respectively, in comparison with the CB-DAWM schemes, as shown in columns 3 and 4 of Table 2.

Increased payload (P_I) in the single-level WB-DAWM scheme as compared to the single-level CB-DAWM scheme,

$$P_I = \left(\frac{1050 - 882}{882} \right) \times 100 = 19.05\%.$$

3.3. Imperceptibility test

One of the most important properties of a watermarking scheme is imperceptibility which is measured by subjective and objective methods. Imperceptibility of audio signals is also known as inaudibility. Objective measurement of inaudibility of watermarked audio signals is performed by determining SWR of watermarked audio signals. Inaudibility of watermarked audio signals can also be measured by determining the ODG values of watermarked audio signals using the PEAQ basic model. The ODG values (measured using the PEAQ basic model) mimic the listening test (subjective quality measurement) ratings and have values from -4.0 for very annoying quality to zero for imperceptible (inaudible) difference quality. Average SWRs achieved for the 14 audio signals for single-level and multi-level embedding schemes in the proposed WB-DAWM method are 30 dB and 25.33 dB, respectively. Since the human ear sensitivity below 100 Hz is more than 20 dB lower than the maximum sensitivity (which is around 3 kHz), the embedded mother wavelet is not perceived at these values of SWRs (30 dB and 25.33 dB). The average ODG values for single-level and multi-level (three-level) watermarked audio signals are -0.067 and -0.080, respectively. These ODG values show that both single-level and multi-level WB-DAWM schemes are imperceptible (inaudible) because

the measured ODG values are close to zero which is in the imperceptible range.

3.4. Robustness measurement

To evaluate robustness of the proposed schemes, various audio watermarking attacks such as filtering, sampling rate alteration, compression, noise addition, amplitude scaling, and cropping are applied. Robustness of the WB-DAWM scheme can be evaluated by measuring correlation between the original embedded and recovered watermark data. The proposed single-level and multi-level Wavelet-Based Digital Audio Watermarking (SL-WB-DAWM and ML-WB-DAWM) schemes are compared with the corresponding Chirp-Based Digital Audio Watermarking (SL-CB-DAWM and ML-CB-DAWM) schemes.

3.4.1. Filtering

Low-pass, high-pass, and band-pass filtering using Finite Impulse Response (FIR) digital filters of order 50 was applied on watermarked audio signal and the watermark was extracted from the filtered signal.

Low-Pass Filtering

It has been found that single-level and multi-level WB-DAWM schemes are robust against LPF attack for increasing the cutoff frequency (ω_{nL}) of low-pass filter (LPF) due to the watermark occupying the low frequency. In the proposed multi-level scheme, watermarks for all the three levels are extracted without any error because all the three watermarks occupy low frequency ranges. The value of ω_{nL} for LPF varies between 0.1 (4410 Hz) and 0.9 (39690 Hz). The embedded watermark is extracted from the low-pass filtered watermarked audio signal without any error because the value of ω_{nL} is much higher than the frequency of

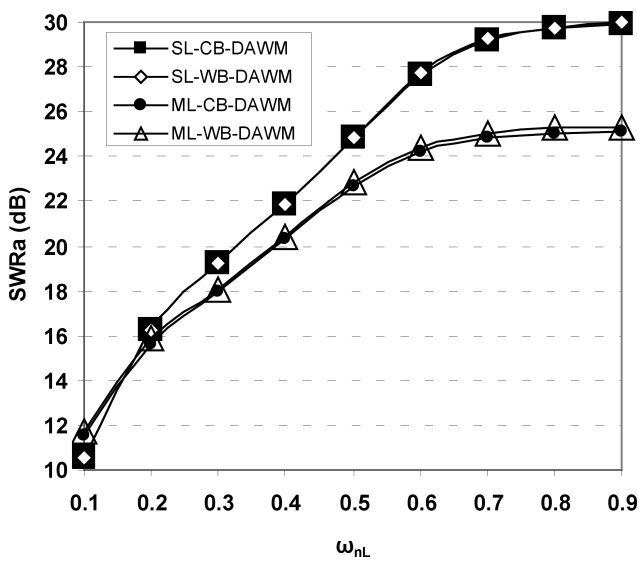


Fig. 5. Variation of SWR_a with increasing ω_{nL} for single- and multi-level schemes.

embedded watermark. This implies that even if 90% of the bandwidth (BW) is lost in filtering, the watermark can still be recovered although the signal becomes useless. The SWR_a for single-level and multi-level schemes after LPF (results shown in Fig. 5) increases with increasing ω_{nL} because the BW of the filtered audio signal increases with increasing the value of ω_{nL} .

High-Pass Filtering

It is evident from the results shown in Fig. 6 that the single-level WB-DAWM scheme is robust under high-pass filtering for a normalized cutoff frequency of the High-Pass Filter (HPF) ω_{nH} up to 0.06 (2646 Hz) because the embedded watermark is extracted without any error. The reason for this is that even though the wavelet-based watermark is of low frequency, the HPF, due to smooth transition, does not remove the watermark but attenuates it. However, for higher cutoff frequencies, errors start to occur, due to which the embedded watermark gets severely attenuated and is not recoverable. As the value of ω_{nH} increases, attenuation in the watermarked frequency band also increases, causing more bits in error and lowering the SWR_a of the watermarked audio signal. Our proposed WB-DAWM (single-level and multi-level) schemes (referred to Fig. 6 and Table 3) are more robust under HPF as compared to the CB-DAWM (single-level and multi-level) methods. In the single-level WB-DAWM scheme (results shown in Fig. 6), the embedded watermark is extracted without any error for ω_{nH} equal to 0.06 (2646 Hz). At the same time, in the single-level CB-DAWM scheme, the embedded watermark is extracted without any error for ω_{nH} equal to 0.04 (1764 Hz). The multi-level WB-DAWM scheme (as results given in Table 3) is also more robust against HPF for all the three levels of watermarks as compared to the multi-level CB-DAWM scheme.

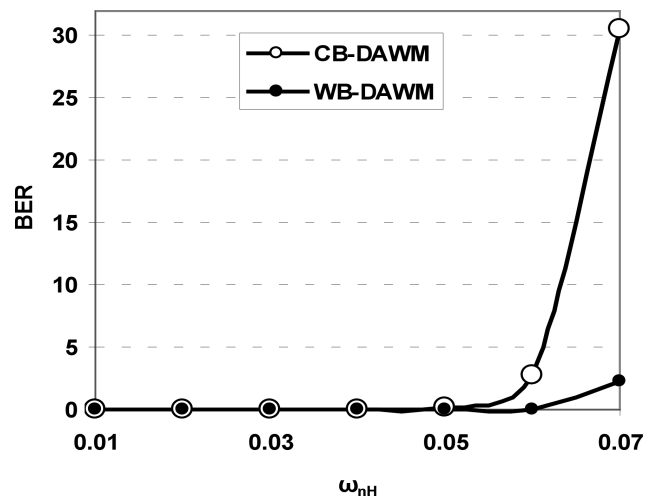


Fig. 6. Watermark extraction performances for single-level schemes under HPF.

Table 3. Results of multi-level watermarking schemes under HPF.

ω_{nH}	CB-DAWM				WB-DAWM			
	BER I	BER II	BER III	SWR _a [dB]	BER I	BER II	BER III	SWR _a [dB]
0.01	0	0	0	15.22	0	0	0	15.24
0.02	0	0	0	9.78	0	0	0	9.78
0.03	0.72	0	0	6.74	0	0	0	6.74
0.04	1.07	0	0	4.83	0	0	0	4.83
0.05	1.79	0	0	3.59	0	0	0	3.59
0.06	4.53	0	0	2.78	0	0	0	2.78
0.07	31.55	14.76	15	2.27	2.27	0	0	2.26

Band-Pass Filtering

The band-pass filter (BPF) used here is similar to the HPF that was used previously except for the upper cutoff frequency (ω_{n2B}) being kept at 0.9 (39690 Hz). It has been found from simulation results that the embedded watermark can be extracted without any error up to the lower cutoff frequency of BPF (ω_{n1B}) equal to 0.06 (2646 Hz) for a filtered single-level watermarked audio signal. The proposed single-level and multi-level WB-DAWM schemes are also more resilient against band-pass filtering attacks as compared to the corresponding CB-DAWM schemes.

3.4.2. Sampling rate alteration

Various sampling rate alteration processes such as upsampling, downsampling, and resampling are applied to watermarked audio signals and resampled signals correlated with the appropriate wavelet function. These sampling rate alteration processes are applied on both single-level and multi-level watermarked audio signals obtained by using the proposed WB-DAWM and CB-DAWM schemes.

Interpolation

It is noticed from the simulation results obtained after upsampling by interpolation that the proposed single-level and multi-level schemes are found robust. By upsampling, no spectral distortion is introduced. Therefore, all the watermarks can be extracted without any error and SWR_a remains unaltered for interpolated watermarked audio signals as given in Table 4.

Decimation

It is revealed from the simulation results obtained after downsampling by decimation that the single-level and multi-level proposed audio watermarking schemes are found resilient. The value of SWR_a decreases with increasing of the decimation factor (N_d) due to the fact that the low-pass filtering operation is performed before downsampling (as referred to Table 5). Even though the spectrum is spread during the decimation process, the watermark is preserved because it is of a very low frequency range. In downsampling by decimation, the SWR_a decreases on an increasing N_d due to a higher rate of removal of frequency components in this process.

Table 4. Variation of SWR_a with increasing N_u under upsampling.

Interpolation Factor (N_u)	Single-Level		Multi-Level	
	CB-DAWM	WB-DAWM	CB-DAWM	WB-DAWM
2	30	30	25.09	25.18
3	30	30	25.09	25.18
4	30	30	25.09	25.18

Table 5. Variation of SWR_a with increasing N_d under downsampling.

Decimation Factor (N_d)	Single-Level		Multi-Level	
	CB-DAWM	WB-DAWM	CB-DAWM	WB-DAWM
2	24.95	24.94	22.24	22.29
3	20.04	20.03	18.27	18.30
4	18.04	18.03	16.59	16.76

Resampling

Resampling a signal by some arbitrary rational factor ($N_r = P/Q$) is equivalent to upsampling (interpolation) by an integer factor (P) followed by downsampling (decimation) by another integer factor (Q). Resampling by varying arbitrary factor (N_r) from 0.1 to 1.2 in a step size of 0.1 is applied to attack the single-level and multi-level watermarked audio signals and the results are shown in Fig. 7. It has been found that the single-level and multi-level proposed audio watermarking schemes are robust to resampling by some arbitrary factor. In both cases, watermarks can be extracted without any errors as they are of a very low frequency. If $P < Q$ (i.e., $N_r < 1$), on increasing the value of N_r , the SWR_a increases up to the SWR of the watermarked audio signal. This case has the same performance as that of merely downsampling by decimation. If $P \geq Q$ (i.e., $N_r \geq 1$), on increasing the value of N_r , the SWR_a remains constant. This case has the same performance as that of merely upsampling by interpolation.

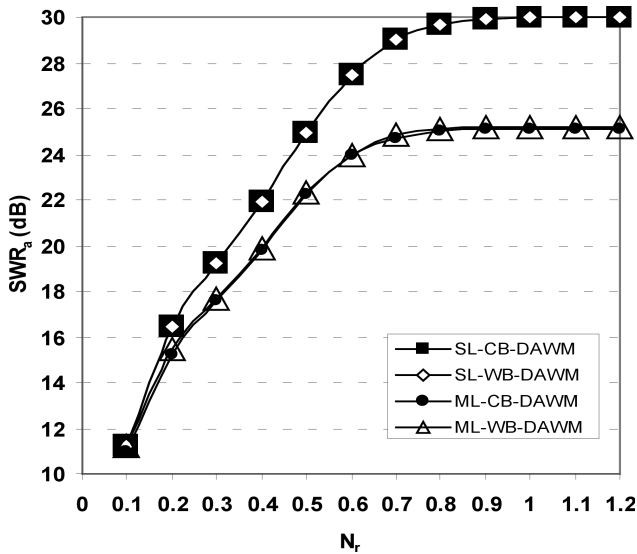


Fig. 7. Variation of SWR_a with increasing N_r under resampling.

3.4.3. MP3 compression

Simulation experiments are carried out for a wide range of MPEG-1 Layer-3 (MP3) compression attacks with Constant Bit Rate (CBR) ranging from 56 kbps to 320 kbps. It has been found that the single-level and multi-level proposed schemes are robust to MP3 compression because the watermarks are extracted without any error. As the bit rate of the MP3 compression algorithm is reduced, a higher compression ratio is obtained. By reducing the bit rate of MP3 compression, more redundant information from the audio signal is eliminated. The SWR_a decreases with reduction in the bit rate of MP3 for the single-level and multi-level schemes as shown in Fig. 8. This is because of

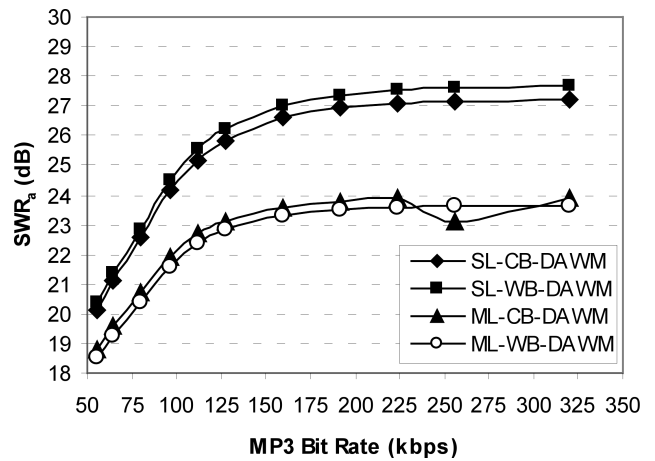


Fig. 8. Variation of SWR_a with increasing bit rate of MP3 compression.

the fact that MP3 removes less significant information from the audio to achieve higher compression (i.e. lower bit rate).

3.4.4. Noise addition

Additive white Gaussian noise (AWGN) of varying noise power is injected into the watermarked audio signals to give Signal-to-Noise Ratio (SNR) in the range -5 dB to 50 dB. In the single-level audio watermarking scheme (referred to in Table 6), as the SNR increases (i.e. the noise decreases), the SWR_a reaches its maximum value (30 dB); in this case, the noise power is negligible. Therefore, the BER in the extracted watermark is also zero for a higher SNR. At the SNR of 10 dB, the SWR_a is also approximately equal to 10 dB. As the SNR decreases (i.e. the noise increases) the noise starts to dominate and the SWR_a becomes equal to or less than SNR. In the multi-level audio watermarking scheme (referred to in Table 7), for the first, second, and third levels of watermarks, the BER in the extracted watermarks is equal to zero at 5 dB, -5 dB, and -5 dB, respectively. The proposed WB-DAWM schemes are more robust against AWGN attacks as compared to the corresponding CB-DAWM schemes, the results are given in Tables 6 and 7.

Table 6. Results for the single-level watermarking schemes under AWGN.

SNR [dB]	CB-DAWM		WB-DAWM	
	BER	SWR_a [dB]	BER	SWR_a [dB]
-5	4.42	-5.01	5.14	-5
0	0.68	-0.01	0.38	-0.01
5	0.23	4.98	0	4.98
10	0	9.95	0	9.95
15	0	14.86	0	14.86
20	0	19.59	0	19.58

Table 7. Results for multi-level watermarking schemes under AWGN.

SNR [dB]	CB-DAWM				WB-DAWM			
	BER I	BER II	BER III	SWR _a [dB]	BER I	BER II	BER III	SWR _a [dB]
-5	7.86	0.48	0	-5.02	4.30	0	0	-5.02
0	2.26	0	0	-0.03	0.57	0	0	-0.03
5	1.67	0	0	4.95	0	0	0	4.95
10	0.60	0	0	9.85	0	0	0	9.86
15	0.12	0	0	14.59	0	0	0	14.58
20	0	0	0	18.82	0	0	0	18.82

3.4.5. Amplitude scaling

Amplitude scaling of single-level and multi-level watermarked audio signals is performed for different scaling factors: from 2 to 10. It is found that the single-level and multi-level watermarking schemes are resilient to amplitude scaling. Amplitude scaling of the watermarked audio signal does not result in spectral modification. Therefore, there is zero BER in the extracted watermark, and the SWR_a remains unaltered for the single-level and multi-level schemes.

3.4.6. Cropping

In cropping attack, some of the samples are removed from the end of the watermarked audio signal; BER increases on increasing the percentage crop (ratio of number of samples cropped to total number of samples in watermarked audio signal) of the audio signal but SWR_a decreases on increasing the percentage crop. If the signal is cropped from the beginning or middle of the audio signal, then the watermark is not detectable from the point beyond which the signal has been cropped, this is due to the offset problem of samples. It has been found from the results obtained after cropping watermarked audio signal that proposed single-level and multi-level WB-DAWM schemes are not resistant against cropping attack

4. Conclusions

A wavelet-based blind digital audio watermarking scheme proposed in this paper is found useful for single-level and multi-level watermark embedding schemes. The proposed scheme has been simulated and tested for various audio signal processing attacks such as filtering, sampling rate alteration, MP3 compression, addition of AWGN, amplitude scaling, cropping, and has been shown to be robust to most of the audio attacks. The proposed schemes (single-level and multi-level) are found resilient for various attacks such as low-pass filtering, upsampling by interpolation, down-sampling by decimation, resampling by some arbitrary rational factor, amplitude scaling, and MP3 compression. These schemes show limited robustness against

high-pass filtering, band-pass filtering, and AWGN attacks. The proposed WB-DAWM schemes are more robust against high-pass filtering, band-pass filtering, and AWGN attacks as compared to corresponding CB-DAWM schemes. These schemes are not robust under cropping attacks. The proposed scheme of embedding mother wavelet functions as watermarks which overlap in time but occupy different frequency bands is an imperceptible and robust watermarking scheme. In the proposed single-level scheme, payload capacity is increased by 19.05% as compared to the single-level chirp-based digital audio watermarking scheme.

References

1. AL-HAJ A., MOHAMMAD A. (2010), *Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition*, European Journal of Scientific Research, **39**, 1, 6–21.
2. ARNOLD M., WOLTHUSEN S., SCHMUCKER M. (2003), *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, Springer-Verlag.
3. BARNI M., BARTOLINI F. (2004), *Watermarking Systems Engineering Enabling Digital Assets Security and Other Applications*, Marcel Dekker Press.
4. BASSIA P., PITAS I. (2001), *Robust Audio Watermarking in the Time-Domain*, IEEE Transactions on Multimedia, **3**, 2, 232–241.
5. BENDER W., GRUHL D., MOROMOTO N., LU A. (1996), *Techniques for Data Hiding*, IBM Systems Journal, **35**, 3–4, 313–336.
6. BLACKLEDGE J., FAROOQ O. (2008), *Audio Data Verification and Authentication Using Frequency Modulation Based Watermarking*, ISAST Transactions on Electronics and Signal Processing, **3**, 2, 51–63.
7. COX I.J., KILIAN J., LEIGHTON F.T., SHAMMOON T. (1997), *Secure Spread Spectrum Watermarking for Multimedia*, IEEE Transactions on Image Processing, **6**, 12, 1673–1687.
8. COX I. J., MILLER M., BLOOM J. (2002), *Digital Watermarking*, Academic Press, USA.
9. DAUBECHIES I. (1990), *The Wavelet Transform, Time-Frequency Localization and Signal Analysis*, IEEE Transactions on Information Theory, **36**, 5, 961–1005.

10. FAROOQ O., DATTA S., BLACKLEDGE J. (2008a), *Robust Watermarking of Audio with Blind Self-Authentication*, 7th WSEAS International Conference on Electronics, Hardware, Wireless and Optical communications, Cambridge, UK, February 20–22, 2008.
11. FAROOQ O., DATTA S., BLACKLEDGE J. (2008b), *Blind Temper Detection in Audio Using Chirp Based Robust Watermarking*, WSEAS Transactions on Signal Processing, **4**, 4, 190–200.
12. SQAM (2008), <http://sound.media.mit.edu/mpeg4/audio/sqam/Dec.2008>.
13. PEAQ (1998), *International Telecommunication Union (ITU): Method for Objective Measurements of Perceived Audio Quality (PEAQ)*, ITU-R BS.1387, 1998.
14. KABAL P. (2002), *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality (PEAQ)*, Technical Report, McGill University, Version 2, 2002.
15. KIROVSKI D., MALVAR H.S. (2003), *Spread Spectrum Watermarking of Audio Signals*, IEEE Transactions on Signal Processing, **51**, 2, 1020–1033.
16. KO B. S., NISHIMURA N., SUZUKI Y. (2005), *Time-Spread Echo Method for Digital Audio Watermarking*, IEEE Transactions on Multimedia, **7**, 2, 212–221.
17. LANGELAAR G., SETYAWAN I., LAGENDIJK R. (2000), *Watermarking Digital Images and Video Data: A State-of Art Overview*, IEEE Signal Processing Magazine, **7**, 5, 20–46.
18. LIE W.N., CHANG L.C. (2006), *Robust and High Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modulation*, IEEE Transactions on Multimedia, **8**, 1, 46–59.
19. LILI L., JIANLING H., XIANGZHONG F. (2007), *Spread Spectrum Audio Watermark Robust Against Pitch-Scale Modification*, IEEE International Conference on Multimedia, 1770–1773.
20. NEUBAUER C., HERRE J. (1998), *Digital Watermarking and Its Influence on Audio Quality*, Proc. of AES Convention 1998, San Francisco, CA, 225–233.
21. OH H.O., SEOK J.W., HONG J.W., YOUN D.H. (2001), *New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking*, ICASSP-2001, 1341–1344.
22. QUAN X., ZHANG H. (2004), *Audio Watermarking Based on a Psychoacoustic Model and Adaptive Wavelet Packets*, 7th International Conference on Signal Processing, **3**, 1, 2518–2521.
23. SHEPPARD N. P., SAFAVI-NAINI, OGUNBONA P. (2001), *On Multiple Watermarking*, ACM Multimedia Conference-2001, Ottawa, Canada, 3–6.
24. VIERU R., TAHBOUB R., CONSTANTINESCU C., LAZARESCU V. (2005), *New Results Using Audio Watermarking Based on the Wavelet Transform*, International Symposium on Signals, Circuits, and Systems, **2**, 1, 441–444.
25. WANG X. Y., ZHAO H. (2006), *A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT*, IEEE Transactions on Signal Processing, **54**, 12, 4835–4840.
26. XIE L., ZHANG J., HE H. (2006), *Robust Audio Watermarking Scheme Based on Non-uniform Discrete Fourier Transform*, IEEE International Conference on Engineering of Intelligent Systems, 1–5.
27. XIONG Y., MING Z.X. (2006), *Covert Communication Audio Watermarking Algorithm Based on LSB*, International Conference on Communication Technology, ICCT-2006, 1–4.
28. XU C., WU J., SUN Q., XIN K. (1999), *Applications of Watermarking Technology in Audio Signals*, Journal of Audio Engineering Society (AES), **47**, 10.