

Analysis of impact of audio modifications on the robustness of watermark for non-blind architecture

Janusz Cichowski · Andrzej Czyżewski · Bożena Kostek

Published online: 7 August 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The aim of this paper is to assess the robustness of the non-blind audio content watermarking scheme proposed by the authors. The authors present the architecture of the designed system along with the employed workflows for embedding and extracting the watermark followed by the implementation phase description and the analysis of the experimental results. Some possible attack simulations on the embedded watermarks are reviewed, and the robustness of the proposed approach is evaluated in the context of the influence of lossy compression on the watermark degradation. Subjective and objective analyses are performed for the algorithm proposed by the authors and compared with the Audio Watermarking Tools (AWT) encoder. Finally, the advantages and drawbacks of the proposed approach are debated followed by the conclusion section outlining possible improvements to the proposed method.

Keywords Non-blind audio watermarking · Discrete wavelet transform · Lossy compression · Quality

1 Introduction

The goal of the described research was to develop the watermarking algorithm for intellectual property protection of audio content, and to evaluate its efficiency in terms of the audio signal quality. The intended field of application of the watermarking algorithm is the *youarchive* system for online audio restoration developed at Gdansk University of Technology [32]. Since the archival audio material to be restored (by automatic removing parasite clicks, crackles, noise, hiss, etc.) might be of different nature and origin, the analysis of

J. Cichowski (✉) · A. Czyżewski
Multimedia Systems Department, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk,
Poland
e-mail: jay@sound.eti.pg.gda.pl

A. Czyżewski
e-mail: andcz@sound.eti.pg.gda.pl

B. Kostek
Laboratory of Audio Acoustics, Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: bokostek@audioacoustics.org

subjective perceptiveness of watermarks took various music genres into consideration. Moreover, in this field of application the immunity of the watermark to signal processing algorithms applied to restore audio and the need to preserve subjective quality of sound seem of primary importance. Consequently, the type of applied watermarking was considered carefully at the stage of choosing an adequate watermarking strategy. Therefore, the non-blind schemes attracted our particular attention, revealing a supremacy over blind watermarking schemes in the above context.

The proposed method extends the conventional function of watermarking as its watermarks may well contain music feature vectors [9], thus forming a metadata transmission channel via which music content necessary in music retrieval processes can be extracted. Before the introduced concepts and research results are presented, an introduction on the watermarking technology developments and their fields of applications will be provided.

Even though recent developments in digital technology—such as high-speed broadband networks, decreasing cost of disk storage or growing capacity of portable memory—changed the way of production, distribution and protection of multimedia content, current expectations for multimedia applications still far exceed the achievement of today's technology. Moreover, illegal copying of digital data may easily lead to media piracy. This implies the need for a much stronger protection of copyrighted data such as audio, music recordings, still photos, movies, software and e-books. The problem of preserving the copyrights has not existed on such a large scale until the development of the Internet, the FTP protocol, and the peer-to-peer software. First methods of copyright protection were designed for the analog domain and are of no use for digital data protection. Hence, copyright holders are seeking novel approaches to protect their property in a way that would be designed for specific content.

The subject of data protection is very important in the Digital Era [16], and it is present in the literature also in the context of watermarking security [21, 22]. The most popular approach for both digital audio and digital video data protection involves watermarking [5, 6, 27, 29, 31, 33]. Various audio watermarking methods have been proposed in the past. The basic idea of digital watermarking lies in inserting imperceptible changes that form the binary data of a watermark into an audio/video signal. In general, watermarking divides into two different ways described in the literature [13] as *active* (achieved with the use of a dedicated multimedia player, which plays music/video only if the algorithm finds a watermark) or *passive* (applied in order to detect piracy via special data embedded as fingerprint used to identify the suspected user's ID, email or IP address). The methods of watermarks implanting are typically classified as blind and non-blind. The blind watermarking allows watermarks to be extracted without the original host signals. However, the drawback of this method is the quality degradation and the increased watermark audibility. Conversely, in the non-blind audio watermarking the original signals are integral to the process of watermarks extraction. The advantage of this method is that watermarks are embedded more imperceptibly and they do not lower the quality of the audio recordings.

Many digital watermarking methods were developed and tested over the recent years. Only a few of them were however designed to secure audio data. It is to notice that despite that recently watermarking technology has been developed intensely in the field of video [31], there is still a need to make progress in practical applications of these techniques to audio.

The spread spectrum watermarking proposed by Bloom and Cox [7] is the most common audio watermarking approach. One bit of information is encoded with a pseudorandom sequence which is statistically neutral. The coded information is embedded in the spectral coefficients of a DCT transform. To recover hidden data, the pseudorandom seed for

subsequent bits of information is required. Dutta, Gupta and Pathak proposed a watermarking technique based on DCT (Discrete Cosine Transform) [11]. Foo and Dong [12] presented a different approach also based on DCT. In this approach, the host audio signals are segmented into frames. Two consecutive frames are assessed, if they are suitable to represent a watermark bit. The proposed scheme adopts a compression-expansion technique to generate a special waveform over two consecutive frames. Foo and Dong also applied the psychoacoustic model to calculate a local auditory mask to ensure that the distortion caused by watermarking is not audible. Moreover, it is shown that the design of the watermarking schemes for mono and stereo audio signals is different. To detect the distortion and extract embedded watermark bits, the correlation-based detection method is used [12].

The most recent approach to audio watermarking is based on the DWT (Discrete Wavelet Transform) domain. An example of a watermarking system based on DWT can be found in the literature [18]. It should also be emphasized that watermarking algorithm is always evaluated taking into account possible attacks and its resistance to such. The embedded watermark should be robust and easily extractable from audio signals even after incidental or intentional attacks, such as noise addition, resampling, filtering or MP3 compression.

We should note that developed audio watermarking techniques may be used for various goals. One of them is echo hiding designed by Ciarkowski and Czyżewski [8], and later by Szwoch et al. [28]. In this approach, the echo-cancelling algorithm embeds a hidden watermark signal into the arriving signal. The detection of a hidden signature allows the algorithm to stop or restart the process of the adaptive echo cancellation. An interesting usage of watermarking is a system designed for reducing pre-echo artifacts by embedding temporal envelope correction as watermarking data, which eliminates the need for any auxiliary communication channel. The restored signal is reconstructed at the receiver from two components: the decoded signal generated by the core decoder, and the temporal envelope of the original signal, which is represented by a few parameters that are transmitted as embedded data in the audio signal [25]. In the approach proposed by Kondo [16], data are embedded using the polarity of the reverberation (echoes) that is added to high frequencies. Each type of a watermarking scheme balances between the watermark capacity, robustness and transparency (imperceptibility). It should be noted that non-blind approaches improve all mentioned above.

Computer industry giants like Apple or Microsoft developed their own mechanisms for digital rights management (DRM) for multimedia data. In Apple's technology called Fairplay [30], watermarks are implanted in protected files as an indicators of the legitimate content. Apple's iTunes software looks for the inserted watermarks to verify the authenticity of the digital content. The mechanism, however, works only with Apple devices and software. Microsoft developed Windows Media DRM services [19]. The Janus DRM [15] algorithm is oriented on portable devices and its modification called Cardea DRM has been designed for network/streaming devices. Both algorithms protect audio visual content using Windows Media Player and the algorithms implemented in specific devices.

The stand alone implementation of the watermarking mechanism is provided as the Audio Watermarking Toolkit (AWT) [4], and it performs both blind and non-blind watermarking. As implementing such commercial DRM systems as Fairplay or Windows Media DRM is practically impossible, the algorithm proposed by the authors was compared with a patented approach delivered as AWT [3]. The demo version of this software allows to perform comparison with the method implemented by the authors.

The paper is organized as follows. In Section 2, an audio watermarking scheme operating in the wavelet transform domain is described and presented through processing workflows provided with a brief description of the developed watermark embedding. In Section 3, the principle of the non-blind extraction employing DWT and the motivation for a downgrading

procedure application are presented. In Section 4, the simulations of possible attacks, the measurement methodology, and the obtained results are included. In Section 5, a comparison of various formats of audio lossy compression is made, and the discussion on the obtained results is presented. In Section 6, the results of subjective tests focused on perceptual audio quality degradation caused by watermarking procedure are shown. In addition, listening tests were performed to check whether the embedded watermarks are perceptually transparent. Section 7 contains results of objective analyses, i.e. spectral analyses as well as PEAQ (Perceptual Evaluation of Audio Quality) analyses performed for the proposed and the AWT algorithms. The summarizing Section contains observations and conclusions focused on how to reduce possible drawbacks of the method in implemented future system implementations. This article is an extended and revised version of the paper presented at the MCSS'12 conference [10].

2 Watermarking embedding process

The method proposed by the authors consists in generating a 10-character text information, and then translating it into a binary sequence. This provides the input for a watermark embedding module. Watermarking takes place in the DWT domain.

A general scheme of the watermark embedding procedure is shown in Fig. 1. As presented in Fig. 1, the original file is divided into non-overlapping frames containing 1024 samples each. Frames can also have a width of 256, 512 or 2048 samples. The designed application accepts files with the sample rate of 44.1 kHz. A wider frame would lead to fewer artifacts generation during the watermark embedding, but at a cost of a smaller watermark bitrate. One bit of a watermark is implanted within audio signal samples of each frame. The watermark is embedded in the second level low frequency DWT transformation of the frame. The values of samples in frames are modified according to the following formula (Eq. (1)):

$$\begin{aligned} s_{LP_{wat}}^2[n] &= 0.9 \cdot s_{LP_{org}}^2[n] + \alpha \cdot \left| s_{LP_{org}}^2[n] \right| \Rightarrow bit = 1 \\ s_{LP_{wat}}^2[n] &= 0.9 \cdot s_{LP_{org}}^2[n] - \alpha \cdot \left| s_{LP_{org}}^2[n] \right| \Rightarrow bit = 0 \end{aligned} \quad (1)$$

where:

$s_{LP_{wat}}^2[n]$	output sample value of the DWT second level low-pass component
$s_{LP_{org}}^2[n]$	original sample value from the current DWT second level low-pass component
α	watermarking strength parameter
bit	current bit value of the embedded information

The α parameter represents the watermarking strength and is inversely proportional to the fidelity of the watermarked file (understood in terms of identity of the content of the watermarked file in comparison to the source file). The α parameter values range from 0 to 10. The current audio sample after the second level DWT decomposition $s_{LP_{org}}^2[n]$ is multiplied by the factor of 0.9 to avoid the audio signal being overridden by watermarking. The audio signal is modified by multiplying the strength parameter α by $s_{LP_{org}}^2[n]$ (current decomposed audio samples) and the result should fall into the range of [0.0–0.1]. If the bit being embed equals to 1, the result of multiplication is added to the current rescaled sample value, contrarily if the bit is 0, the result of the multiplication is subtracted from the current rescaled sample value.

The presented approach uses two embedding modes for stereo signals and one for mono signals. The first mode, called Mono/Stereo 1, embeds the same watermark signal in both

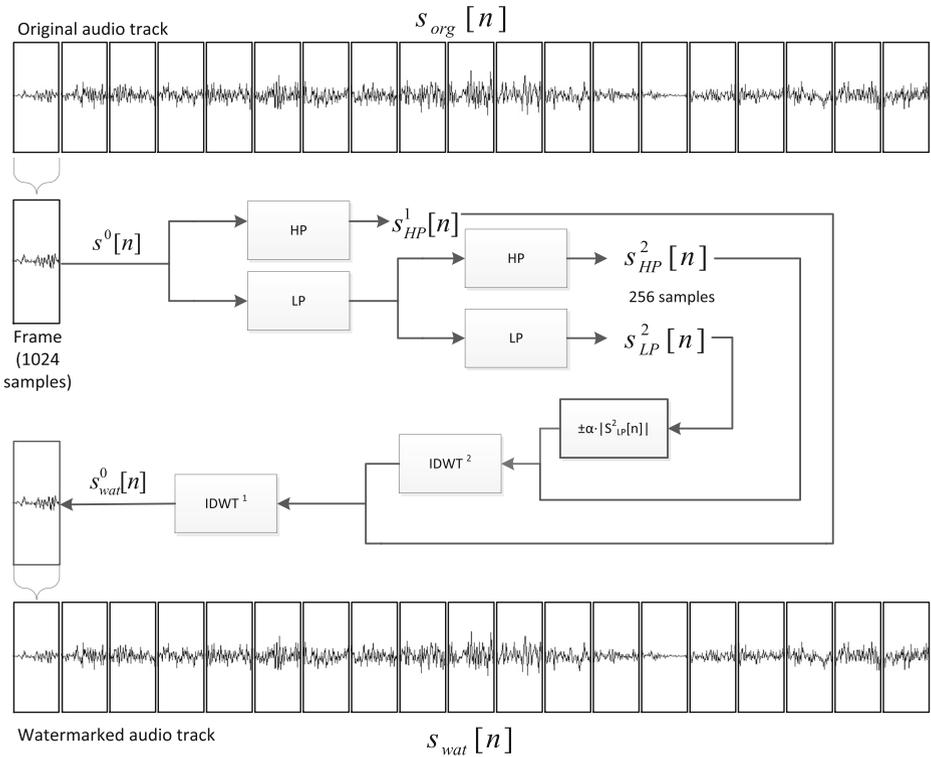


Fig. 1 Watermark embedding process

channels of a stereo signal, which leads to redundancy. The second mode, called Stereo 2, uses two separate media channels for watermarking, which provides a larger watermark bitrate. In this study, the Stereo 1 mode is used. The interdependence of a signal width and a watermark bitrate is shown in Table 1. Once the embedding process has been completed, a frame is transformed back into the time domain and is added into the output stream.

3 Non-blind extraction scheme

The least errorless watermark extraction is as important as watermark embedding. The non-blind extraction mechanism [7] requires the availability of both the original and the watermarked signals to extract the information hidden in the watermark. The general idea of the extraction procedure is presented in Fig. 2.

Table 1 The interdependence of watermarking mode, frame width and watermark bitrate

Frame width [samples]	256	512	1024	2048
- Mono				
- Stereo 1 [bit/s]	172	86	42	22
- Stereo 2 [bit/s]	344	172	86	42

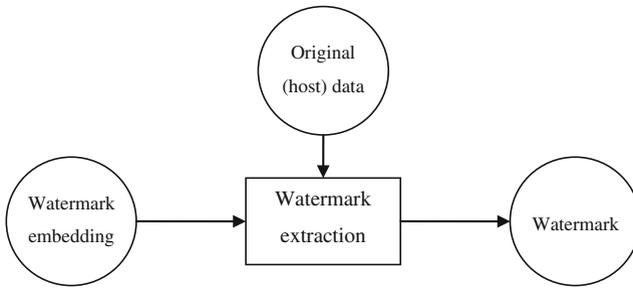


Fig. 2 Non-blind watermark extraction scheme

The disadvantage of providing the original signal for watermark extraction is recompensed by a high imperceptibility and capacity of the watermark, both being desirable parameters in audio applications. Moreover, the watermarking strength can be maintained at adequate level, so it does not introduce any audio quality degradation.

Extracting the watermark is similar to embedding it. The extraction scheme is shown in Fig. 3. Two audio tracks are required to extract the watermark: the original uncompressed (S_{org}) and the watermarked (S_{wat}) one. If the watermarked file is converted into a different format than the original signal, some preprocessing must be performed on, i.e. the downgrading procedure of compressing S_{org} to the same format as S_{wat} must be applied first.

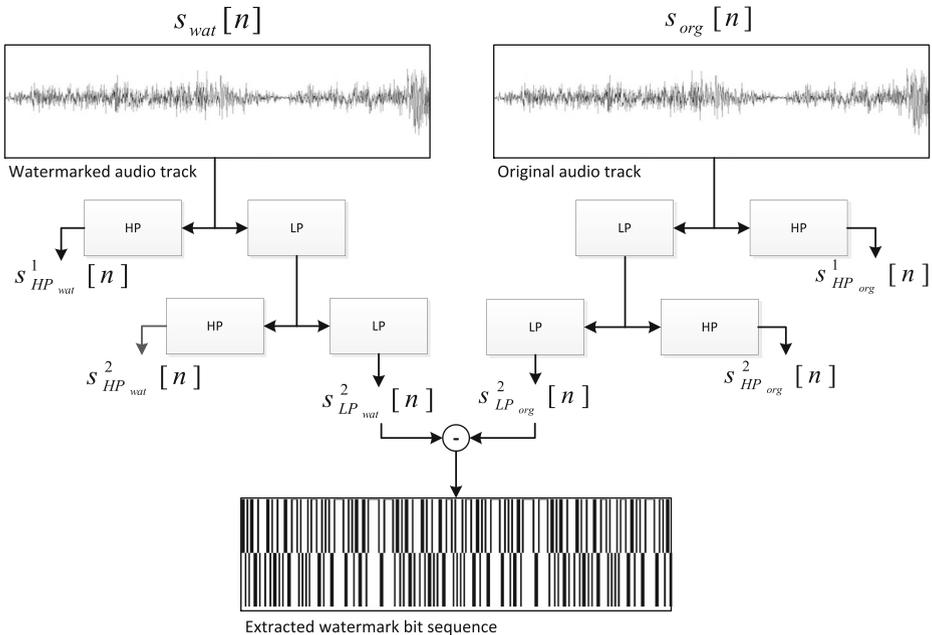


Fig. 3 Watermark extraction process

The difference between the low frequency parts of the second level DWT of the two files represents the watermark signal and is calculated according to formula presented by Eq. 2.

$$bit = \begin{cases} 0, & \Rightarrow t \leq \sum_{n=0}^{255} s_{LP_{wat}}^2[n] - 0.9 \cdot s_{LP_{org}}^2[n] \\ 1, & \Rightarrow t \geq \sum_{n=0}^{255} s_{LP_{wat}}^2[n] - 0.9 \cdot s_{LP_{org}}^2[n] \end{cases} \quad (2)$$

where:

$s_{LP_{wat}}^2[n]$ output sample value of the DWT second level low-pass component
 $s_{LP_{org}}^2[n]$ original sample value from the current DWT second level low-pass component
 t thresholding level
 bit current bit value of the embedded information

The obtained difference signal is searched for the expected start sequence of the watermark and afterwards divided into non-overlapping frames containing 256 (quarter of 1024) samples each. One watermark bit is extracted from each frame. The signal differences of the whole frame are summed. Finally, the binary sequence is translated into text and can be presented in a human-readable form.

The results of thresholding the signal containing differences between the original and watermarked audio stream are presented in the bottom part of Fig. 3 as a pseudo binary sequence. The procedure focuses on the extraction of the watermark insertion signature, which is embedded before each binary sequence of the watermark. Locating the insertion sequence allows to identify the first bit of the watermark and starts proper watermark extraction. The insertion signature is an 8-bit binary sequence, selected in a way which minimizes the probability of its accidental repetition inside the watermark bitstream. If the audio data are unprocessed (not modified) the analysis and watermark extraction are errorless. Yet, the variety of possible audio signal modifications that may happen in real life force the assumption of the error-prone watermark extraction. Errors at the stage of signature insertion cause the problem with finding the first bit of the watermark sequence and in consequence induce the increased possibility of a false positive detection of the beginning of the watermark. To solve this, the authors used the autocorrelation function given by Eq. (3):

$$ACF[m] = \frac{1}{N-m} \sum_{n=1}^{N-m} x^*[n] \cdot x[n+m] \quad (3)$$

where:

$ACF[m]$ auto correlation function value for m shifted samples
 $x[m]$ current sample value in the thresholded signal
 $x^*[m]$ conjugate value of current sample
 N number of signal samples

The watermark binary sequence can be repeatedly embedded in an audio signal. The repetitions of the watermark are presented as local maxima of the ACF (autocorrelation function). The ACF values are within the range of [0.0–1.0]. A value close to 0.0 means no signal correlation for a specific shift, while a value close to 1.0 depicts a strong signal correlation. An example of the autocorrelation analysis for a non attacked watermark sequence is presented in Fig. 4.

The peaks visible in Fig. 4b correspond to the beginning of the watermark binary sequences. There are more than one peak because the watermark was embedded repeatedly.

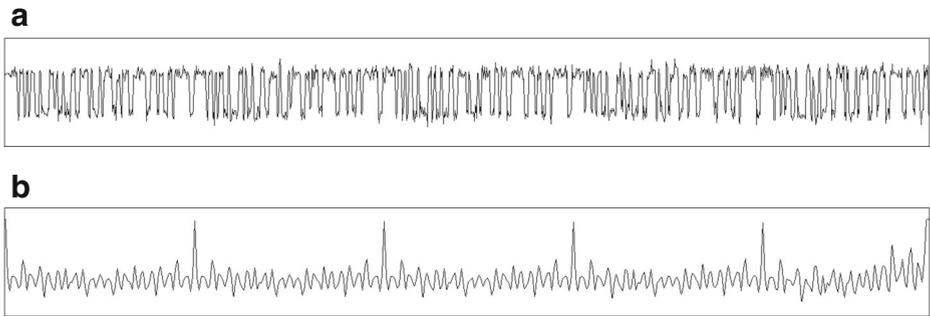


Fig. 4 Real watermark signal extracted from non attacked audio data (a), autocorrelation function of the extracted signal (b)

The distance between peaks is strongly related with the watermark embedding capacity. The information about the peak localization and the validation of the distance between subsequent maxima are sufficient for an accurate localization of the first bit of the watermark.

4 Simulation of attacks

There are several possible attacks that can affect the robustness of the watermark, i.e. resampling, filtration, time warping, cropping, lossy compression, etc. All types of unexpected processing may pose a threat to the extraction of a bit sequence hidden in the embedded watermark. Simple modifications of audio signals are treated as potential watermark attacks described in the literature [17, 23]. Possible processing is not always aimed at watermark deletion or corruption. However either intentional or unintentional, the modifications are sufficiently dangerous and can lead to watermark degradation.

It is infeasible to embed a watermark being robust to all types of attacks. However, within the purpose of the presented research study, the authors performed some most common simulated attacks from real life scenarios. The results of the simulations are presented in Fig. 5. This figure contains extracted difference signals related with watermarks binary sequences (left column) and autocorrelation functions (right column) presenting localizations of the first bit of the watermarks binary sequences. Various types of audio processing were taken into consideration, including audio lossy compression (Fig. 5b), D/A-A/D conversion (Fig. 5c), low-pass filtration (Fig. 5d), high-pass filtration (Fig. 5e), downsampling (Fig. 5f), upsampling (Fig. 5g), denoising (Fig. 5h) and time warping (Fig. 5i). Audio signal attacks were simulated, and then the watermark was extracted. The attacked audio samples were normalized before the watermark extraction to be compatible with the reference signal, the latter being presented in Fig. 5a.

Audio lossy compression (see Fig. 5b) and D/A-A/D conversion (Fig. 5c) were applied to the whole audio signal, other modifications were performed only on a part of the signal, i.e. from the middle to the end of the audio track. The signal differences in the non attacked part of the recordings are unnoticeable in comparison to the attacked part of the track. The differences are also present in the non attacked part. However, they cannot be seen in Fig. 5d–i due to scaling. The crucial issue for the watermark extraction is the autocorrelation

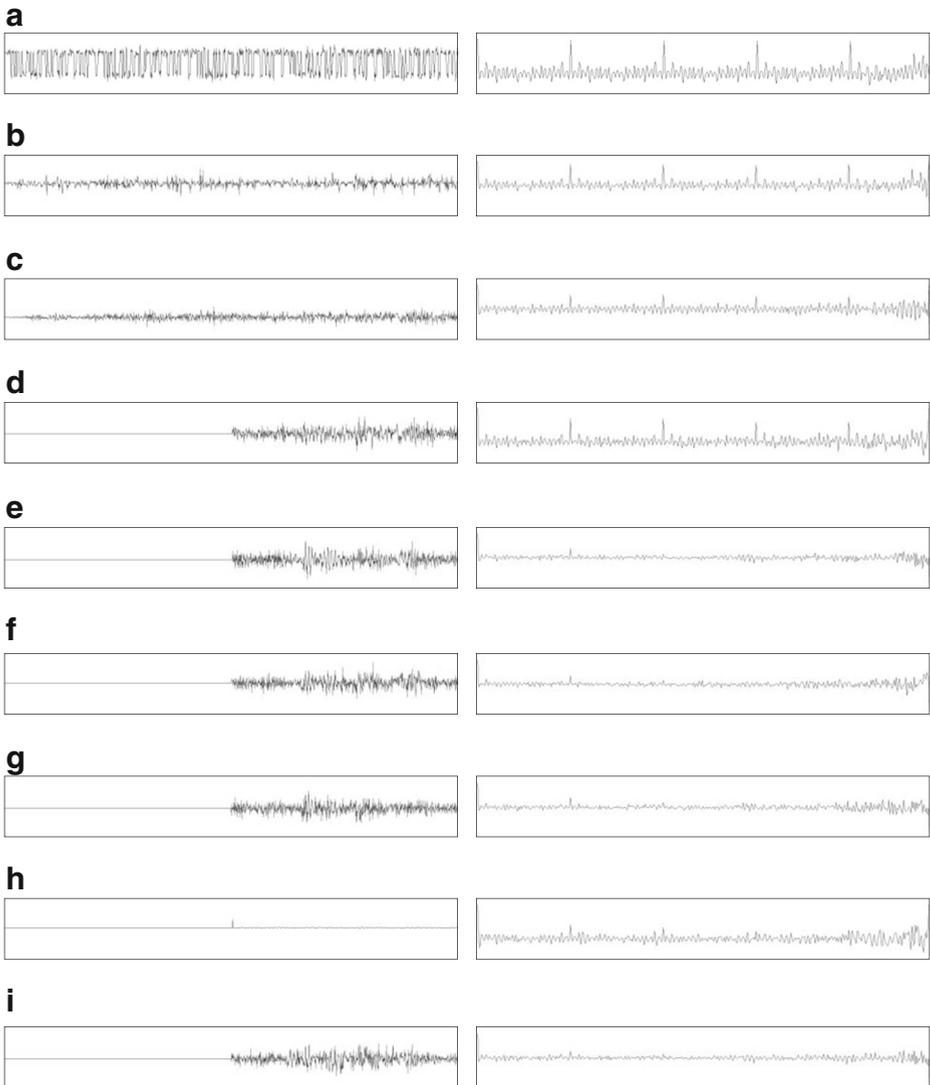


Fig. 5 Attacked watermark extraction: (*left column*) - extracted difference signal, (*right column*) - autocorrelation function of a binary sequence

function, which is required to designate the first bit of the watermark properly. Distinct peaks, visible in the autocorrelation function in Fig. 5 (right column), signify that signal modifications did not affect the watermark. Maxima visible only at the left part of ACF mean that the simulated attack partially misled the watermark extractor. The position of the first bit of the watermark in the attacked part of the track indicates that the watermark cannot be detected.

Simulations of attacks show that the proposed watermarking scheme is not robust to all types of the simulated modifications. The lossy audio compression and low-pass filtration do not harmfully affect the watermark, while the other types of modifications are more destructive.

The lossy audio compression is the most possible attack type in the digital data flow. The commonly used lossy compression formats such as: AAC (Advanced Audio Coding), MP3 (MPEG-1/MPEG-2 Audio Layer 3) and OGG (Xiph.org - Ogg Vorbis) allow the users to modify data size at the cost of data quality. Encoding a file to another format influences the watermark as shown in Fig. 6b. The waveform in Fig. 6a presents the extracted watermark signal, which shows the difference between the original and the watermarked files in the second level low frequency DWT domain. The length of the text watermark is 10 ASCII characters [24]. The same procedures of embedding and extracting the watermark were done for the signal shown in Fig. 6b, after applying the MP3 lossy compression with CBR (constant bitrate) equal to 320 kbps. The audio sample containing a classical music was used in the experiment.

The waveforms presented in Fig. 6 are significantly different, but both contain the same watermark. The implemented extraction algorithm is able to extract the binary watermark sequence without errors from (a) to (b) waveforms. However, the algorithms had to be tested for different music genres, various lossy compression codecs and diverse compression levels to ensure the correctness of the watermark extraction.

To simulate realistic situations the large test set is required. The set employed for simulations was generated automatically considering the variations of four music genres (classical, jazz, pop, rock), three types of lossy compression formats (AAC, MP3, OGG) and six quality levels for each codec. The recompression was executed using the Audacity [2] open source software. Unfortunately, the chosen codecs do not have a unified compression rate which would ensure a similar audio quality for each level of compression. Different quality rates for different codecs are compensated by the ways described in the Audacity software manual [2]. The reference parameter is MP3 CBR bitrate, the minimum value of '128 kbps' in this codec should provide similar audio quality as for value '5' in the OGG quality scale and the value of '250' in the AAC quality scale. The maximum reference value of '320 kbps' in the MP3 codec should provide a similar audio quality as value '10' in the OGG quality scale and the value of '500' in the AAC quality scale. Six quality values were chosen with a linear interval for each scale. Moreover, each music file was watermarked with three different values of the α parameter as in Eq. (1). First, the value of α was heuristically adopted as 10 (base value), mostly because it provided inaudible changes in audio signals as well as errorless watermark detection which is illustrated in Fig. 6. Two other values were chosen as the half ($\alpha=5$) and the double ($\alpha=20$) of the base value.

Finally, a set of 288 audio files including 72 reference audio files (original files, without watermark compressed with specific codec and quality) and 216 watermarked audio files

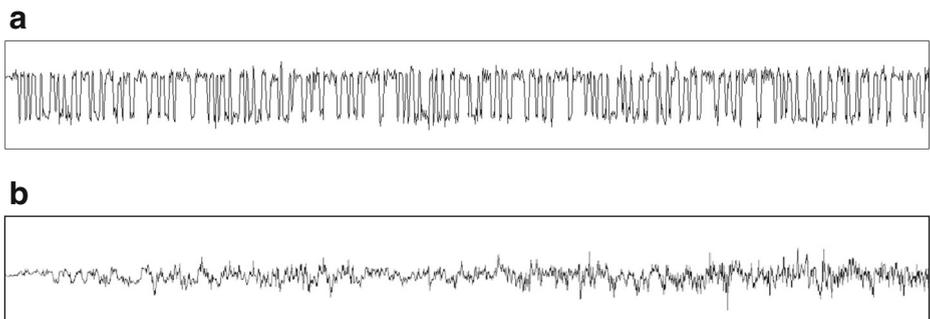


Fig. 6 Watermark signal extracted from: **a** raw audio data, **b** compressed audio data

was generated. The large amount of generated files enabled a simulation of potential attacks as is described in the following Section.

5 Experimental results

In the experiments, the reference watermark signal extracted from the uncompressed files (as shown in Fig. 6) was compared to the watermark signal obtained in the same way from files under the attack with a lossy compression. For this purpose quantitative/empirical PSNR (Peak Signal-to-Noise Ratio) ratio which is often used as the measurement of quality between the reference/original signal and the compressed one. The measurements of differences between the obtained watermark signals are computed according to Eq. (4) [1]:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (4)$$

$$MSE = \frac{1}{l} \sum_{n=0}^l (x[n] - y[n])^2$$

where:

<i>PSNR</i>	peak signal to noise ratio
<i>MSE</i>	mean squared error
<i>MAX</i>	maximum value of used data type
<i>l</i>	total signal length in sample
$x[n]$	current sample value in the reference signal
$y[n]$	current sample value in the analyzed signal

The simplest definition of PSNR origins from MSE which measures the absolute difference between two signals x and y . When MSE equals 0 then signals are identical and PSNR is not computed, otherwise the MSE is compared to the MAX value (in the case of float precision data within the range of $[-1.0-1.0]$, MAX equals 1.0). The PSNR allows to present the results in the dB scale. Lower PSNR values mean that signals differ more, thereby proving that higher values are better in the context of the executed simulations and measures.

The results obtained during the simulations are presented in Fig. 7. The x -axes are related to each other, separately for each codec described in the previous Section, with the normalized quality rates.

The first iteration of measurements (with the base value of $\alpha=10$) was very promising, because no errors occurred. In order to measure the robustness of the proposed watermarking method, we utilized BER (bit error rate) defined as the number of incorrect watermarks to the number of total watermarks. BER was equal to zero in each test case, which proved an errorless watermark extraction. The charts in Fig. 7 present the influence of the compression on the watermarks, and show differences with regard to music types. The PSNR level directly depends on the type of music. The highest PSNR value was obtained for classical music, and the lowest for rock and pop. The increase of PSNR is observed for MP3 and OGG formats. Contrarily, the AAC encoder provides a constant PSNR value irrespective of the resulting quality. In further measurements, the α parameter values were changed to $\alpha=5$ and next to $\alpha=20$ (see Eq. 1). These results are presented in Fig. 8.

The obtained results show errorless watermark extraction even after lossy audio compression has been applied. The PSNR values vary by ca. ± 3 dB compared to the results

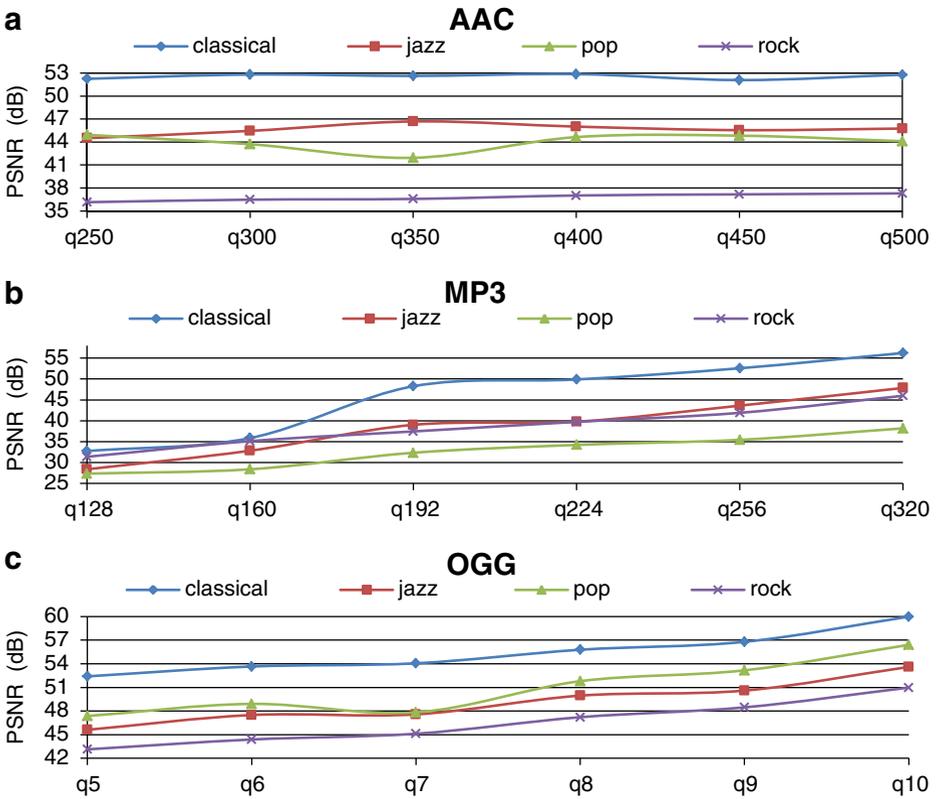


Fig. 7 Results of compression attack simulations for $\alpha=10$: **a** AAC, **b** MP3, **c** OGG

obtained for $\alpha=10$. The watermarks are properly decrypted for each test case, although the encoder and music genres influence PSNR.

The influence of the watermarking strength on watermark extraction was tested with various values of parameter α . Regardless of α value, no errors in the extracted binary watermark were detected. Nevertheless, higher values caused a significant reduction of PSNR for each measurement.

6 Subjective tests

Signal degradation can be assessed through a variety of evaluation techniques including subjective quality measures. Thus, audio signals with the watermark added were tested subjectively. A group of experts had to choose a better audio recording from a pair of audio samples, i.e. the original and the watermarked. There were 20 experts taking part in two test series for three values of the α parameter to enhance the reliability of the experiment. The same subjective test was repeated after 6 months with the same group of experts. Each series contained two pairs of audio samples of specific music genres. If the watermarked file was assessed as better than the original one, one point was added to the pool—otherwise one point was subtracted. The final result for a specific genre shows differences between audio tracks that were compared.

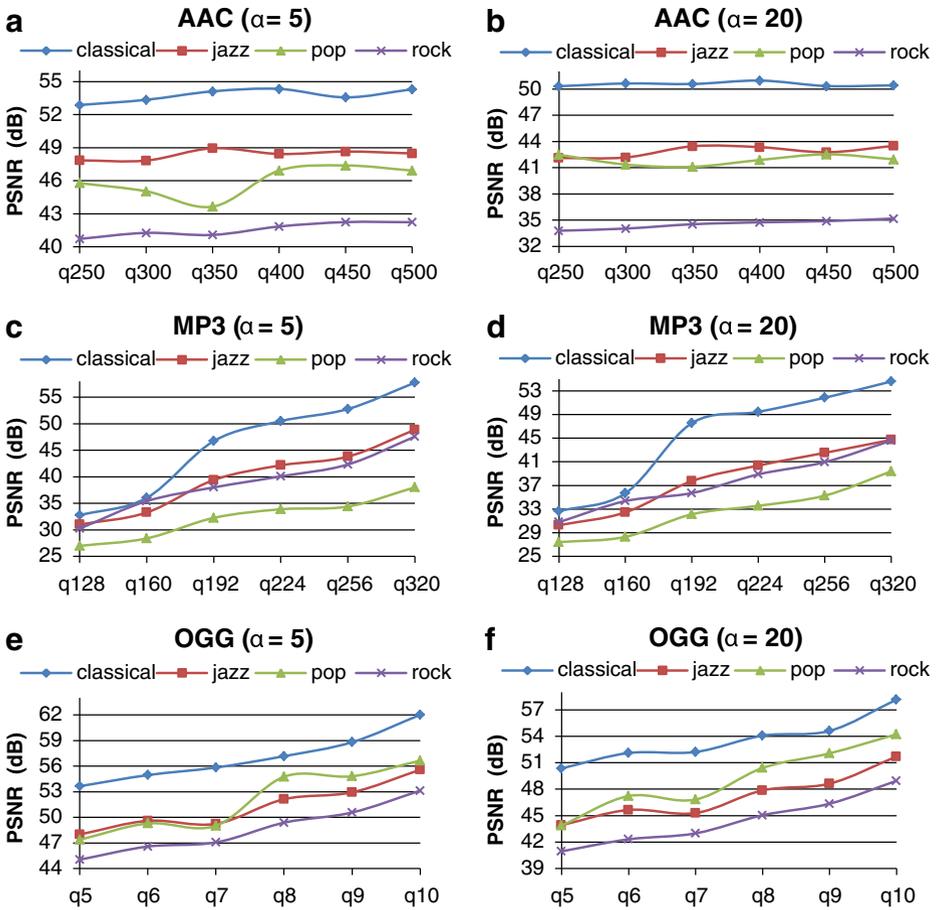
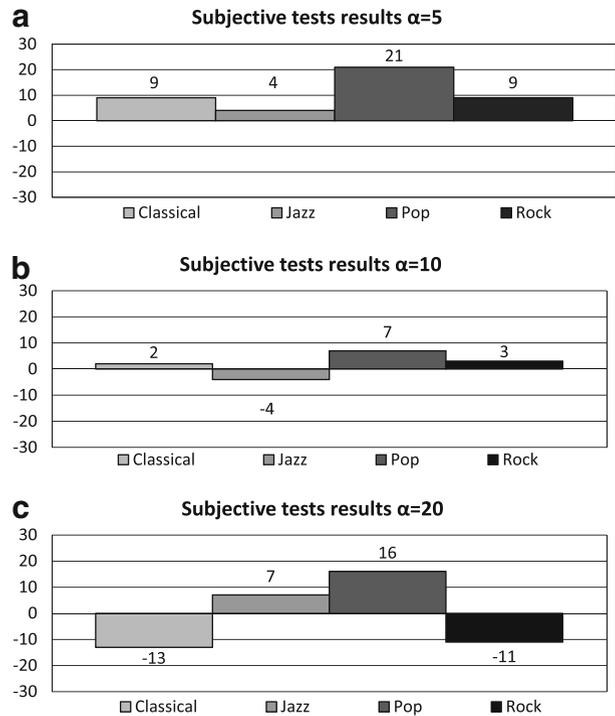


Fig. 8 Results of compression attack simulations for: **a** AAC, $\alpha=5$; **b** AAC, $\alpha=20$; **c** MP3, $\alpha=5$; **d** MP3, $\alpha=20$; **e** OGG, $\alpha=5$; **f** OGG, $\alpha=20$

Positive values mean that the watermarked file was assessed better than the original file, while a negative values signify that the watermark was audible to the expert, a value close to zero means that the watermark was not detected and the audio files were comparable. The aim of the experiments was also to compare the influence of the watermarking strength for three different values of the α parameter, $\alpha=5$, $\alpha=10$ and $\alpha=20$. The obtained results are presented in Fig. 9.

The obtained results demonstrate that the embedded watermark is not always perceptually transparent. Moreover, the results differ for music genres. The degradation of quality caused by the watermarking differs for various cases and depends on both music genre and the watermarking strength. The smallest value of α parameter does not alter the signal quality. In this case the experts assessed each audio sample similarly. The audible distortions were not reported for the watermarked recordings. The resulting sum of ratings has a positive value for each music genre. Moreover, the watermarked audio tracks were assessed higher than the original tracks. This probably occurred due to the

Fig. 9 Results obtained in subjective audio tests **a** $\alpha=5$, **b** $\alpha=10$ **c** $\alpha=20$



slight audio signal colorization which may have a positive impact on the audio quality for all music genres. Increasing the watermarking strength from 5 to 10 led to a positive degradation rate for all audio tracks. The subjective degradation grade for *Jazz* is lower than zero, meaning that experts observed an audio quality degradation for this genre. Other recordings were evaluated positively. Since α equal 10 seems to be the most neutral, all rates are near zero. In the last part of the experiments, larger values of α (10 and 20) were examined. Such modification brought a significant difference in the experts' judgments. Tracks representing *Classical* and *Rock* genres were assessed significantly lower than *Jazz* and *Pop* genres. The artifacts and quality degradation were observed by the experts in those cases. Nevertheless, it is unexpected that *Jazz* and *Pop* were assessed significantly higher despite the use of stronger signal modifications (when α equaled 20).

It may therefore be concluded that a signal quality depends on genre and on the dynamics of the music excerpt. Audible distortions caused by the watermarking may be perceived more for audio genres with a more regular harmonic structure such as *Classical*. Obviously, the strength of the watermarking procedure is crucial for audio quality, greater value of α parameter means higher audio quality degradation.

Since the results seem not entirely conclusive, the explanation for this may be that when signals are of a sufficiently high quality, the performance-based evaluation may not be adequate, because the evaluators turn to opinion-based rating scales. This means that the listeners might tend to express their preference rather than to rate signal quality.

7 Objective analysis

As mentioned before, the results obtained during subjective tests were not entirely conclusive. Therefore, it was decided to perform the objective analysis of audio signal quality degradation. Moreover, the proposed algorithm, its efficiency and quality were compared with one of the state of the art algorithm, i.e. the AWT algorithm [3], which embeds the watermark in time domain by time-shifting the carrier signal blocks in one or several frequency sub-bands. The watermark extraction is similar to the extraction in the proposed architecture. Both the original host signal and the watermarked audio signal must be used.

The watermarked audio examples were obtained using the AWT demo software [4] which enables embedding watermarks of the length used in the described experiments. The software allows choosing the aggressiveness of AWT watermarking procedure, which has the same functionality as the α parameter. The default value of aggressiveness was set to 1.0, however, the audio tracks were also watermarked with the values of 0.5 and 2.0. The obtained tracks were analyzed accordingly to the errorless watermark extraction after signal modifications presented in Section 4. The results of the extraction are collected in Table 2 for both algorithms. Similarities between the proposed approach and the AWT watermarking justified the comparison of algorithms.

The watermarks embedded with the AWT algorithm are robust for all types of attacks except time warping. The proposed algorithm is significantly less resistant to modifications of the watermarked signal. Errorless watermark extraction is possible only after low-pass filtration and lossy compression, other types of modifications destroy synchronization and making watermark extraction impossible.

The ability of errorless watermark extraction is significantly better in the case of AWT than for the proposed methodology. Deeper analysis of the algorithm influence on signal quality is necessary. Therefore, spectral analysis as well as PEAQ (Perceptual Evaluation of Audio Quality) analysis were performed [14]. However only tracks watermarked with the highest strength (α equals 20 or *aggressiveness* equals 2.0) were taken into consideration. An additional analysis was performed of an audio track containing a 440 Hz tone. Figure 10 presents the spectral analysis of 1-s excerpts coming from the recordings of the original and watermarked tracks. The left column, holds the analysis of the original track. The center column presents the analysis of the track watermarked with the proposed method, and the right column shows the analysis of recordings from the track watermarked with the AWT algorithm.

Table 2 Watermark extraction for the proposed and the AWT algorithm

Attack	Proposed algorithm (DWT domain)	AWT algorithm (time domain)
D/A - A/D conversion	✗	✓
Lossy compression	✓	✓
Low-pass filtration	✓	✓
High-pass filtration	✗	✓
Downsampling	✗	✓
Upsampling	✗	✓
Denoising	✗	✓
Time warping	✗	✗

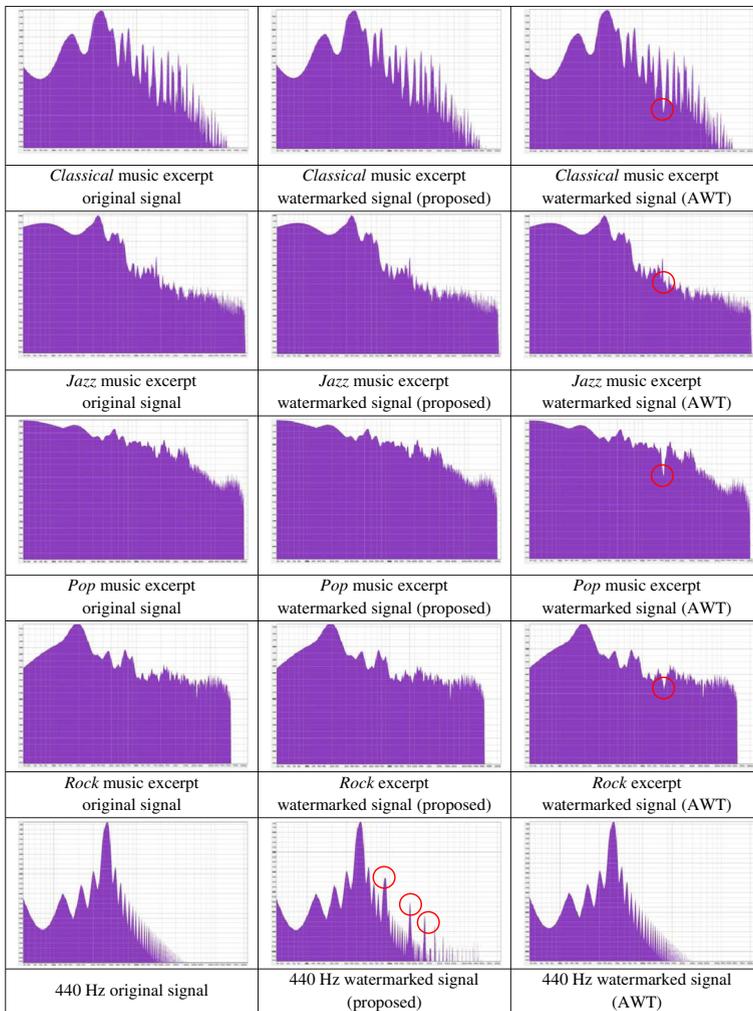


Fig. 10 Spectra of the analyzed music excerpts, *left column* represents original host signal, *center column* – signal watermarked with proposed algorithm, *right column* – signal watermarked with AWT algorithm

Spectra of *Classical*, *Jazz*, *Pop* and *Rock* music recordings watermarked with the proposed algorithm are slightly different than their host signal spectrum. For both signals, only differences in amplitude are observable due to the scaling factor application (see Eq. 1). However, the spectrum of the watermarked tone differs significantly in relation to its host sinusoid. High harmonics and high-frequency noise occurred in this case. The differences are marked with red circles. Spectra of signals watermarked with the AWT algorithm are also similar to their host signal spectrum but for music samples the degradation poses notch filtration around 1,800 Hz as marked with red circles. However, the spectrum of a simple tone looks identical to its original signal spectrum.

To characterize the perceived audio quality as experts would do in listening tests, the PEAQ analysis that enables comparing audio signals based on a human hearing perceptual

model was applied. To characterize the perceived audio quality as in listening tests, the PEAQ analysis that enables comparing audio signals based on a human hearing perceptual model was applied. Moreover, the perceptual audio quality degradation caused by the proposed algorithm was compared with the quality degradation induced by the AWT algorithm. The result of this comparison is presented in the ODG (Objective Difference Grade) scale. Values of the ODG may range from 0.0 to -4.0 , where 0.0 signifies that signals are identical (without differences), and -4.0 means that differences between signals are not acceptable. The measurements were performed employing the advanced mode of PEAQ and the Opera software [20]. The results of the PEAQ analysis are presented in Fig. 11.

Differences between the signals revealed in objective measurements are presented in Fig. 11. The PEAQ analysis only partially confirms the results of the previous subjective assessments. It proves that the quality of music watermarked with the proposed method depends on music genre. The highest quality degradation occurs for the *Classical* genre. The differences between signals for *Jazz* and *Pop* are comparable (see Fig. 11a). Unexpectedly to the results of subjective assessments, the least degraded is *Rock* (see Fig. 9c). The spectral analysis of the 440 Hz tone (the original and watermarked) confirms that watermarking mostly affects quality of audio signals having harmonic structure (see Fig. 11c). The PEAQ analysis for the original and the watermarked sinusoids shows that both signals are significantly different ($ODG < -3.7$). Watermarking with the AWT algorithm generates significantly more observable distortions as can be seen in Fig. 11b. The AWT has negative influence mostly on *Jazz* and *Classical* genres, and then on *Pop* and *Rock*. The quality degradation for the 440 Hz tone is similar to the quality degradation observed for music genres.

Differences in quality between the proposed algorithm and the AWT algorithm are obvious. The quality degradation for music genres caused by the proposed algorithm falls within the range of $[-0.03 - -0.25]$ meaning that it is practically not reduced. On the other hand,

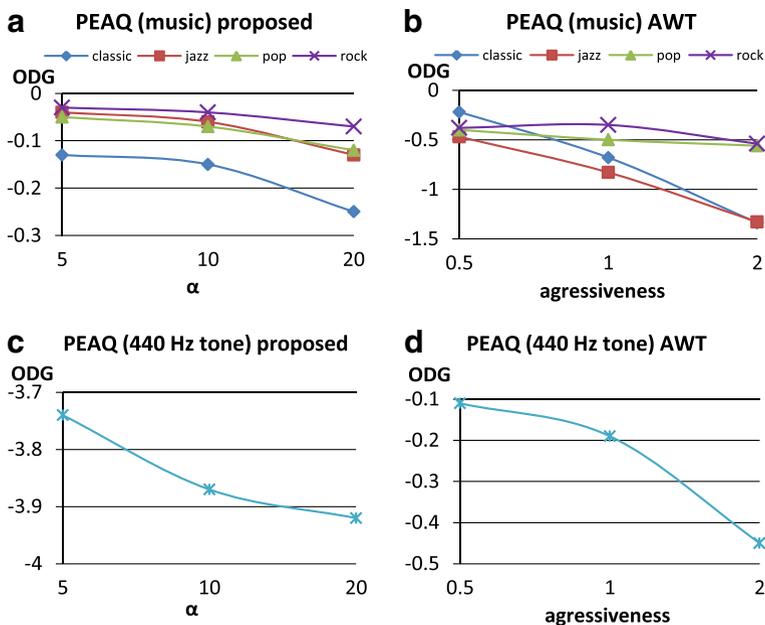


Fig. 11 Results of PEAQ analysis for music genres (a), 440 Hz tone (c) watermarked with the proposed methodology and music genres (b), 440 Hz tone (d) watermarked employing AWT algorithm

the degradation after the AWT watermarking brings results within the range of $[-0.22 - -1.34]$. For human perceptiveness, the ODG lower than -1.0 means significant deterioration of quality. However, the analysis of the 440 Hz tone quality proves that for these types of signals the AWT method is considerably better than the proposed scheme.

8 Conclusions

The proposed watermarking system is still at the stage of development, but the results obtained so far are very promising. The extraction of watermark from files that were attacked with lossy compression were successful – the obtained BER was equal to 0. It brought repeatable, quantitative data that could further be exploited in the development of new watermarking techniques. The results of subjective assessment of the watermark influence on its audibility with regard to music genre were not entirely conclusive. However, the results of two sessions of subjective tests performed after a 6 months break were mutually consistent. The objective analysis of watermarked signals showed that the inconsistency in the evaluation of watermarked music excerpts with regard to music genre may occur due to the introduced high frequency noise (and harmonics), which in the case of some music genres can be regarded as acceptable by some listeners.

In comparison to the AWT method, the proposed methodology is less resistant to several types of attacks. Actually, only signals modified by lossy compression and low-pass filtration can be decoded errorless. On the other hand, the perceptual quality degradation caused by the proposed algorithm is significantly lower than for the AWT algorithm. The future research will be orientated towards improving the proposed algorithm robustness against other types of signal processing attacks, such as: multiple compression and/or stereo downmix to mono, because these attacks are in general very damaging to audio signals.

Moreover, further tests involving other music genres should be performed to ensure the imperceptiveness of the inserted watermark and to verify the perceived quality of the modified audio files.

Acknowledgments The research was founded within the project No. SP/I/1/77065/10 entitled: “Creation of universal, open, repository platform for hosting and communication of networked resources of knowledge for science, education and open society of knowledge”, being a part of Strategic Research Program “Interdisciplinary system of interactive scientific and technical information” supported by The National Centre for Research and Development (NCBiR, Poland).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Al-Haj AM (2012) Advanced techniques in multimedia watermarking: image, video and audio applications. IGI Global, New York
2. Audacity (2012) Audacity manual. http://manual.audacityteam.org/man/Main_Page. Accessed 18 December 2012
3. Audio Watermarking Toolkit patent - U.S. Patent no. 8,116,514
4. AWT (2013) Audio Watermarking Toolkit homepage. <http://audiowatermarking.info>. Accessed 27 May 2013
5. Bhatnagar G, Raman B, Wu QMJ (2012) Robust watermarking using fractional wavelet packet domain. IET Image Process 6(4):386–397. doi:10.1049/iet-ipr.2010.0400

6. Bhatnagar G, Wu QMJ, Raman B (2011) A new aspect in robust digital watermarking. *Multimed Tools Appl*. doi:10.1007/s11042-011-0788-z
7. Bloom JA, Cox IJ, Fridrich J, Kalker T, Miller ML (2008) *Digital watermarking and steganography*. Morgan Kaufman, Boston
8. Ciarkowski A, Czyżewski A (2010) Performance of watermarking-based DTD algorithm under time-varying echo path conditions. *Intell Interact Multimed Syst Serv* 6:69–78. doi:10.1007/978-3-642-14619-0
9. Cichowski J, Czyżyk P, Kostek B, Czyżewski A (2013) Low-level music feature vectors embedded as watermarks. *Intell Tools Build Sci Inf Platf Stud Comput Intell*. doi:10.1007/978-3-642-35647-6_27
10. Czyżyk P, Cichowski J, Czyżewski A, Kostek B (2012) Analysis of impact of lossy audio compression on the robustness of watermark embedded in the DWT domain for non-blind copyright protection. *5th Int Conf on Multimed Commun, Serv and Secur* 1:36–46. doi:10.1007/978-3-642-30721-8_4
11. Dutta MK, Gupta P, Pathak VK (2009) Perceptible audio watermarking for digital rights management control. *7th Int Conf on Inf, Commun and Signal Process* 1:55–59. doi:10.1109/ICICS.2009.5397484
12. Foo SW, Dong Q (2010) Audio watermarking based on compression-expansion technique. *World Acad Sci Eng Technol* 37:680–686
13. Furht B, Kirovski D (2006) *Multimedia encryption and authentication techniques and applications*. Auerbach Publications, Florida
14. ITU-R BS.1387-1 (2001) Method for objective measurements of perceived audio quality
15. Janus Patent - U.S. Patent No. 7,010,808
16. Kondo K (2011) A data hiding method for stereo audio signals using interchannel decorrelator polarity inversion. *J Audio Eng Soc* 59(6):379–395
17. Lang A, Dittmann J, Spring R, Vielhauer C (2005) Audio watermark attacks: from single to profile attacks. *Proc of the 7th Workshop on Multimed and Secur* 1:39–50. doi:10.1145/1073170.1073179
18. Maha C, Maher E, Chokri BA (2008) A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain. *3rd Int Symp on Commun, Control and Signal Process* 1:1138–1143. doi:10.1109/ISCCSP.2008.4537396
19. Microsoft Windows Media Digital Rights Management homepage. <http://www.microsoft.com/windows/windowsmedia/licensing/default.mspx>. Accessed 27 May 2013
20. OPERA™ Voice and Audio Quality Analyzer, <http://www.opticom.de/products/opera.html>. Accessed 08 April 2013
21. Pérez-Freire L, Pérez-González F (2009) Spread spectrum watermarking security. *IEEE Trans Inf Forensic Secur* 4(1):2–24. doi:10.1109/TIFS.2008.2009603
22. Pérez-Freire L, Comesana P, Troncoso-Pastoriza JR, Pérez-González F (2006) Watermarking security: a survey. *Trans Data Hiding Multimed Secur* 4300:41–72
23. Petitcolas FAP, Anderson RJ, Kuhn MG (1998) Attacks on copyright marking systems. *Proc of the 2nd Int Workshop on Inf Hiding*, Springer-Verlag, London, 1:218–238
24. RFC 20: ASCII format for Network Interchange, ANSI X3.4, October 1969
25. Samaali I, Mahé G, Turki M (2012) Watermark-aided pre-echo reduction in low bit-rate audio coding. *J Audio Eng Soc* 60(6):431–443
26. Singh J, Garg P, De A (2012) Multiplicative watermarking of audio in DFT magnitude. *Multimed Tools Appl*. doi:10.1007/s11042-012-1282-y
27. Subramanyam AV, Sabu E (2012) Partially compressed-encrypted domain robust JPEG image watermarking. *Multimed Tools Appl*. doi:10.1007/s11042-012-1272-0
28. Szwoch G, Czyżewski A, Ciarkowski A (2009) A double-talk detector using audio watermarking. *J Audio Eng Soc* 57(11):916–926
29. Taneja N, Bhatnagar G, Raman B, Gupta I (2012) Joint watermarking and encryption for still visual data. *Multimed Tools Appl*. doi:10.1007/s11042-012-1037-9
30. Venkataramu R (2007) Analysis and enhancement of Apple's fairplay digital rights management. Project report. The Faculty of The Department of Computer Science, San Jose State University
31. Wassermann J (2013) New robust video watermarking techniques based on DWT transform and spread spectrum of basis images of 2D hadamard transform. *Multimed Commun Serv Secur Commun Comput Inf Sci* 368:298–308. doi:10.1007/978-3-642-38559-9_26
32. Youarchive audio reconstruction system. <http://www.youarchive.net>. Accessed 27 May 2013
33. Tang Z, Wang S, Zhang X, Wei W, Zhao Y (2012) Lexico-graphical framework for image hashing with implementation based on DCT and NMF. *Multimed Tools Appl*. doi:10.1007/s11042-009-0437-y



J. Cichowski is a native of Gdansk, Poland. He received his M.Sc. degree in Sound and Visual Engineering from Gdansk University of Technology in 2011. He graduated from the Multimedia Systems Department. His subject of M.Sc. thesis was related to digital watermarking and hiding image sensitive objects in surveillance systems with regards to privacy issues. At present he is a Ph.D. student of Multimedia Systems Department. He co-authored research papers published and presented in conferences related to privacy enhancement and anonymization for smart surveillance in the public domain. Mr. Cichowski is a member of the Polish Section of the Audio Engineering Society.



A. Czyżewski is a native of Gdansk, Poland. He received his M.Sc. degree in Sound Engineering from the Gdansk University of Technology in 1982, his Ph.D. degree in 1987 and his D.Sc. degree in 1992 from the Cracov Academy of Mining and Metallurgy. He joined the staff of the Sound Engineering Department of the Gdansk University of Technology in 1984. In December 1999 Mr. President of Poland granted him the title of Professor. In 2002 the Senate of his University approved him to the position of Full Professor. He is an author of more than 400 research papers published in international journals or presented in congresses & conferences around the World. In 1991, he published a monograph devoted to digital audio operations, in 1998 he published in Poland his book entitled “Digital Sound” which won him the Prize of Ministry of Education of Poland. He is also author of 10 Polish patents in the domain of computer science and 5 international patent applications. Prof. Czyżewski serves as Head of the Multimedia Systems Department of Gdansk University of Technology; Director of the Doctoral Studies (1994–2007) at the Faculty of Electronics, Telecommunications & Informatics. He holds Fellowship of the Audio Engineering Society and he is a member of: IEEE, International Rough Set Society, and others.



B. Kostek holds professorship at the Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology (GUT), Poland. She is Head of the Audio Acoustics Laboratory. She received her M.Sc. degrees in Sound Engineering (1983) and Organization and Management (1986) from GUT. She also received postgraduate DEA degree (1988) from Toulouse University, France. In 1992 she supported her Ph.D. thesis with honors at GUT, and in 2000 her D.Sc. degree at the Research Systems Institute, Polish Academy of Sciences. In 2005 President of Poland granted her the title of Professor. She published over 450 scientific papers in journals and at international conferences. She has served as the Editor-in-Chief of *Archives of Acoustics* (2007–2012) and *J. Audio Eng. Soc.* EIC since 2011. She was the recipient of many prestigious awards for research, including those of the Prime Minister of Poland for outstanding research achievements, prizes of the Polish Academy of Sciences and Ministry of Science. She also received the Audio Eng. Soc. Fellowship Award in 2010. Her research activities are interdisciplinary, however the main research interests focus on audio signal processing, human-computer interaction, cognitive bases of sound and vision processing, QoE.