

Robust audio watermarking in the time domain

P.Bassia and I.Pitas

Dept. of Informatics, University of Thessaloniki

Thessaloniki 540 06, GREECE

Tel: +30 31 996361; fax: +30 31 996304

e-mail: vbassia@zeus.csd.auth.gr pitas@zeus.csd.auth.gr

ABSTRACT

The audio watermarking method presented below offers copyright protection to an audio signal by modifying its temporal characteristics. The amount of modification embedded is limited by the necessity that the output signal must not be perceptually different from the original one. The watermarking method presented here does not require the original signal for watermark detection. The watermark key is simply a seed known only by the copyright owner. This seed creates the watermark signal to be embedded. Watermark embedding depends also on the audio signal amplitude in a way that minimizes the audibility of the watermark signal. The embedded watermark is robust to MPEG audio coding, filtering, resampling and requantization.

1 Introduction

The outstanding progress of digital technology has increased the ease with which digital data is reproduced and retransmitted. However, since the advantages of such a progress are broadly available, they offer equally increasing potential to both legal and unauthorized data manipulation. Consequently, the necessity arises for copyright protection of digital products against unauthorized recording attempts, known as data piracy.

Current research in image, audio and video copyright protection exploits the fact that the human visual and audio perception cannot detect slight changes in certain temporal or frequency domains of the image and the audio signal, respectively. This property is called *masking*, according to which a faint but perceptible signal becomes non-perceptible in the presence of another one under certain conditions.

Most research methods consider a watermark signal produced in a unique way by a function of one or more input keys. These keys can be both owner and signal dependent and generate a signal which is embedded on the original one. The embedding signal is known as a *watermark* or *copyright label*. Temporal and frequency characteristics of the original signal should be taken into account in the watermark casting process to reduce perceptible distortions in the watermarked signal. Each

individual that produces or possesses digital data owns a unique key that identifies its legal possession and is required for the watermark detection. Besides copyright purposes, a watermark serves authentication purposes, as well.

Methods that are discussed in this paper investigate the watermarking potential in audio signals, taking into account the specifications of the human audio perception. An audio watermark is a perceptually inaudible modification of the audio signal, based on one or more keys determined by the copyright owner [1]. Whenever a question about ownership arises, keys are used to determine the rightful owner.

Our watermark is embedded in the spatial domain and it modifies each audio sample according to its amplitude. Unlike other methods, it does not require the original signal for its detection. This way the owner of the data does not have to keep double copies of both original and watermarked products.

A watermark has to be statistically undetectable by others to prevent the efforts of its unauthorized removal. This condition is fulfilled if the potential number of keys that produce distinct watermarks is large enough to ensure statistical safety. The detection scheme should be as statistically reliable as possible. False rejection or acceptance of the existence of the watermark should be minimal. Finally, a watermark has to be robust to signal manipulation and impossible to be removed without significant alteration of the signal. In other words, a pirate should have to destroy the audio signal before he accomplishes to destroy the watermark. The robustness should extend to common signal processing operations, such as filtering, compression, resampling, requantization, cropping, noise, D/A conversion. Our watermark scheme fulfills most of the above conditions up to a satisfactory level and current research is being held out to improve the robustness of the algorithm to any form of unauthorized manipulation.

In the audio watermarking area there are methods that use the frequency domain [1, 2]. Some of them exploit the frequency characteristics of the audio signal in order to embed the watermark, minimizing audible dis-

tortions even for high amplitude watermarks. However, most of the above methods require the original signal in order to detect the watermark.

In images, there are methods that cast a watermark either in the spatial [3, 4, 5] or in the frequency domain [2, 6]. The casting signal is generated in a random way and digital data is often divided in casting subsets in order to increase the robustness of the detection scheme to signal processing. In both cases, the amplitude of each sample of the watermarking signal is either constant or calculated as a function of the amplitude of the original sample. Our audio watermarking method has certain similarities with a method used for image watermarking [3].

2 Watermark Embedding

The watermark embedding scheme proposed in this paper modifies original digital audio signals, which are represented as 16-bit or 8-bit sample sequences, by changing the least significant bits of each sample. The result is a slight amplitude modification of each sample in a way that does not produce any perceptual difference. Let us assume an audio signal of N samples $x(i)$, $i = 1, \dots, N$, where N is at least 88200 (corresponding to one second of 44.1kHz stereo audio signal). In order to embed a watermark we modify each sample using a function $f(x(i), w(i))$, where $w(i)$ is the watermark signal in the range $[-\alpha, \alpha]$ and α is a constant. Function f takes the basic audio masking properties into account to prevent watermark audibility. The watermarked sample $y(i)$ is therefore:

$$y(i) = x(i) + f(x(i), w(i)) \quad (1)$$

The *signal to noise ratio* is calculated by

$$SNR = 10 \log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - y(n)]^2}$$

Here, it is important to denote that the random generator w should provide statistically equal numbers of discrete output values in order for the detection procedure to function more accurately. The robustness of the watermark generally increases with the amplitude of the watermarking signal, but the noise poses a limit to this increase.

3 Watermark Detection

Let us denote by S the following sum:

$$S = \sum_{i=1}^N y(i)w(i) \quad (2)$$

By combining (1), (2) we get:

$$S = \sum_{i=1}^N [x(i)w(i) + f(x(i), w(i))w(i)] \quad (3)$$

The first sum in (3) is zero if the random generator produces equal numbers of discrete output values and the signal mean value m_x is equal to zero. In case some random output values are more frequently produced this difference, denoted as Δw , must be taken into account.

Therefore, in (3)

$$S = \sum_{i=1}^{N-\Delta w} x(i)w(i) + \sum_{i=1}^{\Delta w} x(i)w(i) + \sum_{i=1}^N f(x(i), w(i))w(i) \quad (4)$$

$\sum_{i=1}^{N-\Delta w} x(i)w(i)$ is approximated by zero as explained above. If no watermark has been embedded on the signal, S is approximately equal to $\frac{\Delta w}{N} \sum_{i=1}^N x(i)w(i)$. On the other hand, if the signal is watermarked, S is approximately equal to $\frac{\Delta w}{N} \sum_{i=1}^N x(i)w(i) + \sum_{i=1}^N f(x(i), w(i))w(i)$. However, $x(i)$ is of the original signal which cannot be used in the detection process. $x(i)$ can be replaced by $y(i)$ in the last two terms of (4) without significant error. This replacement leads to the replacement of $\sum_{i=1}^{\Delta w} x(i)w(i)$ by $\frac{\Delta w}{N} S$.

Therefore, by subtracting the amount $\frac{\Delta w}{N}|S|$ from S and dividing the result by $\sum_{i=1}^N f(y(i), w(i))w(i)$, the result r is approximately normalized to 0 or 1. The watermark detector used in this method produces the *detection value* r given by:

$$r \triangleq \frac{S - \frac{\Delta w}{N}|S|}{\sum_{i=1}^N f(y(i), w(i))w(i)} \quad (5)$$

The detection value theoretically lies between 0 and 1, yet the approximation of $x(i)$ by $y(i)$ introduces an inaccuracy which slightly expands the interval $[0, 1]$ to $[0 - \varepsilon, 1 + \varepsilon]$. Experimentally, a watermark threshold may be embedded above 0.5, in order to decide whether a certain watermark exists in the signal. The threshold used can be increased if we require increasing certainty in relation to the watermark detection.

Figure 1 illustrates the empirical pdf of the detection value r in a watermarked and a non-watermarked signal. The empirical pdf of a non-watermarked signal is represented by the solid curve, whereas the dashed curve shows the empirical pdf function of the watermarked signal. Both distributions have been calculated using 1000 different watermarks with $SNR = 26$.

The embedding of multiple watermarks causes incremental audible distortion. The maximum number of multiple watermarks to be embedded without audible distortion depends on the amplitude of each watermarking signal. The worst case considers that the first watermark reaches the threshold of noise, so the second one will be audibly perceptible even if it is of low amplitude. However, multiple watermarks are being detected in a signal with equal success. In conclusion, a multiple watermarking scheme is possible as soon as the combination of the watermarks is chosen carefully in a way that the whole acoustic result is not audibly perceptible.

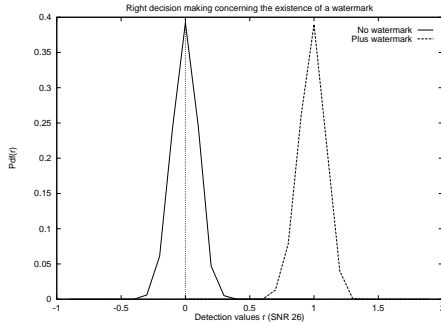


Figure 1: Pdf of detection values in non-watermarked and watermarked signals.

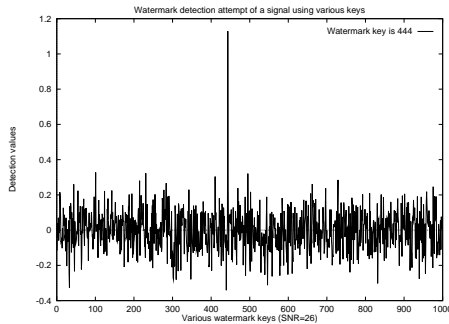


Figure 2: Detection values in a watermarked signal using various seeds (Key is 444).

The detection algorithm gives no false alarm even when it is tested on a watermarked signal using other wrong watermark keys. For example, 1000 keys, of which only one was valid, have been used to detect a watermark in a watermarked signal. The results are being shown in Figure 2.

4 Audio Watermark Robustness to Signal Manipulation

4.1 Robustness to MPEG2 audio compression

The robustness of the watermark technique described above, has been tested using Layers II and III of the MPEG Audio. Several 16-bits signed stereo 44.1KHz watermarked signals were encoded using 80 kbps rates in Layer III and 48kbps rates in Layer II. The lower the rate the bigger is the compression ratio. The watermark resists the encoding-decoding process as shown in Figure 3(a). In this Figure 1000 distinct watermarks ($SNR = 26$) were embedded into an audio signal producing equal number of distinct watermarked signals. All were compressed by MPEG2 audio Layer III with a compression rate of 80kbps. Watermarking detection after decompression indicates a slight decrease of the detection values in the watermarked signals. The solid curve shows the pdf of the detection values after MPEG2 in comparison with the dashed curve which indicates the pdf of the same values before MPEG2. Since the detection ratio is always above the 0.5 threshold, we have

100% success in watermark detection in this experiment. Layer II 48kbps causes an audible distortion to the watermarked signal, yet the watermark detection is being held out successfully.

4.2 Robustness to Filtering

The robustness of the watermark procedure described in this paper was studied under moving average and other low-pass filtering. Watermarked audio files were filtered by a moving average filter of length 20 which introduces a noticeable audible distortion, yet the watermark is detected. Figure 3(b) shows the alteration in detection values introduced by the use of the above mean filter. In general, the detection values are increased in the filtered signals and this is the reason why the pdf of the detection values of the filtered signals in 3(b) is translated to the right in comparison with the non-filtered respective pdf.

Our watermark is also robust to LowPass Filtering. 1000 watermarked audio signals sampled at 44100Hz were filtered by a 25th order Hamming LowPass filter with cut-off frequency 2205Hz. The solid curve of figure 3(c) displays the pdf of the detection values after filtering, while the dashed curve indicates the respective pdf before filtering. It is obvious that the mean value of the solid (filtered) curve is increased in comparison with the dashed (non-filtered) one. The deviation of the solid curve is increased as well, therefore slightly reducing the robustness of the detection scheme after LowPass Filtering. In both experiments, we have 100% success in watermark detection.

4.3 Robustness to Resampling and Requantization

Watermarked audio signals sampled at 44100Hz have been resampled down to 22050Hz and 11025Hz and back again to their initial frequency. Although the above processing caused noticeable distortion in relation to the original signals, the watermarks remained easily detectable. Figure 4 shows how the watermark is retained in 1000 watermarked signals that have been resampled down to 11025Hz and back to their initial 44100Hz frequency. In this experiment we have 100% success in watermark detection.

Requantization of the original 16-bit audio signal down to 8-bit samples and backwards conserves the embedding watermark despite the loss of information during the processing. The watermarks resist the requantization process because it is amplitude adaptable with respect to the original signal. Figure 5 shows how the watermark is retained in 1000 watermarked signals that have been requantized down to 8-bit and back to 16-bit. The deviation of the requantized pdf is increased as in LowPass Filtering, thus reducing the robustness of detection. In this experiment we have 99,8% success in watermark detection. All above experiments were held using the same watermark parameters $SNR = 26$.

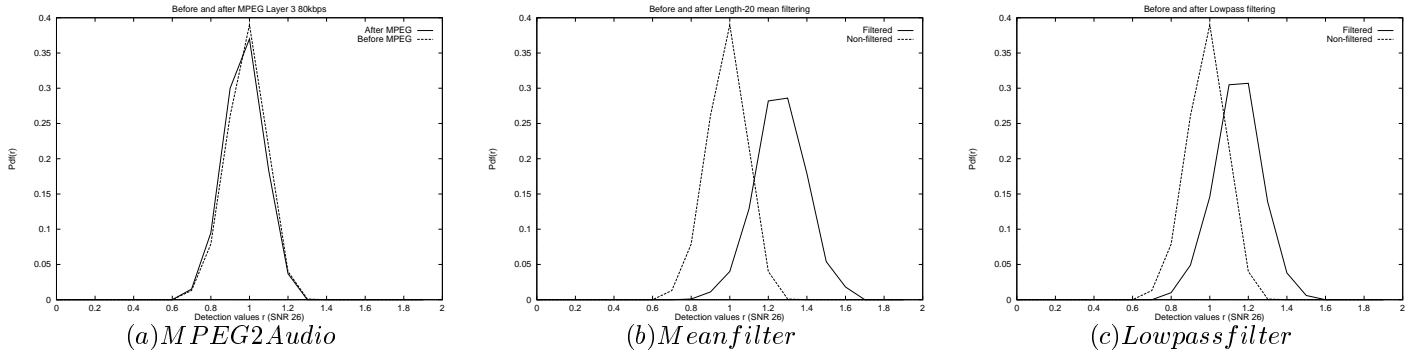


Figure 3: Empirical pdf of detection values before and after MPEG2 Audio, Mean and LowPass filtering on watermarked audio signals.

5 Conclusions

The watermarking scheme presented above embeds a watermark in the time domain of a digital audio signal by slightly modifying the amplitude of each audio sample. The characteristics of this modification are determined both by the original signal and the copyright owner. The detection procedure does not use the original signal. Our watermarking scheme is statistically imperceptible and resists MPEG compression plus other forms of signal manipulation, such as filtering, resampling and requantization.

References

- [1] L. Boney, H. Tewfik, K. N. Hamdy. *Digital Watermarks for Audio Signals, EUSIPCO 96, Trieste, Italy, 1996*
- [2] I.J.Cox, J.Kilian, T.Leighton, T.Shamoon. *Secure spread spectrum watermarking for Multimedia, Tech. Report, NEC research Institute, 95-10, p.1.*
- [3] G. Voyatzis, I. Pitas. *Embedding Robust Logo Watermarks in Digital Images, DSP'97, vol. 1, pp. 213-216*
- [4] N. Nikolaidis, I. Pitas. *Copyright protection of images using robust digital signatures, ICASSP-96, vol. 4, pp. 2168-2171*
- [5] I. Pitas. *A method for Signature Casting on Digital Images, Proc. 1996 IEEE Int. Conference on Image Processing (ICIP 96), vol III, pp 215-218*
- [6] A. Bors, I. Pitas. *Image Watermarking using DCT domain constraints, Proc. 1996 IEEE Int. Conference on Image Processing (ICIP 96), vol III, pp 231-234*
- [7] D.Pan *A Tutorial on MPEG/Audio Compression, IEEE Multimedia Summer 1995*

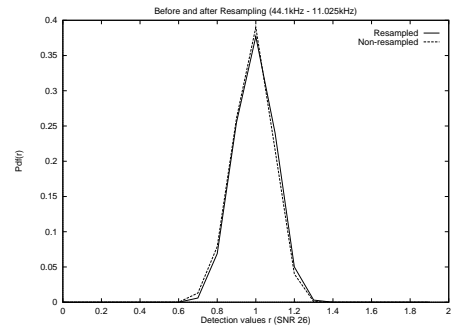


Figure 4: Empirical pdf of detection values before and after Resampling on watermarked audio signals.

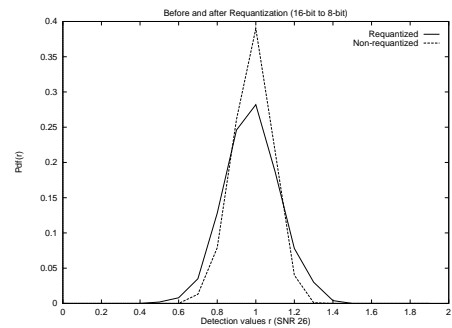


Figure 5: Empirical pdf of detection values before and after Requantization on watermarked audio signals.