

# ROBUST MULTIBIT AUDIO WATERMARKING IN THE TEMPORAL DOMAIN

*Charalampos Laftsidis, Anastasios Tefas, Nikolaos Nikolaidis, Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece,  
e-mail: pitas@zeus.csd.auth.gr

## ABSTRACT

In this paper, a technique for temporal embedding of a multibit watermark on an audio signal is proposed. The method is based on generating a chaotic sequence which is used for modifying the audio samples. A chaotic sequence generates the watermark, while the parameters of the embedding procedure are chosen so as to minimize the perceivable distortion of the initial signal, while preserving the power of the watermark at a detectable level. The watermark is resistant to lowpass manipulation attacks, such as filtering or compression, and cropping attacks as well.

## 1. INTRODUCTION

The progress of digital technology and its spreading among people has made reproduction of multimedia information a common task. This has also made illegal manipulation of people's intellectual property possible. Watermarking techniques provide a way to authenticate the rightful owner of a multimedia piece. Comprehensive reviews on existing watermarking techniques that have been developed can be found in [3, 6].

In this paper, an audio watermarking technique is proposed. The watermark is required to be inaudible. It should also be detectable after the watermarked signal has been manipulated through any sort of time or frequency domain operation. Thus, the quality of the signal should be considerably degraded before the watermark is unretrievable. The proposed method is a multibit extension of the watermarking technique, presented in [1]. The original audio signal is not required during detection as in [4, 2] and the method withstands attacks of lowpass nature, such as compression. The proposed method is also robust against cropping of the watermarked signal. Chaotic number generators are used in the watermark generation procedure since they attain controllable spectral characteristics [7, 5].

## 2. WATERMARK GENERATION

A chaotic random number generator creates the multibit watermark, which is to be embedded on the original signal. A chaotic function  $F$  is applied recursively on an initial value  $z[0]$ , i.e. the watermark key.

$$z[i + 1] = F(z[i]) \quad (1)$$

where  $F$  is a nonlinear transformation, that maps scalars to scalars. In the proposed implementation the Renyi map  $F : U \rightarrow U$ ,  $U \subset \mathcal{R}$  was used. The procedure is repeated until a series of  $N$  numbers is produced:

$$z[i + 1] = (\lambda \cdot z[i]) \bmod \pi, \quad \lambda \in \mathcal{R}, i = 0 \dots N - 2 \quad (2)$$

where parameter  $\lambda$  controls the spectral characteristics of the sequence. An appropriate threshold is chosen, in order to generate a bipolar signal  $w(i)$  of values +1 and -1, which is the basic watermark and which is used to generate the multibit watermark. This threshold is selected so that symbols +1 and -1 are equiprobable.

The chaotic nature of the watermark ensures that low correlation values occur between different watermarks. Furthermore, the chaotic sequences have controllable spectral characteristics and superior performance than pseudorandom sequences [5]. As this watermark is intended to be inaudible and robust to lowpass processing, it should not contain high-frequency components. By choosing  $\lambda > 1$  and close to 1, the resulting watermark has lowpass characteristics.

## 3. WATERMARK EMBEDDING

During this procedure the basic watermark  $w(i)$  is modified, so as to encode an  $M$ -bit long bitstream  $m$ , and is embedded on the host signal  $f$ , in order to obtain the watermarked signal  $f_w$ . The host signal  $f$  is divided in  $M$  segments of  $N$  samples each. For synchronization purposes two additional segments are needed. Thus, the minimum length of the original signal should be at least  $(M + 2) \cdot N$  samples.

An important feature of the proposed method is that the basic watermark is not embedded directly in the beginning

of the audio signal, but on the sample, among the first  $N$  ones, that turns out to be optimal. This assures that the watermark is safely detectable with minimal interference from the host signal. In order to find the most appropriate sample for embedding, the following technique is used.

The host signal  $f$  is divided into  $M + 2$  segments of  $N$  samples:  $f_k(i) = f(k \cdot N + i)$ ,  $i = 0 \dots N - 1$  and  $k = 0 \dots M + 1$ . The correlation  $c_k(i)$  of the segments  $f_k(i)$  and the basic watermark  $w(i)$  is calculated for every  $k$  and then the sum of the absolute values with respect to  $k$ ,  $C(i)$ , is taken:

$$c_k(i) = \sum_{j=0}^{N-1} f(k \cdot N + i + j) \cdot w(j) \quad (3)$$

$$C(i) = \sum_{k=0}^{M+1} |c_k(i)| \quad (4)$$

This assures that the watermark is immune to cropping attacks, since its detection is not based on its location, as will be shown in the corresponding section. Embedding the watermark on the host signal should begin from that sample  $p$ , among the first  $N$  ones, where the lowest value of  $C(i)$  appears.

$$p = \arg \min_i (C(i)) \quad (5)$$

The basic watermark is repeatedly embedded on the audio signal. Each basic watermark encodes one bit of the sequence. If the bit is equal to one, then the basic watermark is casted as is. Otherwise, its negative is used.

A problem that usually appears in a multibit watermark system is locating the beginning of the bitstream in case of cropping. Casting the first bit on the first basic watermark is not an adequate approach, since a cropping attack makes the recovery of the bitstream impossible. The bits should be encoded, so as to enable their recovery even if only a part of the audio stream is given for detection (provided that it is long enough to contain the full bitstream). Thus, an effective method to locate the first bit of the sequence has been implemented.

One part of the host signal, that will be called as the null-bit, where a basic watermark should have been embedded, is left unwatermarked and denotes that the next basic watermark period contains the first bit of the sequence. It is therefore important that this null-bit is indeed detected as a gap among the watermarked ones. Therefore, the correlation of the basic watermark and the audio signal, at that period, should be significantly different than the correlation values obtained from the other watermarked periods. Thus, the optimal position  $g$  of the null-bit is that period, among the first  $M + 1$  ones, which gives the minimum absolute correlation value with the watermark.

$$g = \arg \min_k (|c_k(p)|) \quad (6)$$

If the host signal is longer than the watermark, the procedure continues embedding the bits until the end of the signal. However, the next null-bit, which will have to appear in bit number  $g + M + 1$ , will probably not be an "optimal" one, in respect to its correlation with the watermark. Therefore, a small modification is made on each of its samples. This ensures that it is safely detected as a null-bit in case the first one is cropped.

In case the host audio signal is stereo, the same procedure is applied to both channels. This results in different values of  $p$  and  $g$  given by (5) and (6) for the two channels, which is like having two independent audio signals. This duplication of the information is used for safer recovery of the bitstream, as mentioned in the detection procedure.

After these steps have been taken, the watermark can be embedded on the host signal. The main idea, as in [1], is that every sample of the host signal is modified slightly, so that the watermark remains inaudible.

At first, the host signal is again divided in segments of  $N$  samples each. This time the first  $p$  samples are not taken into account, resulting in  $M + 1$  segments  $x_k(i)$  of length  $N$ :

$$x_k(i) = f(p + k \cdot N + i), \quad i = 0 \dots N - 1, k = 0 \dots M \quad (7)$$

As described previously, the watermark starts being embedded from sample  $p$ . Thus, until this sample the output  $f_w(i)$  is left unwatermarked:

$$f_w(i) = f(i), \quad i = 0 \dots p - 1 \quad (8)$$

The watermarked signal is then calculated as:

$$f_w(p + k \cdot N + i) = x_k(i) \oplus s_k(i) \cdot \alpha |x_k(i)| \otimes w(i) \quad (9)$$

$i = 0 \dots N - 1, \quad k = 0 \dots M$

where constant  $a$  controls the watermark's intensity in comparison to the intensity of the host signal. Operators  $\oplus$  and  $\otimes$  denote superposition laws, which can be multiplication, power law etc. Function  $s_k(i)$  has a value of -1, 0 or 1 depending on the bit, which is encoded in period  $k$ . So:

$$s_k(i) = \begin{cases} 1 & , \text{if bit 1 is encoded in period } k \\ -1 & , \text{if bit -1 is encoded in period } k \\ 0 & , \text{if } k = g \end{cases} \quad (10)$$

#### 4. WATERMARK DETECTION

During the detection phase the goal is to identify if a test signal  $y$  is watermarked and, if it is, to extract the bit sequence  $\hat{m}$  that it contains. A procedure similar to the embedding one is followed. The correlation of the test signal with the basic watermark is calculated for every segment  $k \in [0, M]$ . This is then normalized to a range of values between 0 and

1, with respect to the expected correlation peak values in case  $y$  is watermarked, according to the equation:

$$c'_k(i) = \frac{\sum_{j=0}^{N-1} y(k \cdot N + i + j) \cdot w(j)}{\alpha \sum_{j=0}^{N-1} |y(k \cdot N + i + j)|} \quad (11)$$

The normalized correlation  $c'_k(i)$  is used in order to derive the mean absolute correlation  $C'(i)$ , in a similar way as it was calculated during watermark embedding:

$$C'(i) = \frac{\sum_{k=0}^{N+1} |c'_k(i)|}{M+1} \quad (12)$$

The maximum value of  $C'$  is the detection value, which is calculated by the algorithm. If a peak exists in this function, i.e. a watermark is indeed embedded in  $y$ , since the correlation of the watermark with the host signal is statistically zero, then the detection value would be approximately equal to 1. By putting an appropriate threshold in these results, the existence or absence of a watermark is verified.

If the watermark is detected in the test signal, the next step is to extract the bitstream that is encoded in the watermark. The purpose is to find which index bears the maximum value of  $C'$  and then detect the sign of the  $c'_k$  function at that index. Therefore, the index of the maximum value of the  $C'$  function is estimated by:

$$\hat{p} = \arg \max_i (C'(i)) \quad (13)$$

This value  $\hat{p}$  should be the same as index  $p$ , which was calculated while embedding the watermark. The embedded bits are derived judging on the sign of samples  $c'_k(\hat{p})$  for every period  $k \in [0, M]$ . This is the reason why it is so important to embed the basic watermarks in those positions of the host signal, which result in a correlation, which is close to zero. Otherwise, a bit can be detected as the opposite or even as a null-bit. Function  $b(k)$  calculates the values of  $c'_k(i)$  at index  $\hat{p}$ .

$$b(k) = c'_k(\hat{p}), \quad k = 0 \dots M \quad (14)$$

In order to estimate the index of the null-bit, the one which corresponds to the minimum absolute value of  $b(k)$  is selected:

$$\hat{g} = \arg \min_k (|b(k)|) \quad (15)$$

Then, the bits are derived in the correct order by shifting function  $b$  according to the position of the null-bit and detecting the sign of each value:

$$\hat{m}(k) = \text{sgn}(b((\hat{g} + k) \bmod (M + 1))), \quad k = 0 \dots M - 1 \quad (16)$$

In case of two-channel host signals, every channel is dealt with independently. This results in two functions,  $b_1(k)$

and  $b_2(k)$ , which can be combined for a safer retrieval of the final bitstream  $\hat{m}$ .

The circular correlation  $o(k)$  is used in order to align the two functions  $b_1(k)$  and  $b_2(k)$ , since they contain the same bit information, but the one is a circularly shifted version of the other. Therefore,  $o(k)$  contains a maximum, which occurs at the index where functions  $b_1(k)$  and  $b_2(k)$  are aligned. After having been aligned they can be combined, so as to extract the bits, which they contain, and locate the null-bit. A safe way to detect the null-bit is by calculating the product of corresponding indices of functions  $b_1(k)$  and  $b_2(k)$ , whereas, the bits can be extracted by using their sum.

## 5. EXPERIMENTAL RESULTS

Experiments have been carried out in order to find the optimal values of the algorithm's parameters. The length  $N$  of the basic watermark has been assigned the value of  $2^{15}$  samples, as an acceptable compromise between the detection performance and signal length required to encode  $M$  bits. Parameter  $\alpha$  has been assigned a value, which gives an SNR value of approximately 25.5dB. This SNR value results in signals with imperceptible distortions as verified by presenting the watermarked signals and the originals to several people of medium and higher music education. The audio signals, which were used for testing the method were classical and ethnic music excerpts. Figure 1 presents the ROC curves of the detection results for detecting one watermark comprised of different numbers of bits in the watermarked audio signal. The detection results improve as the length of the encoded message increases. Encoding 16 bits or more gives acceptable detection results (an Equal Error Rate of less than  $10^{-16}$ ). The bit error rate is 0.4% in the case of encoding 8 bits, whereas for the rest of the experiments it was calculated to be between 0 and 0.4%.

The technique has been found (as expected) to be robust to cropping. Furthermore, experiments prove the method's robustness to resampling and requantization. The robustness of the method to compression has also been tested and the detection results after MP3 compression for various compression ratios are presented in Figure 2 for encoding 64 bits, while the bit error rate during the bitstream recovery was not affected by the compression. The detection results after MP3 compression of the test signal are improved in comparison to the ones, which were taken without compression. The reason for this is that the lowpass watermark, which is embedded, is not significantly affected by such a lowpass manipulation, whereas some high-frequency bands of the host signal are removed. Therefore, the power of the watermark is increased against the power of the host signal. However, increasing the compression ratio further results in removing part of the watermark as well and thus the detection results deteriorate. The experimental results of

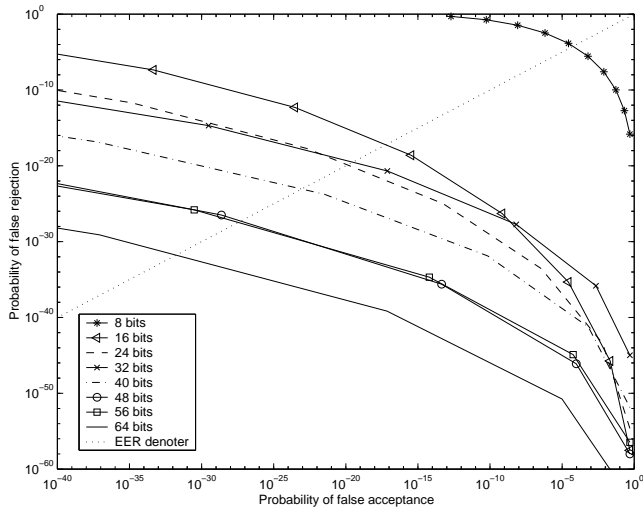


Figure 1: ROC curves for encoding different numbers of bits.

compressing the watermarked signal at 40 kbps (Figure 2) indicate that the performance of the proposed watermarking method is very good although the quality of the audio signal is severely degraded.

## 6. CONCLUSIONS

The proposed method is a novel approach to multibit audio watermarking in the time domain. It is oriented in producing lowpass imperceptible watermarks generated by a chaotic map, which assures that the watermark can not be retrieved by any means without knowing the watermark key and the generator's parameters. The watermark is resistant to attacks, such as compression, resampling, requantization and lowpass filtering. It also resists cropping, since any piece of the audio signal, which still contains a full watermark, is enough for detecting the watermark and extracting the encoded bit sequence.

## 7. REFERENCES

- [1] P. Bassia, I. Pitas and N. Nikolaidis, *Robust audio watermarking in the time domain*, IEEE Transactions on Multimedia, vol.3, no.2, June 2001, pp. 232-241.
- [2] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoan, *Secure spread spectrum watermarking for multimedia*, IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673-1687, 1997
- [3] F. Hartung and M. Kutter, *Multimedia watermarking techniques*, Proceedings of the IEEE, vol.87, no.7, July 1999, pp. 1079-1107.

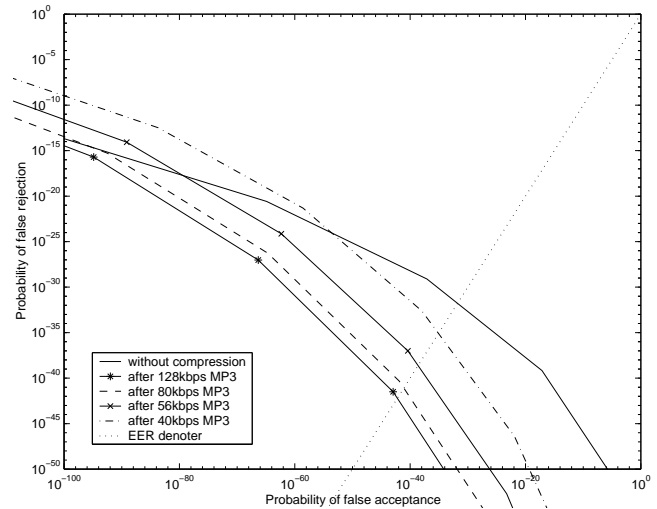


Figure 2: ROC curves for encoding 64 bits, before the attack and after MP3 encoding at different compression ratios.

- [4] W. Kim, J. Lee, and W. Lee, *An audio watermarking scheme robust to MPEG audio compression*, in Proceedings NSIP'99, Antalya, Turkey, June 20-23 1999, vol. 1, pp. 326-330.
- [5] A. Nikolaidis and I. Pitas, *Comparison of different chaotic maps with application to image watermarking*, IEEE Int. Symposium on Circuits and Systems (ISCAS 2000), Geneva, Switzerland, vol. V, pp. 509-512, 28-31 May 2000.
- [6] M. D. Swanson, M. Kobayashi and A. H. Tewfik, *Multimedia data embedding and watermarking technologies*, Proceedings of the IEEE, vol.86, no.6, June 1998, pp. 1064-1087.
- [7] A. Tefas, A. Nikolaidis, N. Nikolaidis, V. Solachidis, S. Tsekeridou, I. Pitas, *Statistical Analysis of Markov Chaotic Sequences for Watermarking applications*, 2001 IEEE International Symposium on Circuits and Systems (ISCAS 2001), Sydney, Australia, 6-9 May, 2001.