ENHANCED TRANSFORM-DOMAIN CORRELATION-BASED AUDIO WATERMARKING

A. Tefas, A. Giannoula, N. Nikolaidis, and I. Pitas

Department of Informatics, Aristotle University of Thessaloniki,
Box 451, Thessaloniki 540 06, GREECE
e-mail: alexia@comm.utoronto.ca,{tefas,nikolaid,pitas}@zeus.csd.auth.gr

ABSTRACT

Various watermarking techniques have been proposed so far, aiming at the copyright protection of audio signals. Little effort has been made, however, in taking under consideration the spectrum of the watermark sequence itself and exploiting its frequency properties. An enhanced audio watermarking technique based on correlation detection, is introduced in this paper, where high-frequency chaotic watermarks are multiplicatively embedded in the low frequencies of the DFT domain. A series of experiments have been conducted to demonstrate both detection reliability and robustness against attacks.

1. INTRODUCTION

Recent trends, such as the growing popularity of MPEG Audio Layer 3 (MP3), which has rendered the distribution of audio files through the Internet easy and fast, have given rise to a generalized effort regarding the protection of the digital content and the intellectual property rights of its owner. This effort has spawned a corresponding digital watermarking imperative, driving the scientific community towards developing more secure and reliable methods of embedding watermarks in digital signals. Watermarking techniques based on correlation detectors have been widely popular in the watermarking community. Theoretical evaluation of their performance with respect to detection reliability has been also attempted. In [1, 2] the statistical properties of the correlation detectors are explored, using pseudorandom watermark signals. Research conducted so far, establishes the dependence of the system detection reliability on the frequency characteristics of the watermark signal.

In this paper, an enhanced transform-domain audio watermarking technique is proposed, where high-frequency watermarks are multiplicatively embedded in the low frequencies of the Discrete Fourier Transform (DFT) domain. A correlation detector is employed in order to detect the embedded signal. The suggested watermarking scheme guarantees robustness against lowpass attacks, together with preservation of the correlation properties and thus, enhancement of the detector reliability. For this purpose, the class of the piecewise-linear Markov chaotic watermarks is proposed, and in particular, the skew tent maps are utilized. Their major advantage is their easily controllable spectral/correlation properties, a fact that makes them a good alternative to the widely used pseudorandom signals [3, 4]. The efficiency of high-frequency watermark embedding in the low-frequency subbands of the DFT domain is demonstrated using a large number of experiments and tests against various attacks. The complete theoretical justification and statistical analysis of the correlation detector can be found in [5].

The outline of this paper is as follows: The proposed audio watermarking model (embedding and detection) is discussed in Section 2. Detailed experimental results are included in Section 3. Conclusions are drawn in Section 4.

The presented work was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 program.

2. WATERMARK EMBEDDING AND DETECTION

Consider x and X, of length N_s, to be the source audio signal and its DFT coefficients, correspondingly. Watermark embedding is performed by modifying the magnitude $F = |X|$ of the DFT coefficients, which can be described by the following formula:

$$W(n) = \begin{cases} W_o(i), & \text{if } aN_s \le n \le bN_s, 0 \le i < N-1 \\ W'_o(i), & \text{if } (1-b)N_s \le n \le (1-a)N_s, \\ & 0 \le i < N-1, \\ 0, & \text{otherwise} \end{cases}$$

where $n = 0, 1, \ldots, N_s - 1$ and coefficients a, b $(0 < a < b \le 0.5)$ control the frequency terms that will be modified. The watermark signal W_o that is used for the construction of W consists of N samples, where $N = \lceil (b-a)N_s \rceil$, and it is generated through an appropriately selected function, $W_o = g(K, N)$, where K denotes the secret key, accessible only to the copyright owner or authorized users. W_o affects a specific low frequency subband of the host signal, around coefficient 0, according to a multiplicative superposition rule. Due to the symmetry of the DFT magnitude, a reflected version of the signal $W'_o(i) = W_o(N-i-1)$ is also embedded in the low frequency components around coefficient $N_s - 1$. Multiplicative embedding is employed for exploiting masking properties of the human auditory system (HAS):

$$F^{'} = F + p\, W\, F \tag{1}$$

where $F^{'}$ is the watermarked audio signal and p is a constant that controls the watermark embedding power.

Correlation detection will be utilized in this paper to examine whether a test audio signal F_t, described by equation (1), contains a watermark W_t or not:

$$c = \frac{1}{N} \sum_{n=0}^{N-1} (F(n)W_t(n) + pW(n)F(n)W_t(n)) \tag{2}$$

In order to reach a decision about the signal being watermarked or not, c is compared against a suitably selected threshold T.

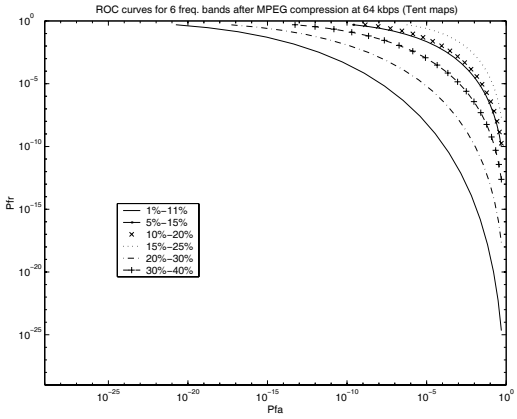

Figure 1: Experimental ROC curves for various frequency bands after MPEG compression at 64 kbps.

The performance of the system can be measured in terms of the probability of false alarm $P_{fa}(T)$ (probability of erroneously detecting the existence of a specific watermark in a signal that is not watermarked or that is watermarked with a different watermark) and the probability of false rejection $P_{fr}(T)$ (probability of erroneously rejecting the existence of a specific watermark in a signal that is indeed watermarked). It was theoretically shown in [5] that the performance of such a correlation-based detector depends on the autocorrelation and cross-correlation functions, or equivalently the power spectrum of the constructed watermarks. Therefore, functions that generate watermarks with desirable spectral/correlation properties, such as certain chaotic signals [6], can be employed to enhance the detection accuracy [3, 7]. For the proposed technique, the class of *skew tent* Markov maps

$$\mathcal{T}(x) = \begin{cases} \frac{1}{\lambda}\, x & , \ 0 \le x \le \lambda \\ \frac{1}{\lambda - 1}\, x + \frac{1}{1-\lambda} & , \ \lambda < x \le 1 \end{cases} \quad \lambda \in (0,1) \quad (3)$$

was selected for the watermark construction. Specifically, if the parameter λ is chosen to be less than 0.5, leading to high-frequency watermarks, the variance of the correlation detector decreases and subsequently, the system performance improves, since it was found that the performance of the proposed watermarking model is uniquely determined by the variance of the correlator. Complete statistical analysis of such a correlation-based watermarking scheme, using chaotic sequences, can be found in [5].

3. EXPERIMENTAL RESULTS

Various experiments were conducted to demonstrate the efficiency of the proposed audio watermarking scheme. For this purpose, a music mono audio signal of approximately 5.94 sec duration, sampled at 44.1 KHz with 16 bits per sample, was utilized. All sets of experiments were performed by employing chaotic watermark signals generated by the skew tent map, using a total number of 10000 keys. The system detection performance was measured in terms of the ROC curves (plots of P_{fa} versus P_{fr}).

In the first set of experiments, the skew tent map parameter λ has been chosen to be equal to 0.3, thus leading to watermarks of highpass spectral characteristics. The just noticeable distortion level for watermark embedding in various frequency bands of the host signal was estimated. Six overlapping frequency subbands of lengths equal to the 10% of the entire host signal were defined. For each of these bands, their symmetric band was also utilized, leading to a total number of watermarked samples equal to the 20% of the host signal. The length of the band was chosen so that it gives a good compromise between robustness and imperceptibility. For each of these frequency bands, the embedding power p was gradually increased and a panel of listeners were asked to verify that no signal distortions could be perceived. The point, where any further increase of the embedding factor p would cause audible artifacts, was considered as the maximum watermark embedding power p that this band could tolerate, i.e. the just noticeable distortion level. The corresponding SNR values (in decibels) and the values of the watermark embedding factor p, for various subbands of the signal spectrum, can be seen in Table 1. It is obvious that watermark embedding in the lowest frequency band, provides great perceptual capacity, enabling to reduce substantially the SNR value without distorting the perceptual quality of the original signal. As the embedding frequency band moves towards the high frequencies, one should increase the SNR level to prevent audible distortions.

Table 1: SNR and the corresponding embedding power p values resulting in inaudible distortions for six frequency bands.

Frequency band	SNR (dB)	Embedding Power p
$a = 1\%, b = 11\%$	23.08	0.27
$a = 5\%, b = 15\%$	41.39	0.15
$a = 10\%, b = 20\%$	41.66	0.18
$a = 15\%, b = 25\%$	45.02	0.15
$a = 20\%, b = 30\%$	45.17	0.25
$a = 30\%, b = 40\%$	47.31	0.80

Although the previous experiment indicates that the lowest frequency band can tolerate distortions of a greater extent, it does not give any clue about the influence of the band selection on the detection performance of the embedded watermark, both with and without attacks. To investigate this influence, the proposed method was tested against lossy MPEG Audio-I layer III compression, for all six frequency subbands defined in the previous experiments. For each embedding frequency band, the audio signal was encoded at a bitrate equal to 64 Kbps, producing a compressed mono audio signal of modest quality, significantly below the standard compression rate of 128 kbps. The corresponding ROC curves are plotted in Fig. 1. Visual inspection demonstrates that embedding in the lowest subband ($a = 0.01$, $b = 0.11$) attains by far the highest robustness to compression. As the frequency band moves towards the high frequencies, the resistance of the technique to MPEG compression declines. This and the previous results lead to selecting the lowest frequency subband as the band of choice for the rest of the experiments.

A large number of experiments were devoted to investigate the robustness of the proposed watermarking scheme to lowpass attacks. In order to compare the performance of the proposed audio watermarking scheme against the performance of alternative techniques, experiments were conducted for two competitive audio embedding schemes. For all methods, watermarks that are just below the audibility threshold have been used, ensuring a fair comparison. A correlation detector (applied in the appropriate domain) was used in all three schemes.

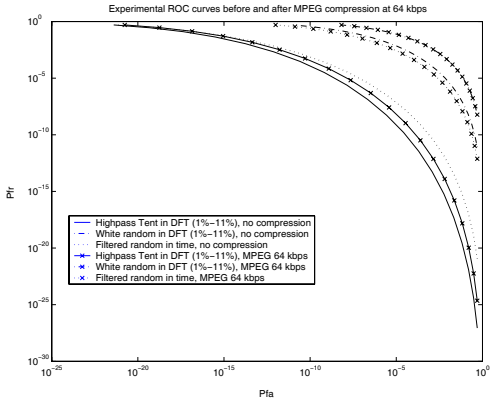

Figure 2: ROC curves for the three watermarking schemes (tent and white watermarks in the DFT domain and prefiltered watermarks in the time-domain) after MPEG compression at 64 kbps.

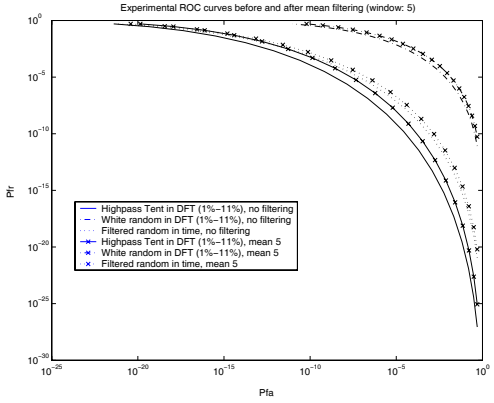

Figure 3: ROC curves for the three watermarking schemes after mean filtering of window size 5.

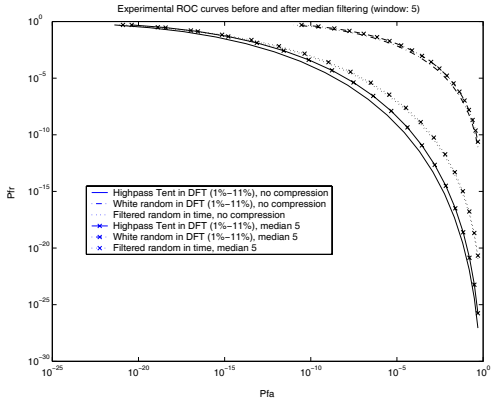

Figure 4: ROC curves for the three watermarking schemes after median filtering of window size 5.

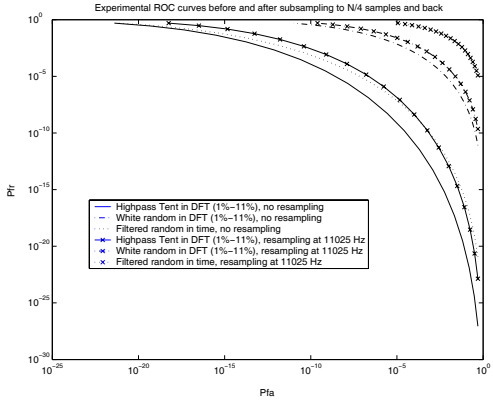

Figure 5: ROC curves for the three watermarking schemes after resampling to $\frac{N}{4}$ samples.

The first alternative embedding scheme involved white pseudorandom watermark sequences ($w(i) \in \{-1, 1\}$) multiplicatively embedded in the same low frequency subband ($a = 0.01$, $b = 0.11$) of the DFT domain, producing watermarked signals with SNR=23 db. The second scheme was based on the time-domain audio watermarking technique presented in [8]: a bipolar white pseudorandom watermark $w(n)$ ($w(n) \in \{-1, 1\}$) was modulated according to the amplitude of the original audio samples $m(n)$ using a multiplicative law:

$$w'(n) = p \mid m(n) \mid w(n) \tag{4}$$

where p denotes the embedding strength. In the next stage $w'(n)$ was shaped using a lowpass Hamming filter of 25th-order with cutoff frequency of 2205 Hz, in order to improve imperceptibility and robustness to lowpass attacks. The resulting filtered watermark signal $w''(n)$ was embedded in the time domain of the original signal $m(n)$:

$$m_w(n) = m(n) + w''(n) \tag{5}$$

thus, producing the watermarked signal $m_w(n)$ (SNR=22 db). Watermarks generated using the previously described procedure will be called hereafter "time-domain pseudorandom watermarks". In all subsequent tests for measuring the detection reliability against various attacks, results will be included for all three watermarking techniques, indicating the superiority of the proposed scheme.

The superior performance of highpass tent watermarks embedded over the low DFT frequencies, against the alternative techniques described above in the case of MPEG-I layer III encoding at 64 kbps is illustrated in Fig. 2. Further experiments involved system performance evaluation for the three techniques under investigation, when the watermarked audio signal is lowpass filtered with mean and median filters of window size 3 and 5. All three techniques prove to be robust against these kinds of attacks, as it can be observed in Figures 3 and 4 for mean and median filtering, respectively (window size 5). The pseudorandom watermarks, used in the two techniques presented for comparison purposes, withstand efficiently these distortions, since in one case, they are embedded in the low frequencies of the host signal and in the other case, they are prefiltered with a lowpass Hamming filter. However, the proposed watermarking scheme still outperforms the other two schemes. Similar results were obtained for filters of length 3.

Next, the watermarked audio signal was subsampled down to 22050 Hz and 11025 Hz and then restored to its initial 44100 Hz sampling frequency, using linear interpolation. The relative performance of the watermarking techniques under consideration,

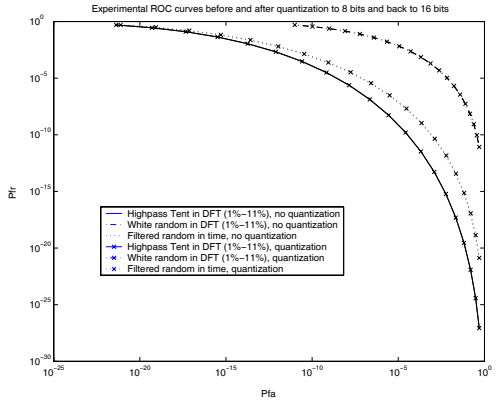

Figure 6: ROC curves for the three watermarking schemes after requantization to 8 bits per sample and backwards.

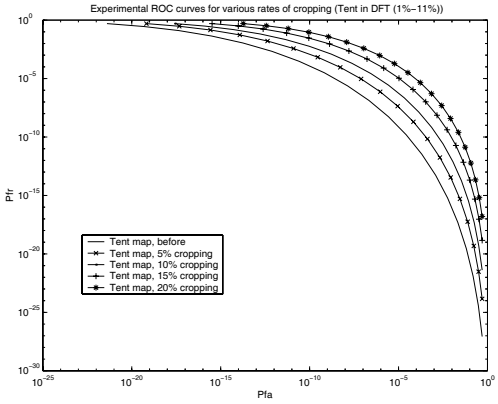

Figure 7: ROC curves after 5%, 10%, 15% and 20% cropping for highpass tent watermarks in DFT (low-band).

against resampling to 11025 Hz and back to 44100 Hz is shown in Fig. 5. Similar results were obtained when subsampling to 22050 Hz was used. By examining the resulting ROC curves, it is obvious that the proposed technique is very robust to resampling. This result can be justified as follows: the extent of deterioration observed in the ROC curves of the alternative techniques is proportional to that of the proposed technique. Since the proposed scheme performs better than the other methods in the undistorted case, its superiority is retained after the attacks. The resulting ROC curves for the other schemes after subsubling are unacceptably poor. Requantization of the original 16-bit audio signal down to 8 bits per sample and backwards, has no effect on the embedded watermark information despite the inevitable loss of data, not only for the proposed scheme but also for the other two schemes, as shown in Fig. 6.

In a last set of experiments, the examined watermarking techniques were tested against cropping. Tests with cropping percentage (percentage of samples removed from the signal) equal to 5%, 10%, 15% and 20% were performed. The corresponding number of samples were removed from the beginning of the audio signal and an equal number of samples were inserted at the end of it, thereby conserving the total number of samples of the signal, since watermark detection is usually applied on signal segments of specific duration (e.g. 1 sec of 44100 samples for a sampling frequency of 44.1 KHz) and not on the whole signal. The time-domain watermark was seriously degraded by cropping of even the smallest percentage, a result that was expected due to synchronization loss caused by such an attack. The other two frequency-based watermarking schemes appeared to cope well with cropping. In fact, the proposed audio watermarking technique proved to be significantly robust even at the highest cropping percentage of 20% (Fig. 7), whereas white random watermarks in the frequency domain deteriorated as this rate increased (the corresponding figures of the two alternative schemes are not shown due to lack of space).

4. CONCLUSIONS

An efficient transform-domain audio watermarking technique based on correlation detection, was presented in this paper. Chaotic watermarks attaining high-frequency spectrum were embedded in the lowest frequency subband of the DFT domain, obeying a multiplicative rule. The statistical properties of the correlation detector were also exploited. The proposed technique guaranteed enhancement of the system detection reliability, inaudibility and great robustness to various attacks. A wide range of experiments were used to establish the superiority of the suggested scheme.

5. REFERENCES

[1] J.R. Hernandez and F. Perez-Gonzalez, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1142–1166, July 1999.

[2] J.-P. Linnartz, T. Kalker, and G. Depovere, "Modeling the false alarm and missed detection rate for electronic watermarks," in *Proc. of 2nd Information Hiding Workshop*, Oregon, USA, April 1998, pp. 329–343.

[3] A. Tefas, A. Nikolaidis, N. Nikolaidis, V. Solachidis, S. Tsekeridou, and I. Pitas, "Performance analysis of correlation-based watermarking schemes employing markov chaotic sequences," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1979 –1994, July 2003.

[4] S. Tsekeridou, V. Solachidis, N. Nikolaidis, A. Nikolaidis, A. Tefas, and I. Pitas, "Statistical analysis of a watermarking system based on bernoulli chaotic sequences," *Elsevier Signal Processing, Sp. Issue on Information Theoretic Issues in Digital Watermarking*, vol. 81, no. 6, pp. 1273–1293, 2001.

[5] A. Giannoula, A. Tefas, N. Nikolaidis, and I. Pitas, "Improving the detection reliability of correlation-based watermarking schemes," in *2003 IEEE International Conference on Multimedia and Expo (ICME 2003)*, Baltimore, USA, 2003, pp. 6–9.

[6] S.H. Isabelle and G.W. Wornell, "Statistical analysis and spectral estimation techniques for one-dimensional chaotic signals," *IEEE Trans. on Signal Processing*, vol. 45, no. 6, pp. 1495–1506, June 1997.

[7] G. Voyatzis and I. Pitas, "Chaotic watermarks for embedding in the spatial digital image domain," in *Proc. of ICIP'98*, Chicago, USA, 4-7 October 1998, vol. II, pp. 432–436.

[8] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 232–241, June 2001.