Robust audio watermarking in the time domain

P.Bassia, I.Pitas, N. Nikolaidis
Dept. of Informatics, University of Thessaloniki
Thessaloniki 540 06, GREECE
Tel: +30 31 996304; fax: +30 31 996304
e-mail: {pitas,nikolaid}@zeus.csd.auth.gr

June 27, 2001

EDICS: 5-SECR

Abstract

The audio watermarking method proposed in this paper offers copyright protection to an audio signal by time domain processing. The strength of audio signal modifications is limited by the necessity to produce an output signal that is perceptually similar to the original one. The watermarking method presented here does not require the use of the original signal for watermark detection. The watermark signal is generated using a key, i.e., a single number known only to the copyright owner. Watermark embedding depends on the audio signal amplitude and frequency in a way that minimizes the audibility of the watermark signal. The embedded watermark is robust to common audio signal manipulations like MPEG audio coding, cropping, time shifting, filtering, resampling and requantization.

This work has been funded by the LTR-ESPRIT European Project 31103 -INSPECT

1 Introduction

The outstanding progress of digital technology has increased the ease with which multimedia data are reproduced and retransmitted. 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 copying and redistribution, known as data piracy.

In the last few years, the problem of copyright protection of multimedia products attracted the interest of scientists and entrepreneurs alike. The most promising solution to this challenging task seems to lie in data hiding techniques, i.e., methods for secretly and imperceptibly embedding signals of certain characteristics in the media of interest. Such signals are usually referred to as watermarks. A watermark completely characterizes the person who applied it and thus marks the host signal as being its intellectual property. Whenever a question about data ownership arises, the product owner/creator is the only one that can detect the watermark and thus prove his ownership on the data. A watermarking system consists of three distinct modules namely the watermark signal generation module, the watermark embedding module and the watermark detection module. The watermark signal is a low amplitude signal that is generated by a non-invertible function that takes as input a set of parameters (watermark key). Sometimes, the host signal is also taken into account in order to generate signal-dependent watermarks. Watermark embedding is performed in the temporal (audio data), spatial (still images), spatiotemporal domain (video) or a suitable transform domain using a properly selected embedding rule (additive, multiplicative etc.). Finally, watermark detection is performed by some sort of correlation detector or statistical hypothesis testing, with or without resorting to the original signal.

A watermarking system should satisfy a set of rather conflicting requirements. First of all, a watermark has to be statistically undetectable by unauthorized persons in order to prevent its removal. This condition is fulfilled if the potential number of keys that produce distinct watermarks is large enough to ensure statistical safety. In addition, the detection scheme should be as statistically reliable as possible. The probability of false watermark rejection or acceptance should be minimal. Furthermore, a watermark has to be robust to signal manipulation and its removal should be impossible without perceptible signal alteration. In other words, a pirate should seriously degrade the signal quality before he accomplishes the watermark removal. The robustness should extend to common signal processing operations, such as filtering, compression, resampling, requantization, cropping, noise corruption, etc. Finally, the watermark should be imperceptible, a requirement which prevents its visual or aural identification and ensures the preservation of the original signal quality. This is usually achieved by implicitly or explicitly taking into account the characteristics of the human visual or auditory system in order to benefit from the so-called perceptual masking, i.e., the decrease in the perceived intensity of a stimulus when this is superimposed over another stimulus.

The algorithmic aspects and requirements described above are common to both image and audio watermarking methods. A generalized watermarking framework that applies to copyright protection watermarking systems for both types of signals is presented in [1, 2]. Despite their similarities, audio and still image watermarking systems exhibit significant differences. First of all, the fact that images are two-dimensional signals provides attackers with more ways of introducing distortions that might affect the watermark integrity. For example, while image watermarking methods have to take into account rotation, removal of rows/columns and nonuniform scaling (scaling with different scaling factors along the two dimensions) audio watermarking methods need not deal with such attacks. Another difference lies in the domain where watermark embedding takes place. Audio watermarking methods usually work on the time or Fourier domain, whereas image watermarking methods often use other embedding domains e.g. the DCT domain which is closely related to the JPEG compression standard or the wavelet transform domain. Finally, due to the different characteristics of human visual and auditory systems, different masking principles should be taken into account in each case. In general, the human auditory system is more sensitive to distortions than the visual system thus making the generation of imperceptible audio watermarks a challenging task.

A review of the multitude of existing multimedia watermarking techniques can be found in [3, 4]. The obvious conclusion from reading these review papers is that research on audio watermarking is not as mature compared to image watermarking. Indeed, only few publications on audio watermarking exist, most of them in the patent literature. However, with the invention of efficient audio compression algorithms, which enabled the distribution of digital music through the internet, the introduction of reliable copyright protection audio watermarking techniques has become a matter of great commercial importance.

A method that performs watermark embedding in the frequency domain in order to take advantage of the frequency masking characteristics of the human auditory system is presented in [5, 6]. Temporal masking is also taken into account. A similar frequency domain embedding technique is described in [7]. In [8] a spread spectrum technique that considers the original signal as the communication channel and the watermark as the transmitted signal is proposed. The authors claim that the watermark must be placed in perceptually significant regions of a signal in order to be robust to attacks. All previously mentioned methods exploit the frequency characteristics of the audio signal in order to embed the watermark, thus minimizing audible distortions even for high amplitude watermarks. However, the original signal is required in order to detect the watermark. In [9] two techniques for watermark embedding in the MPEG audio stream are proposed. The methods work by modifying the scale factors or the encoded samples in each MPEG audio frame. Watermark detection requires the original MPEG audio file. Furthermore, the robustness of the methods to attacks is not discussed. In [10] the authors embed multi-bit information in audio signals in the form of closely spaced echoes. Zeroes and ones are encoded by using different delays between the original signal and the echo. By controlling the echo parameters inaudible distortions are introduced. The method is fairly robust to MPEG audio compression and D/A conversion but has poor results in signals with silence intervals (e.g. speech signals). In [11] a phase coding method is used to embed an inaudible signal. The audio signal is separated into segments and the phase of the first segment is substituted by a reference phase while the phase of subsequent segments is adjusted to preserve their relative phase. Finally, a review on the current audio watermarking literature which includes short descriptions of commercially available or patented audio watermarking techniques is presented in [12].

In the audio watermarking approach proposed in this paper, the set of the original digital audio samples is divided in embedding segments in order to further increase the robustness of the detection scheme to various signal processing operations. The initial watermark signal is perceptually shaped prior to its embedding in the spatial domain of the original audio. This shaping modifies the amplitude of each watermarking sample according to the amplitudes of the corresponding neighboring samples in the original audio. A weight function with low-pass characteristics is used for this purpose. Unlike other methods, e.g. [6], the proposed watermarking method does not require the use of the original signal for watermark detection. Thus, the copyright owner does not have to perform data search in his archives prior to detection. Such searches are very time consuming and render watermarking useless for audio monitoring in digital broadcasting. Another novelty of the method is that a chaotic map is used to generate the watermark sequence, thus obtaining a cryptographicaly secure signal.

The fact that the watermark embedding procedure employs a variety of parameters (e.g. weighting function parameters, embedding rule etc) makes it extremely difficult for an attacker to construct a counterfeit watermark signal like the one described in [13]. Furthermore, even if such a case could occur, the probability of constructing a valid counterfeit watermark can be totally eliminated if we make the assumption that a watermark is valid only if it is associated with a valid key [1]. If such an assumption is adopted then, even if an attacker could manage to construct a counterfeit watermark, he would not be able to invert the watermark generating function to obtain the watermark key. Our watermarking scheme is able to withstand common signal processing operations such as cropping, time shifting, MPEG audio encoding, filtering, resampling and requantization. Ongoing research focuses on ways to improve the robustness of the algorithm to other forms of unauthorized manipulation.

The outline of the paper is as follows. Section 2, introduces the fundamentals of audio masking. The proposed watermark embedding scheme is analyzed in section 3. Although the watermark is embedded in the spatial domain, its construction is also grounded on audio frequency analysis to preserve its inaudibility through masking. In section 4, the detection process is thoroughly analyzed. Subjective tests that prove the watermark imperecptibility are presented in section 5. The potential of embedding multiple watermarks is also examined in this section. In section 6, certain experiments are described that prove the resistance of our watermarking scheme towards various signal processing operations. Finally, a brief conclusion is available in section 7.

2 Audio Masking

The basic principles of audio masking will be presented in this section. Readers that are interested in obtaining a better insight on this subject can consult [14, 15, 16, 17]. The human auditory system can be modeled as a frequency analyzer consisting of a set of 25 bandpass filters (called critical bands [18]) that cover the range 10Hz-20KHz. The bandwidth of the filters increases with the central filter frequency. In the absence of other sounds, the perceived intensity of a single sound, which is called loudness, depends on this sound's pressure level (SPL), frequency and duration. However, when a sound lies in the frequency or temporal neighborhood of another sound, its loudness and hearing threshold is affected by the characteristics of the second sound, a phenomenon known as audio masking. The sound that does the masking is called masker and the sound that is masked is called the maskee. Masking is usually distinguished in frequency masking and temporal masking. *Frequency Masking* is also called simultaneous masking, because it occurs when both sounds are heard at the same time. The auditory result depends on the frequency relation of the two sounds (the closer the two sounds are in frequency the more profound is the masking result) and on the frequency, sound pressure level and tone-like or noise-like characteristics of both signals.

Temporal Masking refers to sounds that are heard at different time instances. Temporal masking can be either forward (pre-masking) or backward (post-masking). When the masker affects a previous sound, the masking is called forward whereas, when the masker affects a subsequent sound the masking is called backward masking. In general, forward masking is not as intense as backward masking. Pre-masking occurs for a duration of 5-20 msec before the masker is turned on. Post-masking occurs for a duration of 50-200 msec after the masker is turned off.

3 Watermark Embedding

The watermark embedding scheme proposed in this paper modifies digital audio signals in the time domain. The result is a slight amplitude modification of each audio sample in a way that does not produce any perceived effect. Let us assume an audio signal of M samples $x(i)$, $i = 0, ..., M-1$. The audio signal is divided in N_S segments of N samples each. Let us denote by $x_k(i)$ the data samples belonging to the k-th segment:

$$x_k(i) = x(k \cdot N + i) \quad i = 0 \dots N-1, \qquad k = 0 \dots N_S - 1 \tag{1}$$

Each of these segments is watermarked with the same bipolar sequence $w(i) \in \{-1, 1\}$, $i = 0, ..., N-1$ which is generated by thresholding a chaotic map in a way similar to the one described in [19]. The chaotic map that has been used in our implementation is the Renyi map [20]. The seed (starting point) of the chaotic sequence generator is the watermark key. By using generators of a strongly chaotic nature we can ensure that the system is cryptographicaly secure i.e. the sequence generation mechanism cannot be inverse engineered even if an attacker can manage to obtain a part of the binary sequence. Furthermore the frequency content of the sequence generated by this particular chaotic map can be controlled by varying a certain parameter of the map. The watermark signal is embedded in each audio segment using the following three-stage procedure.

In the first stage, the watermark signal $w(i)$ is modulated by the audio signal to produce the signal-dependent watermark $w'_k(i)$:

$$w'_k(i) = \alpha|x_k(i)| \oplus w(i) \quad i = 0 \ldots N-1, \qquad k = 0 \ldots N_S - 1 \tag{2}$$

In the previous equation $\oplus$ denotes a superposition law which can be multiplication, power law, etc, and α is a constant which controls the amplitude of the watermark signal. Watermark robustness generally increases with the watermark signal amplitude. The maximum allowable watermark amplitude is limited by the allowable perceived signal distortion.

In the next stage, $w'_k(i)$ is shaped using a lowpass Hamming filter of length (order) L:

$$w''_k(i) = \sum_{l=0}^{L-1} b_l w'_k(i-l) \tag{3}$$

where b_l are the filter coefficients. Despite its simplicity, this shaping process results in inaudible watermarks as will be subsequently described. Finally, the signal-dependent, lowpass shaped watermark signal $w''_k(i)$ is embedded in the original signal segment x_k to produce the watermarked signal segment y_k:

$$y_k(i) = x_k(i) + w''_k(i) \tag{4}$$

It is obvious that the calculation of the watermarked sample $y(i)$ is based both on the neighborhood of the sample $x(i)$, taking simple audio masking principles into account, and on the chaotic signal $w(i)$. The fact that the watermark signal $w(i)$ is embedded in each segment of the audio signal allows watermark detection on cropped or time shifted audio signals as will be explained in the next Section.

Figure 1 shows the power spectral density (*psd*) of two watermark signals and one original audio signal using the Welch averaged periodogram method. Both watermark signals are chaotic ones. The psd of the audible watermark is spread over the frequency range, exceeding the power of the original audio signal in certain frequencies. This explains why the pure chaotic signal is audible. The inaudible watermark signal has undergone through the shaping process described above. The psd of the inaudible watermark lies underneath the psd of the original audio signal in the entire frequency range. Similar tests have been carried out on several music scores.

4 Watermark Detection

In order to proceed to watermark detection we partition the signal y into N_S segments of length N. Let us denote by $S_k(n)$, $n = 0, ..., N-1$, the following sum:

$$S_k(n) = \frac{1}{N} \sum_{i=0}^{N-1} y_k((i+n) \bmod N) w(i) \tag{5}$$

where $k = 0, \ldots, N_S - 1$ is the segment number and $y_k(i)$ are the data samples belonging to the k-th segment. $S_k(n)$ is essentially the correlation of w with y_k, evaluated for all possible circular shifts of y_k.

By combining (4), (5) we get:

$$S_k(n) = \frac{1}{N}\left(\sum_{i=0}^{N-1} x_k((i+n) \bmod N)w(i) + \sum_{i=0}^{N-1} w_k''((i+n) \bmod N)w(i)\right) \tag{6}$$

The expected value of the first sum in (6) is zero if either the watermark mean value m_w or the signal mean value m_x is equal to zero. In case m_w is not zero, i.e., if the total number of -1's and 1's is not the same, the quantity $\Delta w = \sum_{i=0}^{N-1} w(i)$, must be taken into account. Let us denote by B a set of $N_B = |\Delta w|$ index values for which the corresponding $w(i)$ values are equal to the symbol (-1 or 1) with the most occurrences. It is easy to show that:

$$\sum_{i \in B} w(i) = \Delta w$$

Let us also denote by A the set of all index values that do not belong to B. Obviously, the cardinality of A is $N_A = N - |\Delta w|$ and the following equation holds:

$$\sum_{i \in A} w(i) = 0$$

Using the above notation $S_k(n)$ can be expressed as follows:

$$\begin{aligned} S_k(n) = \tfrac{1}{N}(&\textstyle\sum_{i \in A} x_k((i+n) \bmod N)w(i) + \sum_{i \in B} x_k((i+n) \bmod N)w(i) + \\ &\textstyle\sum_{i=0}^{N-1} w_k''((i+n) \bmod N)w(i)) \end{aligned} \tag{7}$$

Let us define the following terms:

$$T_{1,k}(n) = \frac{1}{N}\sum_{i \in A} x_k((i+n) \bmod N)w(i)$$

$$T_{2,k}(n) = \frac{1}{N}\sum_{i \in B} x_k((i+n) \bmod N)w(i)$$

$$T_{3,k}(n) = \frac{1}{N}\sum_{i=0}^{N-1} w_k''((i+n) \bmod N)w(i)$$

It can be easily shown that $E(T_{1,k}(n)) = 0$ where $E()$ denotes the expected value operator. Furthermore, for N_A sufficiently large, we can assume that $T_{1,k}(n) \simeq 0$. For the term $T_{2,k}(n)$ it is easy to show that :

$$\begin{aligned} T_{2,k}(n) &= sign(\Delta w)\tfrac{1}{N}\textstyle\sum_{i \in B} x_k((i+n) \bmod N) = \\ &\tfrac{\Delta w}{N}\tfrac{1}{N_B}\textstyle\sum_{i \in B} x_k((i+n) \bmod N) \end{aligned} \tag{8}$$

Therefore

$$E(T_{2,k}(n)) = \frac{\Delta w}{N} m_X \tag{9}$$

For N_B sufficiently large $T_{2,k}(n) \simeq \frac{\Delta w}{N} m_X$. If no watermark has been embedded in the signal $T_{3,k}(n) = 0$ and thus:

$$S_k(n) \simeq T_{2,k}(n) \simeq \frac{\Delta w}{N} m_X \tag{10}$$

On the other hand, if the signal is watermarked

$$S_k(n) \simeq T_{2,k}(n) + T_{3,k}(n) \simeq \frac{\Delta w}{N} m_X + \frac{1}{N}\sum_{i=0}^{N-1} w_k''((i+n) \bmod N)w(i) \tag{11}$$

For watermark detection we construct the ratio $r_k(n)$

$$r_k(n) \triangleq \frac{S_k(n) - T_{2,k}(n)}{T_{3,k}(n)} \tag{12}$$

The original signal $x(i)$ required for the evaluation of $T_{2,k}(n)$, $T_{3,k}(n)$ is not available in the detection process, but can be replaced by $y(i)$ without significant error.

The value of $r_k(n)$ is calculated for every $n = 0...N-1$, for all signal segments $k = 0,\ldots N_S - 1$. Thus it is easy to prove that the computational complexity of the detection stage increases as a linear function of the segment length N. If y is a cropped version of the watermarked signal, the number of segments N_s will be different from the segments in the embedding procedure. Furthermore, in this case the segment starting/ending points in the detection phase will not coincide with those of the segments defined during the embedding procedure. However, the fact that $S_k(n)$ is calculated for all possible circular shifts $n = 0,\ldots,N-1$ of y_k ensures that a 'synchronization' between w and y_k will be achieved for a certain value n_w of n, provided of course that y is watermarked with the correct watermark. In this case, $r_k(n_w)$ would be approximately equal to one. n_w will be zero for an audio signal that has not been cropped or time-shifted and nonzero in all other cases. If the signal is not watermarked, watermarked with a different watermark or if $n \neq n_w$, $r_k(n)$ would be approximately zero. To obtain detection, the maximum value $r_k(n)$ with respect to n, i.e., the value $R_k = \max_{n=0}^{N-1} r_k(n)$ is evaluated. Obviously R_k will occur for the value $n = n_w$ where synchronization is achieved.

Since the watermark signal has been embedded in all audio segments, the detection ratio $R = max_{k=0}^{N_S-1} R_k$ is evaluated. The decision whether a certain watermark exists in the signal can be taken by comparing R against a suitable threshold. For a certain threshold, the detection performance is described in terms of the false positive probability, (i.e. the probability to detect a watermark in a signal that is not watermarked or is watermarked with a different watermark) and the probability of correct detection (i.e., the probability to detect the watermark embedded in the signal). In the ideal case, the values of these probabilities would be 0 and 100% respectively.

From the previous analysis of the detection process it is obvious that our method is inherently robust to time-shifting and cropping. The minimum number of contiguous audio samples that suffice for a successful watermark detection is equal to the segment length N. Thus, a watermarking system using small segments can withstand cropping of a bigger extend. In our implementation a value of $N = 2^{17}$ was used and, thus, for a mono audio signal with a sampling rate of $44100Hz$, less than 3 seconds suffice for watermark detection. Far smaller segments can also be used. In the same manner, a watermark which utilizes a small N is more robust to random insertion/removal of data samples, since the probability that at least one segment remains intact after such an attack increases as the segment length decreases.

Figure 2 illustrates the empirical pdf of the detection ratio R in a watermarked and a non-watermarked excerpt of Beethoven's Ninth Symphony. In this figure, the empirical pdf of R for a non-watermarked signal is represented by a solid line and has been obtained by evaluating the response of the detector on a non-watermarked signal using 1000 distinct watermark keys. The dashed line represents the empirical pdf of R for a watermarked signal, derived by evaluating the detector response on 1000 watermarked versions of the signal. The fact that the mean value of R for the non-watermarked signal obtains a value of 0.3 (i.e. is not zero) is due to the use of the maximum operator in the evaluation of R which results in taking the worst case value for $r_k(n)$.

The detection algorithm gives no false positives when tested on a watermarked signal using invalid watermark keys. For example, 1000 keys (of which only one was valid) have been used to detect a watermark in a watermarked audio signal. The results are shown in Figure 3.

4.1 Statistical Analysis

In order to statistically approximate our watermark detection scheme, we consider a single signal segment and use the expression:

$$S = \frac{1}{N} \sum_{i=0}^{N-1} y(i)w(i) \tag{13}$$

as our theoretical watermark detection function. In addition, we simplify things by ignoring the perceptual masking stage and supposing that our watermark embedding scheme is either additive:

$$y(i) = x(i) + \alpha w(i) \tag{14}$$

or multiplicative:

$$y(i) = x(i) + \alpha |x(i)| w(i) \tag{15}$$

In the following, we derive the expected value and variance of S for both embedding schemes in the presence or absence of a watermark and therefore show the validity of our watermark detector (12).

Let us consider both the original signal and the watermark signal as sequences of identically distributed random variables. Obviously, the bipolar watermarking signal has zero mean $E(w(i)) = m_w = 0$ and variance $\sigma_w^2 = 1$. Generally, $x(i)$ is supposed to have zero expected value, $E(x(i)) = m_x = 0$. If this is not the case, the signal can be shifted to have zero expected value. However, the results presented below are valid even for $m_x \neq 0$.

It is easy to prove that for an additive watermark:

$$m_S = \begin{cases} \alpha, & \text{watermark present} \\ 0, & \text{no watermark present} \end{cases} \tag{16}$$

while for a multiplicative watermark :

$$m_S = \begin{cases} \alpha m_{|x|}, & \text{watermark present} \\ 0, & \text{no watermark present} \end{cases} \tag{17}$$

The variance of S is calculated for an additive watermark as:

$$\sigma_S^2 = \frac{\sigma_x^2}{N} \tag{18}$$

for both cases of watermark existence and absence, and for a multiplicative one as:

$$\sigma_S^2 = \begin{cases} \frac{\sigma_x^2}{N} + \frac{\alpha^2 \sigma_{|x|}^2}{N}, & \text{watermark present} \\ \frac{\sigma_x^2}{N}, & \text{no watermark present} \end{cases} \tag{19}$$

The above results have been derived using the formulas $E(w^3) = E(w) = 0$ and $E(w^4) = E(w^2) = 1$. An examination of the results reveals that the variance of S is inversely proportional to the number of samples N (segment length) and thus the false positive probability decrease with N, whereas the probability of correct detection increases. Furthermore, in the multiplicative case, the variance of S is bigger for watermarked signals than for non-watermarked signals, a fact which confirms the experimental results presented in Figure 2.

5 Subjective Quality Evaluation

The inaudibility of our watermarking schemes has been certified through listening tests that involved ten persons. Half of them had average musical education. Two classical music themes (Vivaldi's L'Amoroso Concert and Beethoven's Ninth Symphony) plus a pop and an ethnic music theme, of 5 seconds duration each, were used as the test signals. The listeners were presented with the original and the watermarked audio excerpts and were asked to report any differences between the two signals. None of the subjects was able to tell the difference between the two signals, which is equivalent to a Mean Opinion Score (MOS) of 5 in a 5-point impairment scale (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying 1: very annoying). In another test, the listeners were successively presented with the original and the watermarked signal with no indication on the identity of each excerpt and were asked to pinpoint the watermarked one. The percentage of erroneous answers was very close to 50% showing that the two signals cannot be discriminated. This is attributed to the noise masking that we have performed.

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 watermark signal. According to listening tests, up to 4 watermarks of $SNR = 22dB$ each, cause almost no perceptible distortion to the watermarked signal, while more than 8 watermarks, of the same SNR each, result to a perceivable noise in the watermarked signal. The MOS values for 1, 2, 4, 6 and 8 watermarks are presented in Table 1. However, multiple watermarks are being detected in a signal with equal success. In conclusion, a multiple watermarking scheme is possible as far as the power of the combined watermarks is chosen so that the sum of the embedded watermarks is not audible.

6 Audio Watermark Robustness to Signal Manipulation

In order to test the robustness of our method to common signal processing operations, a set of experiments involving several audio excerpts (sampling rate $44100Hz$) were conducted. The empirical pdf (histogram) of the detection ratio R on a non-watermarked signal was obtained by calculating the detector output for 1000 different watermark keys on a signal bearing no watermark. The corresponding pdf for a distorted watermarked signal was obtained by calculating the detector output on 1000 distorted watermarked versions of a signal, each generated with a different watermark key. In all experiments the watermark amplitude parameter a was chosen so that the signal to noise ratio for the watermarked signal is $22dB$. A threshold value of 0.5 was used in all experiments. Using this threshold, no false positive detections were observed. A small subset of the obtained results, concerning two classical music themes, are presented in the following subsections. Similar results were obtained for all the other audio excerpts.

6.1 Robustness to MPEG2 audio compression

The robustness of the proposed watermarking technique to compression has been tested using Layers II and III of the MPEG2 audio compression standard. The watermark resists the encoding-decoding process as shown in Figures 4a,b (Beethoven melody). Figure 4a depicts the pdf of the detection ratio after MPEG2 Layer II compression at 48Kbps (solid curve), in comparison to the pdf of the same ratio before MPEG2 (dashed curve). Results indicate a decrease in the mean value of the detection ratio. However, the detection ratio R is always above the threshold, and thus the watermark is detected with 100% success. When MPEG2 Layer III compression at $80kbps$ is applied, the watermark detection is also being performed with a 100% success, as shown in Figure 4b. The corresponding results for the Vivaldi melody are shown in Figures 5a,b. For this signal, the probability of detection for the MPEG2, Layer III compression was 99.9%

Figure 6a presents the detection ratio for the non-watermarked (left distribution) and the watermarked (right distribution) Vivaldi audio excerpt after MPEG Layer II compression at $48kbps$. The corresponding histograms for MPEG Layer III compression at $80kbps$ are depicted in Figures 6b.

For bitrates lower than $48kbps$ (MPEG2 Layer II) and $80kbps$ (MPEG2 Layer III) the probability of correct detection deteriorates. The fact that the method is more robust to Layer II compression than to Layer III compression can be attributed to the advanced compression techniques used in Layer III which obviously affect the watermark more than the simpler techniques utilized in Layer II.

6.2 Robustness to Filtering

The robustness of the watermark procedure described in this paper to moving average and low-pass filtering was also investigated. Figure 7a depicts the pdf of the detection ratio R for the Vivaldi audio excerpt filtered by a moving average filter of length 20 (solid curve), along with the pdf of the same ratio before filtering (dashed curve). Moving average filtering introduced a noticeable audible distortion, yet the watermark was detected with 100% success.

The watermark proved to be robust to low-pass filtering, more specifically filtering with a 25th order Hamming low-pass filter with cut-off frequency $2205Hz$. The solid curve of Figure 7b displays the pdf of the detection ratio after filtering, while the dashed curve indicates the respective pdf before filtering. It is obvious that the mean value of R decreases after the filtering. However 100% success in watermark detection was obtained. Figures 8a,b display the detection ratio for the non-watermarked (left distribution) and the watermarked (right distribution) Vivaldi audio excerpt after moving average and lowpass filtering, respectively.

6.3 Robustness to Resampling, Requantization and Cropping

Watermarked audio signals with original sampling rate $44,100Hz$ have been subsampled down to $22,050Hz$ and $11,025Hz$ and interpolated back to their initial frequency. Although the above procedure caused noticeable distortion in relation to the original signals, the watermarks remained detectable. Figure 9a shows the empirical pdf of R for the watermarked Beethoven excerpt that has been resampled down to $11,025Hz$ and back to its initial $44,100Hz$ frequency, along with the corresponding pdf for the undistorted signal. In this experiment we had 100% success in watermark detection.

Requantization of the original 16-bit audio signal down to 8 bits per sample and backwards conserves the embedded watermark, despite the resulting loss of information. The watermarks resist the requantization process because they are amplitude dependent with respect to the original signal. Figure 9b shows the empirical pdf for the Beethoven watermarked signal that have been requantized down to 8 bits and back to 16 bits/sample. The watermark was detected with 100% success. The corresponding results for the Vivaldi signal are displayed in Figures 10a (resampling), and 10b (requantization). The probability of detection after resampling was 99.9%. Figures 11a,b present the detection ratio for the non-watermarked (left distribution) and the watermarked (right distribution) Vivaldi audio excerpt after after resampling and requantization, respectively.

Robustness to cropping has also been tested. Figure 12 depicts the empirical pdf for the Vivaldi watermarked signal where a segment of duration 1 second has been removed. The watermark was detected with 100%. probability. The results verified that the method is inherently robust to this kind of attack, as explained in Section 4.

7 Conclusions-Future Directions

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 key.

The detection procedure does not use the original signal. Our watermarking scheme is statistically imperceivable and resists MPEG2 audio compression plus other common forms of signal manipulation, such as cropping, time shifting, filtering, resampling and requantization. However the method is not robust to more sophisticated attacks. For example one would not be able to detect the watermark in a signal that has been subject to a change in the time scale. Ongoing research focuses in robustifying our method against such attacks.

References

[1] G. Voyatzis and I. Pitas, "Protecting digital-image copyrights: A framework", *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 18–24, January/February 1999.

[2] G. Voyatzis and I. Pitas, "The use of watermarks in the protection of digital multimedia products", *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1197–1207, July 1999.

[3] F. Hartung and M. Kutter, "Multimedia watermarking techniques", *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, July 1999.

[4] M.D. Swanson, M. Kobayashi, and A.H. Tewfik, "Multimedia data embedding and watermarking technologies", *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, June 1998.

[5] L. Boney, A.H. Tewfik, and K.N. Hamdy, "Digital watermarks for audio signals", in *Proc. of EUSIPCO'96*, September 1996, vol. III, pp. 1697–1700.

[6] M.D. Swanson, B.Zhu, A.H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking", *Elsevier Signal Processing, Sp. Issue on Copyright Protection and Access control*, vol. 66, no. 3, pp. 337–355, 1998.

[7] 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. I, pp. 326–330.

[8] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia", *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 12 1997.

[9] K. Nahrstedt, "Non-invertible watermarking methods for mpeg video and audio", in *Multimedia and Security Workshop, ACM Multimedia '98*, Bristol, UK, September 1998.

[10] D. Gruhl, A. Lu, and W. Bender, "Echo hiding", in *Proceedings of 1st Information Hiding Workshop*, Cambridge, U.K., May 1996, pp. 295–316.

[11] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", *IBM Systems Journal*, vol. 35, no. 3&4, pp. 313–335, 1996.

[12] M. Swanson, B. Zhu, and A. Tewfik, "Current state of the art, challenges and future directions for audio watermarking", in *Proceedings of 1995 IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, June 7-11 1999, vol. 1, pp. 19–24.

[13] S. Craver, N. Memon, B-L. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques : Limitations, attacks and implications", *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 4, pp. 573–586, 1998.

[14] P. Noll, "Wideband speech and audio coding", *IEEE Communications magazine*, vol. 31, no. 11, pp. 34–44, November 1993.

[15] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception", *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1421, October 1993.

[16] Y.Tohkura S.Nakagawa, K.Shikano, *Speech, hearing and neural network models*, IOS Press, Netherlands, 1995.

[17] B. Moore (editor), *Hearing (Handbook of perception and cognition)*, Academic Press, 1995.

[18] H. Fletcher, "Auditory patterns, review of modern physics, 12", 1940.

[19] G. Voyatzis and I. Pitas, "Chaotic watermarks for embedding in the spatial digital image domain", in *Proceedings ICIP98*, Chicago, USA, 4-7 October 1998, vol. II, pp. 432–436.

[20] R. L. Devaney, *An introduction to dynamical systems*, Penjamine/Cummings, 1986.

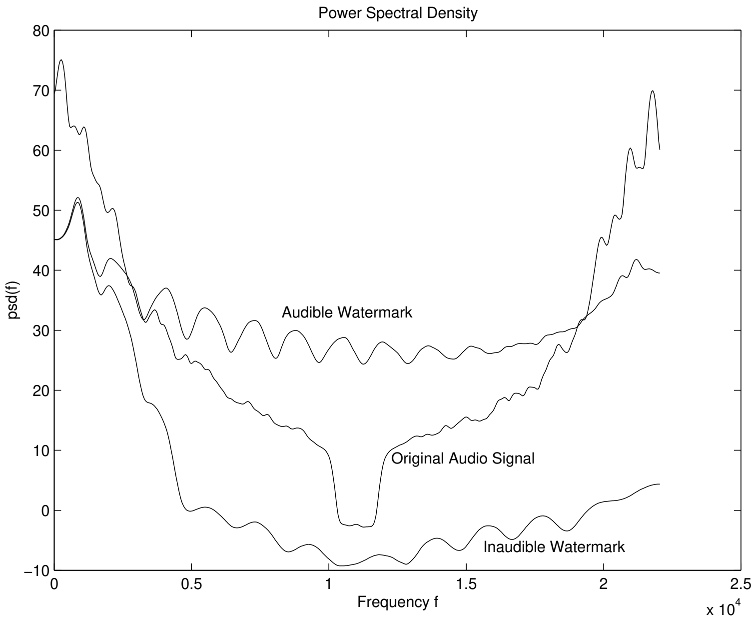


Figure 1: Power spectral density of two watermarks and a music signal (Vivaldi melody).

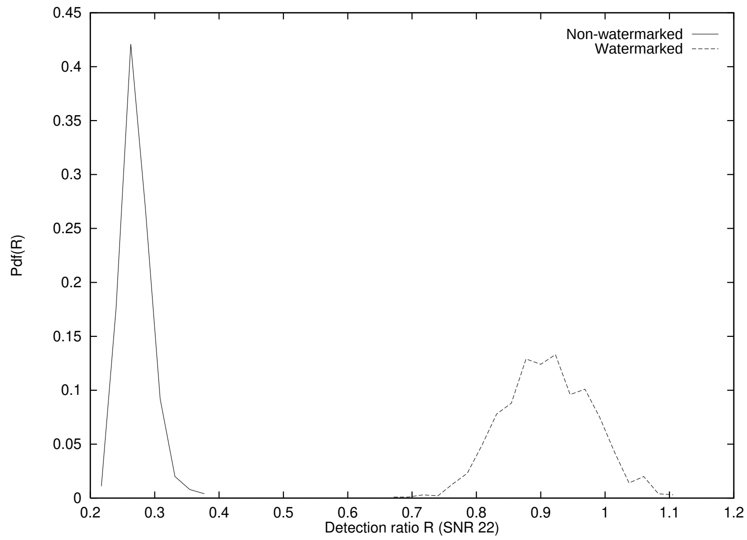


Figure 2: Empirical pdf of the detection ratio in non-watermarked and watermarked signal (Beethoven melody).

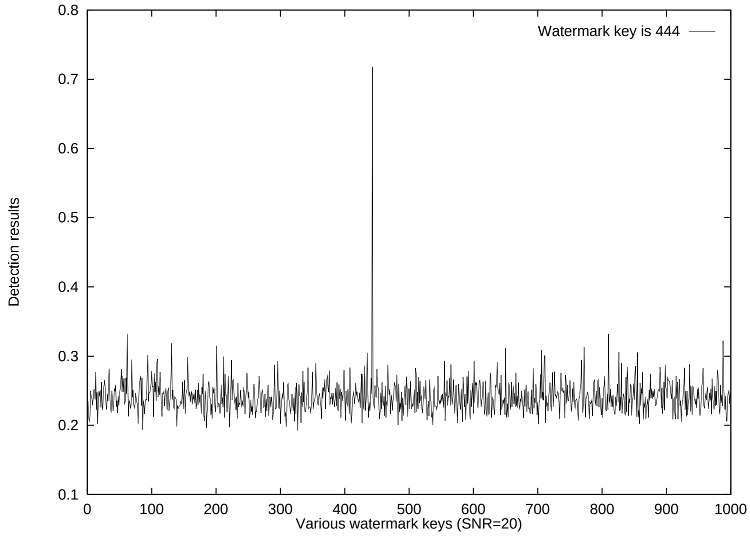
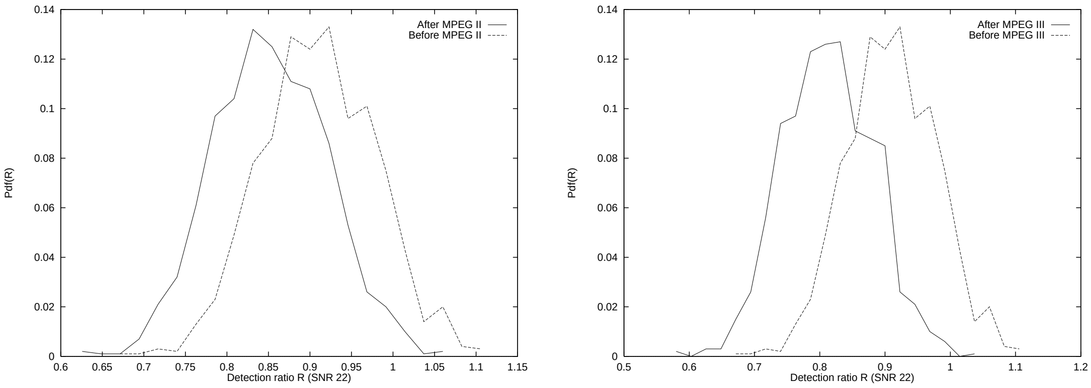
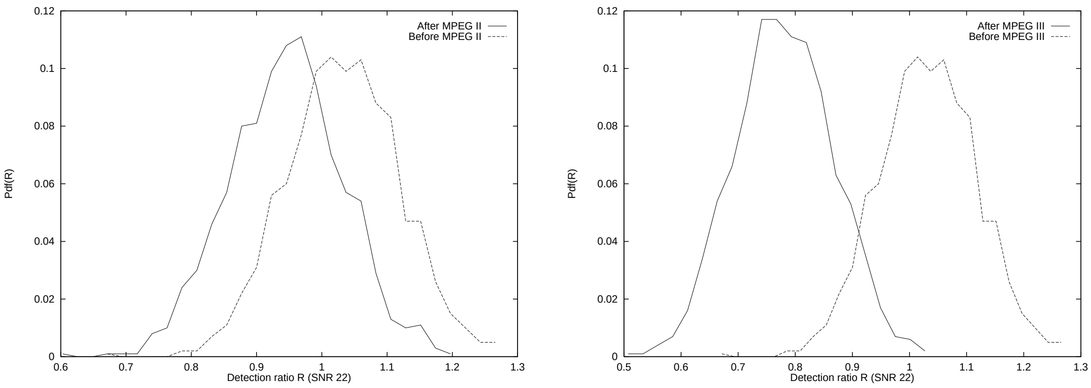


Figure 3: Detection ratios in a watermarked signal using 1000 different keys (valid key is 444).



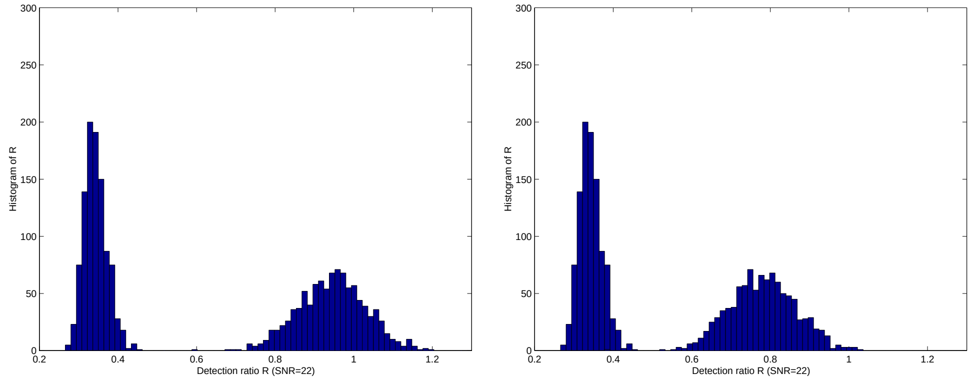
(a) (b)

Figure 4: Empirical pdf of the detection ratio before and after MPEG2 audio compression (Beethoven melody): (a) Layer II 48kbps, (b) Layer III 80 kbps.



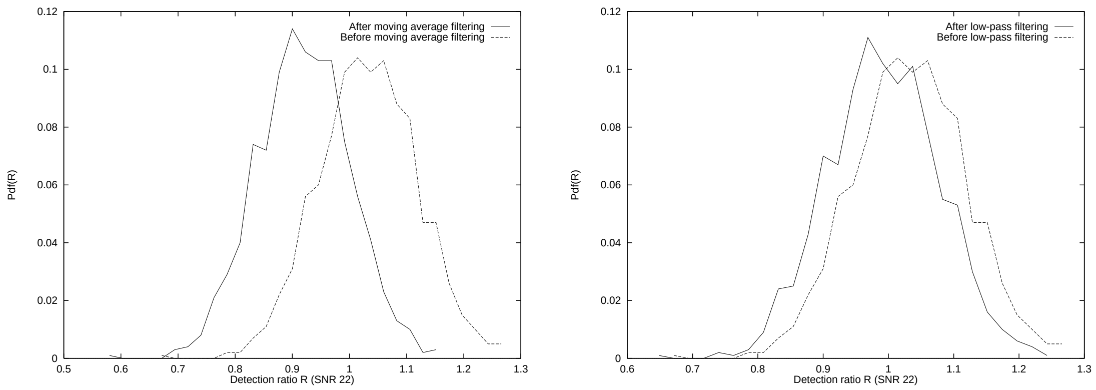
(a) (b)

Figure 5: Empirical pdf of the detection ratio before and after MPEG2 audio compression (Vivaldi melody): (a) Layer II 48kbps, (b) Layer III 80 kbps.



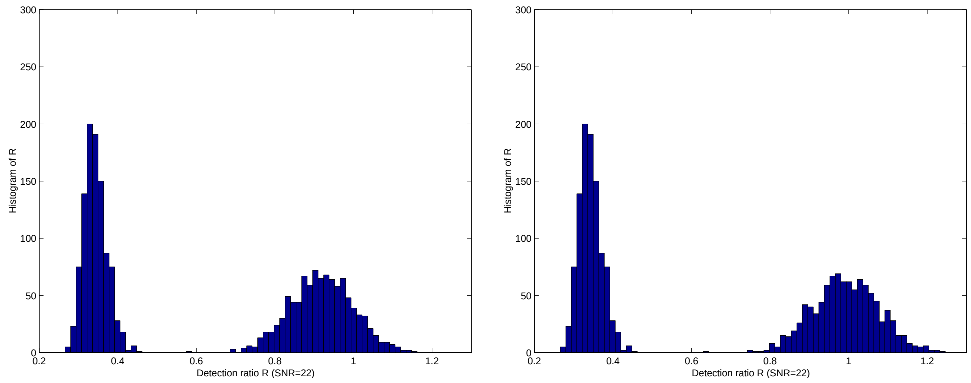
(a) (b)

Figure 6: Histogram of the detection ratio on non-watermarked and watermarked signal (Vivaldi melody) after MPEG2 audio compression: (a) Layer II 48kbps, (b) Layer III 80 kbps.



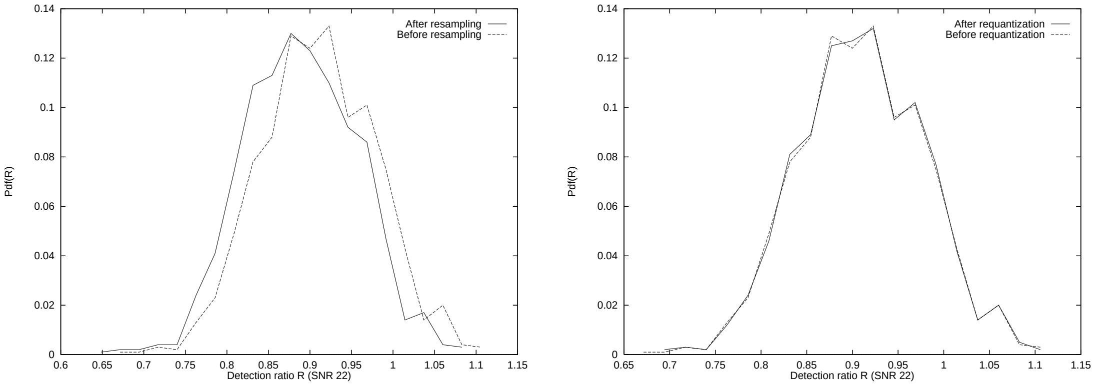
(a) (b)

Figure 7: Empirical pdf of the detection ratio before and after (a) moving average and (b) lowpass filtering of the watermarked Vivaldi melody.



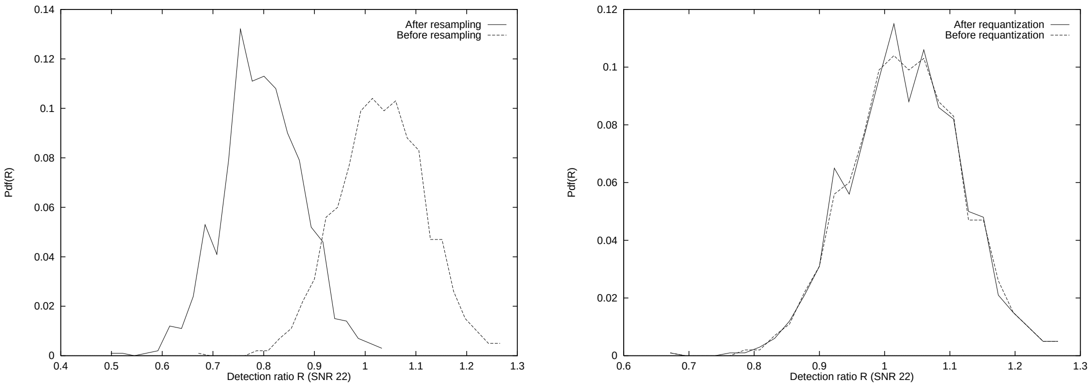
(a) (b)

Figure 8: Histogram of the detection ratio on a non-watermarked and watermarked signal (Vivaldi melody) after (a) moving average and (b) lowpass filtering.



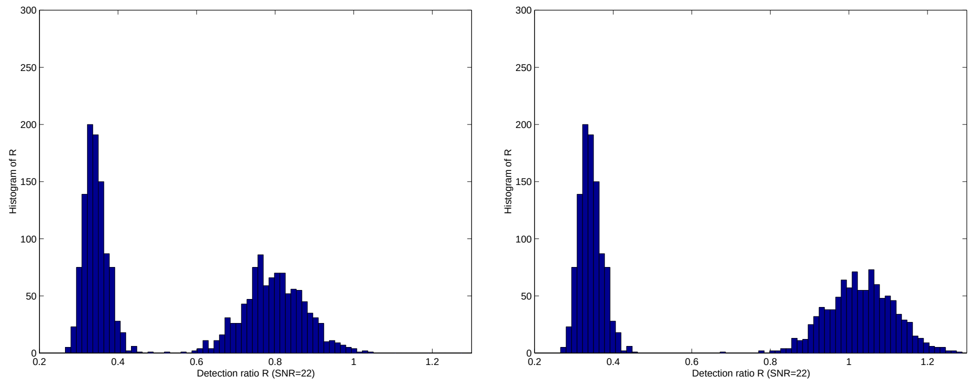
(a) (b)

Figure 9: Empirical pdf of the detection ratio before and after (a) resampling and (b) requantization of the watermarked Beethoven melody.



(a) (b)

Figure 10: Empirical pdf of the detection ratio before and after (a) resampling and (b) requantization of the watermarked Vivaldi melody.



(a) (b)

Figure 11: Histogram of the detection ratio on non-watermarked and watermarked signal (Vivaldi melody) after (a) resampling and (b) requantization.

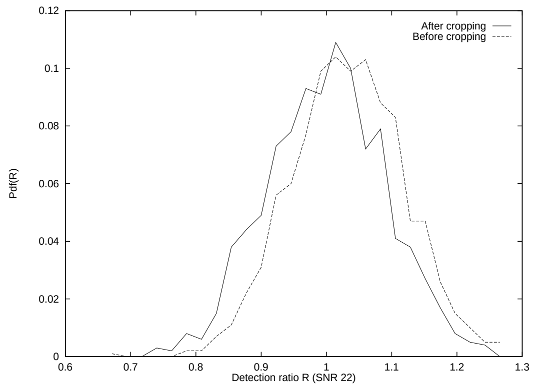


Figure 12: Empirical pdf of the detection ratio before and after cropping of the watermarked Vivaldi melody.

No of watermarks	MOS
1	5
2	4.6
4	4.2
6	3.6
8	3.3

Table 1: Mean Opinion Scores (5-point impairment scale) for a signal watermarked with 1, 2, 4, 6, 8 watermarks