Digital Watermarks for Audio Signals *

Laurence Boney
Departement Signal, ENST, Paris, France 75634
email: boney@email.enst.fr

Ahmed H. Tewfik and Khaled N. Hamdy
Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455
email: tewfik@ee.umn.edu, khamdy@ee.umn.edu

ABSTRACT

In this paper, we present a novel technique for embedding digital "watermarks" into digital audio signals. Watermarking is a technique used to label digital media by hiding copyright or other information into the underlying data. The watermark must be imperceptible and should be robust to attacks and other types of distortion. In addition, the watermark also should be undetectable by all users except the author of the piece. In our method, the watermark is generated by filtering a PN-sequence with a filter that approximates the frequency masking characteristics of the human auditory system (HAS). It is then weighted in the time domain to account for temporal masking. We discuss the detection of the watermark and assess the robustness of our watermarking approach to attacks and various signal manipulations.

1 Introduction

In today's digital world, there is a great wealth of information which can be accessed in various forms: text, images, audio, and video. It is easy to ensure the security of "analog documents" and protect the author from having his work stolen or copied. The question is *how do you copyright or label digital information and preserve its security without destroying or modifying the content of the information.*

Data hiding, or steganography, refers to techniques for embedding watermarks, signatures, and captions in digital data. A watermark could be used to provide proof of "authorship" of a signal. Similarly, a signature is used to provide proof of ownership and track illegal copies of the signal. Watermarking is an application which embeds a small amount of data, but requires the greatest robustness because the watermark is required for copyright protection [1]. One approach to data security is to use encryption [1]; however, once the documents are decrypted, the "signature" is removed and there is *no proof of ownership* such as a label, stamp, or watermark. Note that data hiding does not restrict access to the original information as does cryptography.

The watermark should: **be inaudible** [1, 2]; **be statistically invisible** to prevent unauthorized detection and/or removal by "pirates"; **have similar compression characteristics as the original signal** to survive compression/decompression operations; **be robust to deliberate attacks by "pirates"**; **be robust to standard signal manipulation and processing operations** on the host data, e.g., filtering, resampling, compression, noise, cropping, A/D-D/A conversions, etc; **be embedded directly in the data, not in a header**; **support multiple watermarkings**; **be self-clocking** for ease of detection in the presence of cropping and time-scale change operations.

Observe that a "pirate" can defeat a watermarking scheme in two ways. He may manipulate the audio signal to make the watermark undetectable. Alternatively, he may establish that the watermarking scheme is unreliable, e.g., that it produces too many false alarms by detecting a watermark where none is present. Both goals can be achieved by adding inaudible jamming signals to the audio piece. Therefore, the effectiveness of a watermarking scheme must be measured by its ability to detect a watermark when one is present (probability of detection) *and* the probability that it detects a watermark when none is present (probability of a false alarm) in the presence of jamming signals and signal manipulations.

Several techniques for data hiding in images have been developed [1, 3, 4, 5, 6]. A method similar to ours is proposed in [2], where the N largest frequency components of an image are modified by Gaussian noise. However, the scheme only modifies a subset of the frequency components and does not take into account the human visual system (HVS). The audio watermark we propose here embeds the *maximum* amount of information throughout the spectrum while still remaining perceptually inaudible. It is well-known that detection performance improves with the energy of the signal to be detected. Therefore, we effectively improve the performance of the watermarking scheme by increasing the energy of the watermarked signal while keeping it inaudible.

In [7, 8], we presented a novel technique for embedding digital watermarks into audio signals. Note that our approach is similar to that of the approach of [1], in that we shape the frequency characteristics of a PN-sequence. However, unlike [1] we use perceptual masking models of the HAS to generate the watermark. In particular, our scheme for audio is the only one that uses the frequency masking models of the HAS along with the temporal masking models to hide the copyright information in the signal. We also provide a study of the detection performance of our watermarking scheme. Our results indicate that our scheme is robust to lossy coding/decoding, D/A - A/D conversion, signal resampling, and filtering. In this paper, we present further results showing that our scheme is robust when the watermarks which are

*This work was partially supported by AFOSR under grant AF/F49620-94-1-0461 and by NSF under grant NSF under grant NSF/INT-9406954.

composed of multiple PN-sequences and is robust in the presence of audible distortions due to vector quantization.

Finally, observe that the approach described here for watermarking audio signals can also be used to watermark image and video data with appropriate modifications and extensions (c.f. [7, 9]).

2 Watermark Design

Each audio signal is watermarked with a unique codeword. Our watermarking scheme is based on a repeated application of a basic watermarking operation on processed versions of the audio signal. The basic method uses three steps to watermark an audio segment as shown in Fig. 1. The complete watermarking scheme is shown in Fig. 2. Below we provide a detailed explanation of the basic watermarking step and the complete watermarking technique.

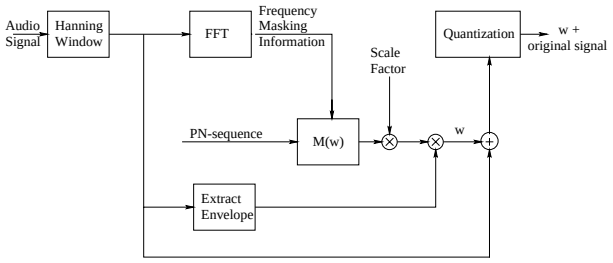

Figure 1: Watermark Generator: First stage for audio

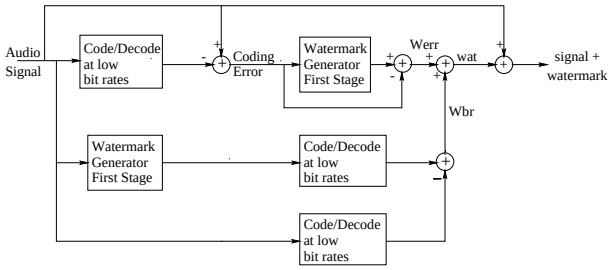

Figure 2: Full Watermark Generator for audio

2.1 The basic watermarking step

The basic watermarking step starts with a PN-sequence. Maximum length PN-sequences are used in our watermarking scheme because they provide an easy way to generate a unique code for an author's identification. Like random binary sequences, PN sequences have 0's and 1's that occur with equal probabilities. The autocorrelation function (ACF) of such a sequence has period N and is binary valued [10]. Because of the periodicity of the ACF, the PN sequence is self-clocking. This allows the author to synchronize with the embedded watermark during the detection process. This is important if the signal is cropped and resampled.

To generate the watermark, we first calculate the masking threshold of the signal using the MPEG Audio Psychoacoustic Model 1, [11]. The masking threshold is determined on consecutive audio segments of 512 samples. Each segment is weighted with a Hanning window. Consecutive blocks overlap by 50%. The masking threshold is then approximated with a 10^{th} order all-pole filter, $M(\omega)$, using a least squares criterion. The PN-sequence, $seq(\omega)$, is filtered with the approximate masking filter, $M(\omega)$, in order to ensure that the spectrum of the watermark is below the masking threshold.

Since the spectral content of the audio signal changes with time, watermarks added to different blocks will be in general different even if they are generated from the same starting PN-sequence. However, it is preferable to use different PN-sequences for different blocks to make the statistical detection by an unauthorized user of the watermark more difficult. Note also that using long PN-sequences or embedding long cryptographic digital signatures also helps in that respect.

Frequency domain shaping is not enough to guarantee that the watermark will be inaudible. Frequency domain masking computations are based on Fourier analysis. A fixed length FFT does not provide good time localization for our application. In particular, a watermark computed using frequency domain masking will spread in time over the entire analysis block. If the signal energy is concentrated in a time interval that is shorter than the analysis block length, the watermark is not masked outside of that subinterval. This then leads to audible distortion, e.g., pre-echoes. To address this problem, we weight the watermark in the time domain with the relative energy of the signal.

The time domain weighting operation attenuates the energy of the computed watermark. In particular, watermarks obtained as above have amplitudes that are typically smaller than the quantization step size. Therefore, the watermark would be lost during the quantization process. Note also that, as observed earlier, detection performance is directly proportional to the energy of the watermark. We have found that it is possible to prevent watermark loss during quantization and improve detection performance by amplifying the watermark by 40 dB before weighting it in the time domain with the relative energy of the signal. We have found experimentally that this amplification does not affect the audibility of the watermark because of the attenuation effect of the time domain weighting operation.

2.2 The full watermarking scheme

As mentioned above, the watermarking scheme must be robust to coding operations. Low bit rate audio coding algorithms tend to retain only the low frequency information in the signal. We, therefore, need to guarantee that most of the energy of the watermark lies in low frequencies. After experimenting with many schemes, we have found that the best way to detect the low frequency watermarking information is to generate a low-frequency watermark as the difference between a low bit rate coded/decoded watermarked signal and the coded/decoded original signal at the same bit rate. Watermarking is done using the basic watermarking step described above. The low bit rate chosen to implement this operation is the minimal bit rate for which near-transparent audio coding is known to be possible for signals sampled at the rate of the original signal. This scheme is more effective than other schemes that attempt to add the watermark on a lowpass filtered version of the signal because the coding/decoding operation is not a linear and does not permute with the watermarking operation. Fig. 2 illustrates the above procedure for signals sampled at an arbitrary sampling rate. The low-frequency watermarking signals is shown as w_{br} in Fig. 2. Here, the subscript br refers to the bit rate

of the coder/decoder.

For best watermark detection performance at higher bit rates, we need to add watermarking information in the higher frequency bands. We do so by producing a watermark w_{err} for the coding error. The coding error is the difference between the original audio signal and its low bit rate coded version. The watermark w_{err} is computed using the basic watermarking step described at the beginning of this section. The final watermark is the sum of the low-frequency watermark and the coding error watermark.

2.3 Listening tests: audibility of the watermarks

We used segments of four different musical pieces as test signals throughout the experiment: the beginning of the third movement of the sonata in B flat major D 960 of Schubert, interpreted by Vladimir Ashkenazy, a castanet piece, a clarinet piece, and a segment of "Tom's Diner" an *a capella* song by Suzanne Vega (svega). The Schubert signal is sampled at 32 kHz. All other signals are sampled at 44.1 kHz. Note that the castanets signal is one of the signals prone to pre-echoes. The signal svega is significant because it contains noticeable periods of silence. The watermark should not be audible during these silent periods.

The quality of the watermarked signals was evaluated through informal listening tests. In the test, the listener was presented with the original signal and the watermarked signal and reported as to whether any differences could be detected between the two signals. Eight people of varying backgrounds, including the authors, were involved in the listening tests. One of the listeners had the ability to perceive absolute pitch and two of the listeners had some background in music.

In all four test signals, the watermark introduced no audible distortion. No pre-echoes were detected in the watermarked castanet signal. The quiet portions of svega were similarly unaffected.

3 Detection of the Watermark

Let us now describe the watermark detection scheme and the detection results that we have obtained. In the experimental work described below, we used shaped inaudible noise to simulate attacks by pirates and distortions due to coding. We also tested the effects of filtering, coding, D/A - A/D converting and re-sampling on the detection performance of the proposed scheme. The detection results that we report below are based on processing 100 blocks of the observed signal of 512 samples. Note that this corresponds to 1.6 sec at the 32 kHz sampling rate and 1.16 sec at the 44.1 kHz sampling rate.

Our detection scheme assumes that the author has access to the original signal and the PN-sequence that he used to watermark the signal. It also assumes that the author has computed the approximate bit rate of the observed audio sequence $r(k)$. To decide whether the given signal $r(k)$ has been watermarked or not, the author subtracts from $r(k)$ a coded version s_{br} of the original audio signal $s(k)$. The signal s_{br} is produced by coding $s(k)$ at the estimated bit rate of $r(k)$ using the MPEG coding procedure. Note that $r(k)$ itself may have been coded using a different coding algorithm. The difference between the output of the MPEG coding algorithm operating on the original signal at the estimated bit rate and that of the actual coding algorithm at the true bit rate will appear as an additive noise signal.

Next, the author needs to solve the following hypothesis testing problem:

- $H_0 : x(k) = r(k) - s_{br}(k) = n(k)$
- $H_1 : x(k) = r(k) - s_{br}(k) = w'(k) + n(k)$.

Here, $n(k)$ denotes an additive noise process that includes errors due to different coding algorithms and signal manipulations, intentional jamming signals and transmission noise. The signal, $w'(k)$, is the modified watermark. Since the precise nature of $n(k)$ is unknown, we solve the above hypothesis testing problem by correlating $x(k)$ with $w'(k)$ and comparing the result with a threshold. Note that one needs to estimate time-scale modifications prior to correlations if such modifications have been performed on the signal. Fig. 3 shows the result of correlating a watermark corresponding to a segment of the Schubert audio piece with itself, the jammed watermark corrupted by frequency shaped noise of maximum masked intensity and shaped noise of maximum masked intensity alone. In all cases, the signal was not coded. The figure clearly indicates that reliable detection is feasible.

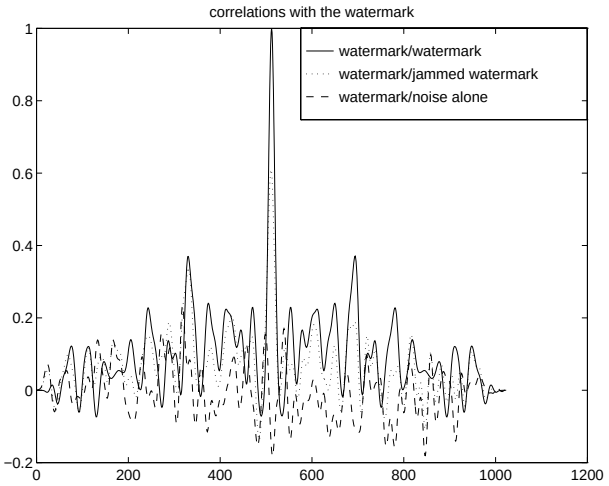

Figure 3: Detection of the watermark in Schubert with additive noise

3.1 Generation of the Additive Noise

Noise which has the same spectral characteristics as the masking threshold provides an approximation of the worst possible additive distortion to the watermark. This type of distortion is a good worst case model for distortions due to intentional jamming with inaudible signals and mismatches between the actual and assumed coding algorithms.

The noise that we have used in our experiments was generated in the same way as the watermark. Specifically, the masking threshold is first shifted $+40dB$ and multiplied by the discrete Fourier transform of a Gaussian white noise process. The resulting noise is then weighted in time by the relative energy of the signal. After quantization, we filter this shaped noise by the masking threshold and requantize it. The resulting noise is almost completely inaudible and is a good approximation of the maximum noise that we can add below the masking threshold.

3.2 Summary of Detection Results

Let us now summarize the detection results that we have obtained. Each group of results is meant to illustrate the robustness of our approach to a specific type of signal manipulation.

Robustness to MPEG coding for single and multiple watermarks
To test the robustness of our watermarking approach to coding, we added noise to several watermarked and non-watermarked audio pieces and coded the result. The watermarks were generated by using different PN sequences in different audio segments. The noise was almost inaudible and was generated using the technique described above. The coding/decoding was performed using a software implementation of the ISO/MPEG-1 Audio Layer III coder with several different bit rates. We then attempted to detect the presence of the watermark in the decoded signals. Table 1 shows that P_{detect} is 1 or nearly 1 in all cases and $P_{falsealarm}$ is nearly 0 in all cases.

We reported in [8] other detection results corresponding to an earlier implementation of our watermarking scheme that used the same PN sequence to watermark all segments of an audio piece. We also reported in that reference the results of detecting multiple watermarks added to a single audio piece. There are many instances where it is useful to add multiple watermarks to a signal. For example, there may be multiple authors for a piece of music, each with his/her own unique id. When detecting specific watermark, the other watermarks are considered to be noise. The results of [8] indicate that with one or more watermark, P_{detect}, is 1 or nearly 1 in all cases. Equally important, the probability of false alarm, $P_{falsealarm}$ is nearly 0 in all cases. These results, together with the ones presented here, establish the robustness of our scheme to MPEG coding and multiple watermarking.

Robustness to VQ distortion
We also tested the robustness of our watermarking approach to VQ coding. The codebooks consisted of 16 bit codewords. The audio signals were processed through codebooks of various sizes: 64, 128, 256, and 512 codewords. Although the signal was noticeably distorted, the watermark detection was unaffected, as shown in Table 2: P_{detect} is 1 or nearly 1 in all cases and $P_{falsealarm}$ is nearly 0 in all cases.

In [8], we also show that our watermarking scheme is robust to signal resampling. We are currently assessing the robustness of our scheme to time-scale modifications of the signal.

4 Conclusions

Our method for the digital watermarking of audio signals extends the previous work on images. Our watermarking scheme consists of a maximal length PN-sequence filtered by the approximate masking characteristics of the HAS and weighted in time, our watermark is imperceptibly embedded into the audio signal and easy to detect by the author thanks to the correlation properties of PN-sequences. Our results show that our watermarking scheme is robust in the presence of additive noise, lossy coding/decoding, VQ distortion, multiple watermarks, resampling, and time-scaling.

References

[1] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," *Proc. of the SPIE*, 1995.

[2] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia." Tech. Rep. 95-10, NEC Research Institute, 1995.

[3] O. Bruyndonckx, J.-J. Quisquater, and B. Macq, "Spatial method for copyright labeling of digital images," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 456–459, 1995.

[4] I. Pitas and T. Kaskalis, "Applying signatures on digital images," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 460–463, 1995.

[5] E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 452–455, 1995.

[6] F. Boland, J. OŔuanaidh, and C. Dautzenberg, "Watermarking digital images for copyright protection," *IEE Intl. Conf. on Image Proc. and Its Apps., Edinburgh, 1995.*

[7] L. Boney, A. Tewfik, K. Hamdy, and M. Swanson, "Digital watermarks for multimedia." Submitted to U.S. Patent Office, February 1996.

[8] L. Boney, A. Tewfik, and K. Hamdy, "Digital watermarks for audio signals," in *IEEE Int. Conf. on Multimedia Computing and Systems*, (Hiroshima, Japan), June 1996.

[9] M. Swanson, B. Zhu, and A. Tewfik, "Transparent robust image watermarking," in *to appear ICIP'96*, (Lausanne, Switzerland), Sept. 1996.

[10] S. Haykin, *Communication Systems, 3rd Edition.* John Wiley and Sons, 1994.

[11] "Codage de l'image animee et du son associe pour les supports de stockage numerique jusqu'a environ 1,5 mbit/s," tech. rep., ISO/CEI 11172, 1993.

Table 1: **Multiple PN sequence watermark with MPEG distortion**

Bit Rate	Watermark	*Schubert*	*Clarinet*	*Castanet*
kbits/sec	Threshold	0.65	0.47	0.54
48	P_{detect}	0.9922	na	na
	$P_{falsealarm}$	0.0117	na	na
64	P_{detect}	0.9961	1	1
	$P_{falsealarm}$	0	0	0.0031
128	P_{detect}	1	1	1
	$P_{falsealarm}$	0	0	0
160	P_{detect}	1	1	1
	$P_{falsealarm}$	0	0	0
224	P_{detect}	1	1	1
	$P_{falsealarm}$	0	0	0
320	P_{detect}	na	1	1
	$P_{falsealarm}$	na	0	0
	# of trials	257	83	639

Table 2: **Watermark detection with VQ distortion**

Bit Rate	Signal	*Clarinet*	*Castanet*	*Svega*
bits/sample	Threshold	0.54	0.46	0.52
6	P_{detect}	1	1	1
	$P_{falsealarm}$	0	0.0087	0
7	P_{detect}	1	1	1
	$P_{falsealarm}$	0.0010	0.01	0
8	P_{detect}	1	1	1
	$P_{falsealarm}$	0	0.0007	0
9	P_{detect}	1	0.9997	1
	$P_{falsealarm}$	0	0.0890	0
	# of trials	3000	3000	3000