

Audio Watermark Detection Improvement by Using Noise Modelling

NEDELJKO CVEJIC, TAPIO SEPPÄNEN*, DAVID BULL

Dept. of Electrical and Electronic Engineering *MediaTeam, Information Processing Lab.
 University of Bristol University of Oulu
 Merchant Venturers Building, BS8 1UB P.O. Box 4500, 4STOINF, 90014 Oulu
 UNITED KINGDOM FINLAND

Abstract: - In this paper we present an improved error correction approach for digital audio watermarking. It uses a turbo coding algorithm that takes into account temporal variations of the host audio's statistical properties, using a turbo decoder that estimates the unknown host audio distribution in the watermarked audio. Experimental results showed nearly one order of magnitude of the BER decrease during watermark extraction, compared to the basic turbo decoding that does not use noise estimation.

Key-Words: - Audio watermarking, watermark detection, noise modelling, turbo codes

1 Introduction

Digital watermarking is a process that embeds a perceptually undetectable signature to multimedia content (e.g. images, video and audio sequences). Embedded watermark contains certain information, as signature, logo or an ID number, related uniquely to the owner or the distributor or the multimedia file. Watermarking algorithms were primarily developed for digital images and video sequences; interest and research in audio watermarking started slightly later.

In the past few years, several algorithms for embedding and extraction of watermarks in audio sequences have been presented [1,2]. All of the developed algorithms take advantage of perceptual properties of the human auditory system (HAS), foremost occurrence of masking effects in the frequency and time domain, in order to add watermark into a host signal in a perceptually transparent manner. Embedding of additional bits in audio signal is a more tedious task than implementation of the same process on images and video, due to the dynamical superiority of HAS in comparison with the human visual system [2].

Information modulation is usually carried out using the quantisation index modulation (QIM) [1] or the spread-spectrum (SS) [2,3] technique. SS modulation augments a low amplitude sequence, which is detected by a correlation receiver. The basic approach to watermarking in the time domain is to embed a pseudo-random noise (e.g. the PN sequence) into the host audio by modifying the amplitudes accordingly. Recently, we have developed a spread-spectrum audio watermarking algorithm in time domain [4], presented in Figure 1. The procedure used a time domain embedding

algorithm and properties of spread spectrum communications as well as temporal and frequency domain masking in the HAS. Matched filter technique, based on auto-correlation of the embedded PN sequence, is optimal in the sense of signal to noise ratio (SNR) in the additive white Gaussian noise (AWGN) channel [5]. However, the host audio signal's statistical properties are generally far from the properties of the AWGN, which leads us to the optimal detection problem, since correlation based receivers are optimal in AWGN. In that sense, an advanced noise model of the audio watermark channel is needed in order to increase performance of the watermarking algorithm based on the SS technique.

In this paper we focus on an improved error correction approach for audio watermarking. It uses a turbo coding algorithm that takes into account temporal variations of the host audio's statistical properties in order to increase the overall watermark detection performance of the watermarking system.

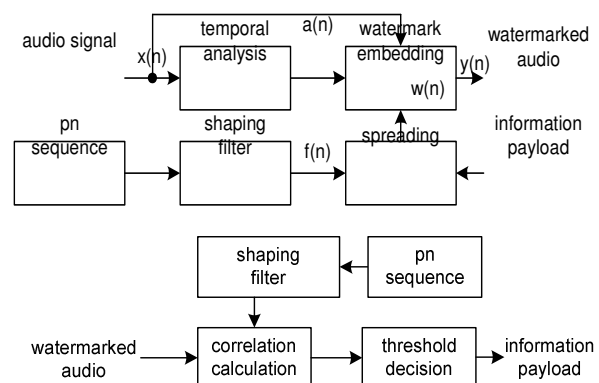


Fig. 1. Watermark embedding scheme (a) and extraction scheme (b)

2 Noise Characteristics of the Host Audio

In a correlation detection scheme, it is often assumed that the host audio signal is AWGN. However, real audio signals do not have white noise properties as adjacent audio samples are highly correlated. Therefore, presumption for optimal signal detection in the sense of signal to noise ratio is not satisfied, especially if extraction calculations is performed in short time windows of audio signal. Figure 2 depicts a histogram estimate of the probability density function (PDF), performed on 1024 successive samples of a short clip of the host audio (Celine Dion), wherein a watermark bit is embedded. It is obvious that the PDF of the host audio is not smooth and has a large variance.

Figure 3 presents the values of skewness and kurtosis of the PDF of the windows of 1024 samples in time domain, taken from the host audio $x(n)$. Taken that μ is mean, μ_3 is the third order moment, μ_4 the fourth order moment and σ the standard deviation of a distribution, the skewness is defined by $S=\mu_3/\sigma^3$ and is an indicator of the PDF symmetry (for a symmetric PDF, $S=0$) [6]. Kurtosis is defined by $K=\mu_4/\sigma^4-3$ and is an indicator of the PDF Gaussianity (for a Gaussian PDF, $K=0$). Therefore, this particular digital communications scheme, in which we communicate watermark bits through the "noisy" host audio, does not obey the AWGN hypothesis.

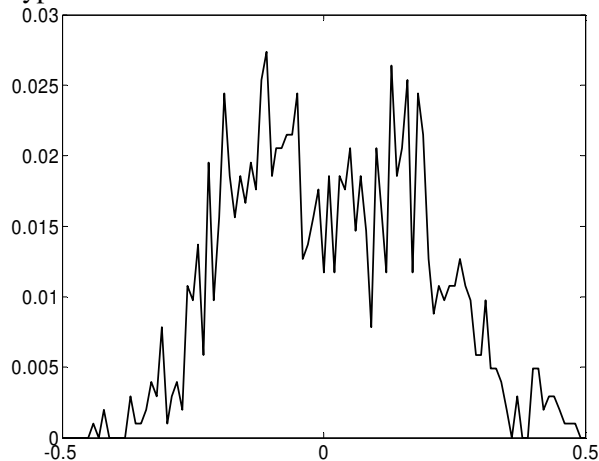


Fig. 2. Histogram PDF estimation of the host audio signal

It is shown below that both Gaussian and Laplacian distributions in the PDF occur in the host audio $x(n)$. The Generalized Gaussian Distribution (GGD) is often used as a model for noise PDF in digital watermarking. It is defined as [6]:

$$p(x) = \frac{A}{\sigma} \exp\left(-\left|b \frac{x - \mu}{\sigma}\right|^\alpha\right)$$

where α is the shape parameter, σ the standard deviation, μ the mean value,

$$A = \frac{\alpha}{2} \frac{[\Gamma(3/\alpha)]^{1/2}}{[\Gamma(1/\alpha)]^{1/2}}, b = \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}$$

and Γ is the Gamma function. When $\alpha=1$ a Laplacian distribution is obtained, while $\alpha=2$ yields a Gaussian distribution. In the extreme cases, for $\alpha \rightarrow 0$ $p(x)$ becomes an impulse function, whereas for $\alpha \rightarrow \infty$, $p(x)$ approaches a uniform distribution. The shape parameter α rules the exponential rate of decay: the larger α , the flatter the PDF, the smaller α , the peak of the PDF is more emphasized. There are many methods for parametric estimation of a GGD. In this paper, we use the method based on the moments estimation and that gives an approximation of the reciprocal of the function $M(\alpha)$ defined as [6]:

$$M(\alpha) = \frac{(E|X|)^2}{E(X^2)}$$

where X is a GGD random variable.

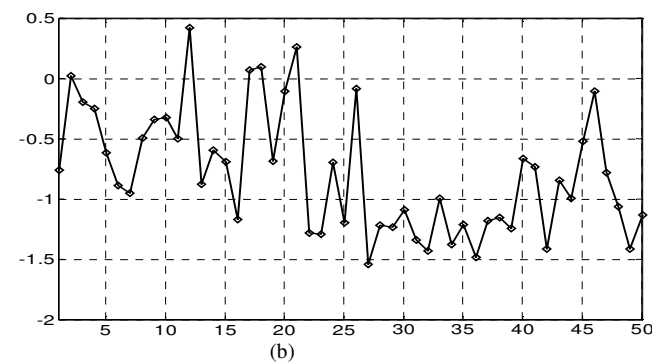
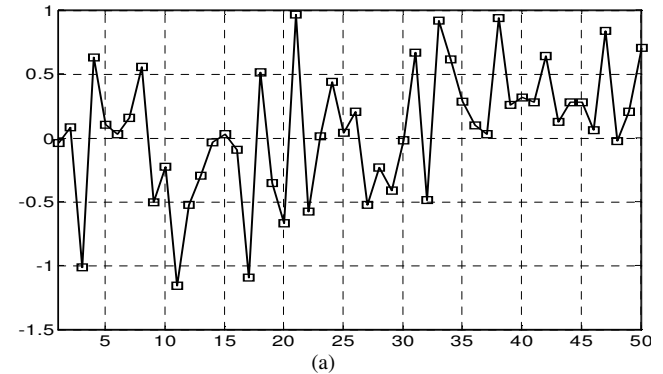


Fig. 3. Skewness (a) and kurtosis (b) values for 50 adjacent blocks with 1024 samples of the host audio

We calculated a piecewise estimate of the GGD for the sample audio extract on a window of 1024 samples. The data show that the shape parameter α varies significantly around the value $\alpha=2$, corresponding to the Gaussian distribution. In fact, α varies between the values of the theoretical Laplacian PDF and the theoretical Gaussian one for 10 different windows of 1024 samples. Similar observations can be done from other audio samples

as well. It is concluded therefore that the GGD is a preferred model for additive noise in audio watermarking as well.

In the audio watermarking framework, there is a low SNR at the watermark detection side, due to the perceptual constraints. If a low bit error rate (BER) is desired, powerful error correcting codes must be used, at the cost of the decreased watermark bit rate. For example, convolutional error correcting codes have been extensively used in image watermarking [7] due to their low computational complexity. We tested the audio watermarking scheme presented in Figure 1 with a convolutional code (R=1/2, K=3). No significant gain in BER was obtained (for detailed results see Section 4.) This result is due to the fact that the tested convolutional code is not appropriate for the SNR range of the presented audio watermarking scheme and the characteristic piecewise stationarity of the host audio. Therefore, a different error correction strategy with real time decoding metric should be used. Reasonable choices are powerful error correcting codes, suitable to low SNR values at the detection side of the given audio watermarking scheme, as turbo codes and low density parity check codes.

3 Noise-Adaptive Turbo Decoding Algorithm

The existing methods of channel coding do not take into account the knowledge of temporal variations of the channel statistics and the decoding algorithms are generally based on the AWGN hypothesis. In order to compensate for the temporal variations we have employed turbo codes that are able to adapt to host audio distribution variations [8]. These turbo decoders have a simple on-line procedure for estimating the unknown noise distribution from each block of the watermarked audio. The procedure is performed in two steps:

1. Quantization of the watermarked audio.
2. Estimation of the host audio distribution from the histogram of the quantized watermarked audio.

The purpose of quantization is to reduce the problem of estimating the conditional PDF to the problem of estimating the conditional probability mass function (PMF), which is, in general, considerably simpler. As proposed in [9] we have used an N-level uniform quantizer (where N is an integer power of 2). For $N \leq 16$, the quantization thresholds are defined as:

$$T_j = \begin{cases} -\infty, & \text{if } j = 0 \\ j / 2^{q-2} - 2, & \text{if } j = 1, 2, \dots, 2^q - 1 \\ +\infty, & \text{if } j = 2^q \end{cases}$$

where $q = \log_2 N$. For $N \geq 32$, the subsequent thresholds are used:

$$T_j = \begin{cases} -\infty, & \text{if } j = 0 \\ j / 2^{q-4} - 8, & \text{if } j = 1, 2, \dots, 2^q - 1 \\ +\infty, & \text{if } j = 2^q \end{cases}$$

The quantization step size is partially dependent on the number of levels, in order to keep the quantize support region within a reasonable range.

Let $y(n)$ be watermarked audio at the detection side, then we denote the quantized watermarked audio block as $\hat{y}(n)$ and watermark sequence as $w(n)$. Along with the quantized input, the turbo decoder is provided with the conditional PMFs $p(\hat{y}(n) | w(n)) = +1$ and $p(\hat{y}(n) | w(n)) = -1$. For each block of samples, we calculate the histogram $h(\hat{y})$ of $\hat{y}(n)$ and symmetrize it by $h_s(\hat{y}) = (h(\hat{y}) + h(-\hat{y})) / 2$. When $N \geq 32$, the expression for $p(\hat{y} | w) = +1$ is given as follows:

$$p(\hat{y} | w = +1) = \begin{cases} h_s(y), & \text{if } \hat{y} \geq +1 \\ h_s(2 - y), & \text{if } -6 \leq \hat{y} \leq +1 \\ (\hat{y} + a)h_s(a) / 2, & \text{if } \hat{y} < -6 \end{cases}$$

where $a = 8 - 16/N$. For $N \leq 16$, the exact formula for $p(\hat{y} | w) = +1$ is:

$$p(\hat{y} | w = +1) = \begin{cases} h_s(\hat{y}) & \text{if } \hat{y} \geq +1 \\ h_s(2 - \hat{y}) & \text{if } -2 \leq \hat{y} < +1 \end{cases}$$

If $p(\hat{y} | w) = +1$ is zero, it is set to a small number.

Watermark bits were encoded before they were embedded into the host audio and iteratively decoded using the soft output values from correlator during watermark extraction process [10]. Watermark bits were divided in frames of 200 bits and encoded using multiple parallel-concatenated convolutional code. Interleaving inside frame was random and five decoding iterations of soft output values were performed in turbo decoder. Each recursive systematic code was an optimum (5,7) code, giving a punctured code rate of $R = 1/2$. Frame length and code rate were chosen as a compromise between low computational complexity requirements of the watermarking algorithm and demand for long iterations during turbo decoding process.

4 Experimental results

The influence of described turbo codes on the BER of the watermarking system has been tested using a large set of songs from different music styles

including pop, rock, classic and country music. All music pieces have been watermarked using the described algorithm, with overall watermark to signal ratio from -26.5 dB to -28.1 dB. Subjective quality evaluations of the watermarking method has been done by blind listening tests involving ten persons that listened to the original and the watermarked audio sequences and were asked to report dissimilarities between the two signals, using a 5-point impairment scale. (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying 1: very annoying.) The average mean opinion score was 4.6 and the standard deviation 0.41.

The watermark extraction was performed using the scheme in Figure 1(b) and the results are presented in Figure 4. The soft output values from correlator during watermark extraction process were used as inputs to convolutional and turbo decoders or for hard threshold decision, in uncoded detection. The results thus show that the given convolutional code does not improve the detection performance on the watermark extraction side. On the other hand, for a fixed watermark capacity, turbo code is able to decrease significantly the BER. If the proposed host audio PDF estimation is used, the detection performance of the watermark extraction system is significantly improved, compared with the basic turbo coding scheme.

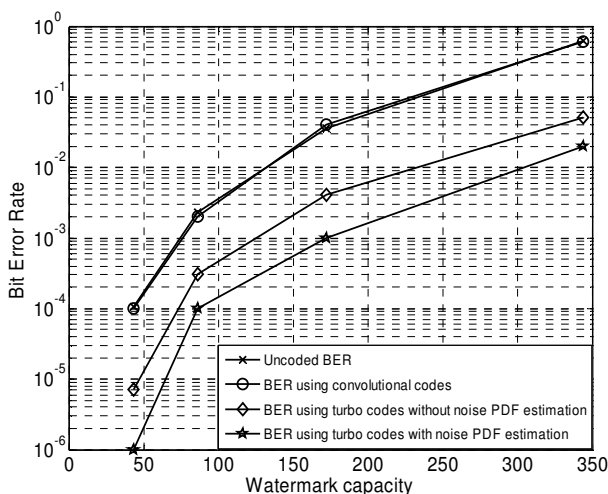


Fig. 4. BER vs. watermark data rate (in bits/second) for different decoding algorithms

5 Conclusions

Careful noise modelling of audio signals must be performed in order to get best results in watermarking performance. It was demonstrated that the noise model may switch between Laplacian and Gaussian distributions in a random manner within a

short excerpt of music. An adaptive noise estimation function should be included as part of audio watermarking algorithms.

An improved error correction approach for audio watermarking was presented that uses a noise-adaptive turbo decoding algorithm to take into account temporal variations of the host audio's statistical properties. The implemented turbo decoder estimates the unknown probability density distribution of the host audio, as one component of the watermarked audio. Experimental results showed improved watermark extraction, compared to the results obtained by basic turbo decoding that does not use noise estimation.

References:

- [1] Chen, B., Wornell, G.W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, Vol. 47, No. 4, 2001, pp. 1423-443.
- [2] Cox, I.J., Kilian, J., Leight, F.T., Shamoon, T. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, Vol. 6, No. 12, 1997, pp. 1673-87.
- [3] Cvejic, N., Seppänen, T. Spread spectrum audio watermarking using frequency hopping and attack characterization. *Signal Processing*, Vol. 84, No. 1, 2004, pp. 207-213.
- [4] Cvejic, N., Keskinarkaus, A., Seppänen, T. Audio watermarking using m-sequences and temporal masking. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, NY, 2001, 227-230.
- [5] Kundur, D., Hatzinakos, D. Diversity and Attack Characterization for Improved Robust Watermarking", *IEEE Transactions on Signal Processing*, Vol. 49, No. 10, 2001, pp. 2383-6.
- [6] Scharf, L. *Statistical Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 2002.
- [7] Hernandez, J. R., Delaigle, J. F., Macq, B. Improving data hiding by using convolutional codes and soft-decision decoding. *Proc. SPIE Security and Watermarking of Multimedia Contents*, Vol. 3971, San Jose, CA, 2000, pp. 24-48.
- [8] Saied-Bouajina, S., Larbi, S., Hamza, R., Slama, L.B., Jidane-Saidane, M. An error correction strategy for digital audio watermarking scheme. *Proc. International Symposium on Control, Communication & Signal Processing*, Hammamet, Tunisia, 2004, pp. 739-742.
- [9] Xiaoling, H., Nam, P. Turbo decoders which adapt to noise distribution mismatch. *IEEE*

Communications Letters, Vol. 2, No. 12, 1998, pp. 321-323.

[10] Cvejic, N., Tujkovic, D., Seppänen, T. Increasing capacity of an audio watermark channel using turbo codes. *Proc. IEEE International Conference on Multimedia & Expo*, Baltimore, MD, 2003, pp. 217-220.