Experimental Validation for Hiding Data Using Audio Watermarking

1Mamoun Suleiman Al Rababaa, 2Ahmad Khader Haboush, 3As'ad Mahmoud Al Naser
1Computer Science Department, Al al-Bayt University.
2Computer Science Department, Jerash University.
3Computer Science Department, Al-Balqa’ Applied University, Ajlun University College.

Abstract: Digital watermarking has been proposed as a new, alternative method to enforce intellectual property rights and protect digital media from tampering. Digital watermarking is defined as imperceptible, robust and secure communication of data related to the host signal. We include here theoretical consideration and experimental validation, including digital signal processing, psychoacoustic modeling and communications theory. The main technical solutions include algorithms for embedding high data rate watermarks into the host audio signal, using channel models and the detection and modeling of attacks.

Key words: audio watermarking, digital rights management, information hiding, Steganography.

INTRODUCTION

Simple protection mechanisms that were based on the information embedded into header bits of the digital file are useless because header information can easily be removed by a simple change of data format, which does not affect the fidelity of media. Modern software and broadband Internet provide the tools to perform it quickly and without much effort and deep technical knowledge. One of the more recent examples is the hack of the Content Scrambling System for DVDs [5,6].

Digital watermarking has been proposed as a new, alternative method to enforce the intellectual property rights and protect digital media from tampering. It involves a process of embedding into a host signal a perceptually transparent digital signature, carrying a message about the host signal in order to "mark" its ownership.

Information hiding (or data hiding) is a more general area, encompassing a wider range of problems than the watermarking (Cox, I., 2003). The term hiding refers to the process of making the information imperceptible or keeping the existence of the information secret. We can define watermarking systems as systems in which the hidden message is related to the host signal.

The primary focus of this paper is the watermarking of digital audio (i.e., audio watermarking), including the development of new watermarking algorithms and new insights of effective design strategies for audio steganography.

Watermarking algorithms can be characterized by a number of defining properties (Cox, I., 2003). Six of them, which are most important for audio watermarking algorithms:

Perceptual Transparency:
Audio signal received by the watermarking algorithms from the audio has signal keeping the perceptual similarity between the original and watermarked audio sequence.

Watermark Bit Rate:
The bit rate of the embedded watermark is the number of the embedded bits within a unit of time and is usually given in bits per second (bps) (Yu, H., 2001).

Robustness:
The robustness of the algorithm is defined as an ability of the watermark detector to extract the embedded watermark after common signal processing manipulations.

Corresponding Author: Mamoun Suleiman Al Rababaa, Computer Science Department, Al al-Bayt University.
**Blind or Informed Watermark Detection:**
In some applications, a detection algorithm may use the original host audio to extract watermark from the watermarked audio sequence (informed detection) (Yu, H., 2001).

**Security:**
Watermark algorithm must be secure in the sense that an adversary must not be able to detect the presence of embedded data, let alone remove the embedded data (Yu, H., 2001).

**Computational Complexity and Cost:**
The implementation of an audio watermarking system is a tedious task, and it depends on the business application involved. The principal issue from the technical point of view is the computational complexity of embedding and detection algorithms and the number of embedders and detectors used in the system (Cox, I., 2003).

**Paper Hypothesis (R.H):**
P.H1: To obtain a distinctively high watermark data rate, embedding algorithm can be implemented in a transform domain, with the usage of the least significant bit coding.
P.H2: To improve detection performance, a spread spectrum method can be used, cross correlation between the watermark sequence and host audio decreased and channel coding introduced.
P.H3: To achieve the robustness of watermarking algorithms, an attack characterization can be introduced at the embedder, improved channel model can be derived and informed detection can be used or watermark decoding.

**Paper Assumptions:**
The general research assumption is that the host signal is divided into two channel signals (right and lift signal), the process of embedding and extraction of watermarks can be modeled as a communication system, where the watermark embedding is modeled as a transmitter, and the distortion of watermarked signal as a communications channel noise and watermark extraction as a communications detector (Yu, H., 2001).

A central assumption in the security analysis of the proposed algorithms is that an adversary that attempts to disrupt the communication of watermark bits or remove the watermark does not have access to the original host audio signal. The watermarking process has to be characterized. This contains:
- The set of watermark messages M.
- The embedding function depending on the watermark message m and key K and constrained to the embedding distortion $d_{emb}$.
- The decoding function, which depends on the key K.

**Frequency Masking:**
Frequency (simultaneous) masking is a frequency domain phenomenon where a low level signal, e.g. a pure tone (the maskee), can be made inaudible (masked) by a simultaneously appearing stronger signal (the masker), e.g. a narrow band noise, if the masker and maskee are close enough to each other in frequency (Zwicker, E. and H. Fastl, 1999). A masking threshold can be derived below which any signal will not be audible.

It should be pointed out that the distance between masking level and masking threshold is smaller in noise-masks tone experiments than in tone-masks-noise experiments due to human auditory system’s (HAS) sensitivity toward additive noise. Noise and low-level signal components are masked inside and outside the particular critical band if their sound pressure level (SPL) is below the masking threshold. Noise contributions can be coding noise, inserted watermark sequence, aliasing distortions, etc. Without a masker, a signal is inaudible if its SPL is below the threshold in quiet, which depends on frequency and covers a dynamic range of more than 70 dB as depicted in the lower curve of Figure 1. The qualitative sketch of Figure 2 gives more details about the masking threshold. The distance between the levels of the masker (given as a tone in Figure 2) and the masking threshold is called signal-to-mask ratio (SMR). Its maximum value is at the left border of the critical band. Within a critical band, noise caused by watermark embedding will be audible as long as signal-to-noise ratio (SNR) for the critical band is higher than its SMR. Let $SNR(m)$ be the signal-to-noise ratio resulting from watermark insertion in the critical band m, the perceivable distortion in a given sub-band is then measured by the noise to mask ratio (Noll, P., 1993):
The noise-to-mask ratio NMR(m) expresses the difference between the watermark noise in a given critical band and the level where a distortion may just become audible, its value in dB should be negative.

A global masking threshold can be computed that describes the threshold of just noticeable distortion (JND) as a function of frequency (Zwicker, E. and H. Fastl, 1999).

The calculation of the global masking threshold is based on the high resolution short-term amplitude spectrum of the audio signal, sufficient for critical band-based analysis and is usually performed using 1024 samples in FFT domain. In a first step, all the individual masking thresholds are determined, depending on the Signal level, type of masker (tone or noise) and frequency range. After that, the global masking threshold is determined by adding all individual masking thresholds and the threshold in quiet (Noll, P., 1993). The effects of the masking reaching over the limits of a critical band must be included in the calculation as well.

Fig. 1: Frequency masking in the human auditory system (HAS), reference sound pressure level is $p_0 = 2 \times 10^{-5}$ Pa.

$$NMR(m) = SMR - SNR(m)$$

(1)
A General Model of Digital Watermarking:

A watermark message \( m \) is embedded into the host signal \( x \) to produce the watermarked signal \( s \). The embedding process is dependent on the key \( K \) and must satisfy the perceptual transparency requirement, i.e. the subjective quality difference between \( x \) and \( s \) (denoted as embedding distortion \( d_{emb} \)) must be below the just noticeable difference threshold. Before the watermark detection and decoding process takes place, \( s \) is usually intentionally or unintentionally modified. The intentional modifications are usually referred to as attacks, an attack produce attack distortion \( d_{att} \) at a perceptually acceptable level. After attacks, a watermark extractor receives attacked signal \( r \).

The watermark extraction process consists of two sub-processes, first, watermark decoding of a received watermark message \( m^* \) using key \( K \), and, second, watermark detection, meaning the hypothesis test between (Yu, H., 2001):

- Hypothesis \( H_0 \): the received data \( r \) is not watermarked with key \( K \).
- Hypothesis \( H_1 \): the received data \( r \) is watermarked with key \( K \).

The detection process verifies two hypotheses on the received content:

- \( H_0 \): watermarked audio content, so it is Gaussian white noise residual signal of host audio after decorrelation process.
- \( H_1 \): consists of de-correlated host audio and watermark as Decorrelation pre-processing was implemented, we can assume that the output of de-correlation filter \( y \) for a given \( w \) has the Gaussian distribution and the Likelihood Ratio Test can be performed.

Statistical Modeling of Digital Watermarking:

In order to properly analyze digital watermarking systems, a stochastic description of the multimedia data is required. The watermarking of data whose content is perfectly known to the adversary is useless. Any alteration of the host signal could be inverted perfectly, resulting in a trivial watermarking removal. Thus, essential requirements on data being robustly watermark-able are that there is enough randomness in the structure of the original data and that quality assessments can be made only in a statistical sense.

Let the original host signal \( x \) be a vector of length \( L_x \). Statistical modeling of data means to consider \( x \) a realization of a discrete random process \( x \) (Eggers, J. and B. Girod, 2002). In the most general form, \( x \) is described by a \( L_x \)-dimensional probability density function \( p_x(x) \).

\[
L_x \prod_{n=1}^{L_x} p_x(x_n) \quad (2)
\]

With \( p_x(x_n) \) being the \( n \)th marginal distribution of \( x \). A further simplification is to assume independent identically distributed (IID) data elements so \( p_x(x_n) = p_x(x) = p(x) \). Most multimedia data cannot be modeled adequately by an IID random process (Eggers, J. and B. Girod, 2002).

Watermark Decoding:

In the performance evaluation of the watermark decoding, digital watermarking is considered as a communication problem. A watermark message \( m \) is embedded into the host signal \( x \) and must be reliably decodable from the received signal \( r \). Low decoding error rates can be achieved only using error correction codes. For practical error correcting coding scenarios, the watermark message is usually encoded into a vector \( b \) of length \( L_b \) with binary elements \( b_n = 0, 1 \). Usually, \( b \) is also called the binary watermark message, and the decoded binary watermark message is \( \hat{b} \). The decoding reliability of \( b \) can be described by the word error probability (WEP)

\[
P_w = P_r(m \neq \hat{m}) = P_r(b \neq \hat{b}) \quad (3)
\]

or by the bit error probability (BEP)

\[
P_b = \frac{1}{L_b} \sum_{n=1}^{L_b} P_r(b_n \neq \hat{b}_n) \quad (4)
\]
The WEP and BEP can be computed for specific stochastic models of the entire watermarking process including attacks. The predicted error probabilities can be confirmed experimentally by a large number of simulations with different realizations of the watermark key \( K \), the host signal \( x \), the attack parameters and a watermark message \( m \). The number of measured error events divided by the number of the observed events defines the measured error rates, word error rate, WER and bit error rate BER.

Performance limits can be derived with methods borrowed from the information theory. For example, the maximum watermark rate which can be received in principle without errors is determined by the mutual information \( I(r|m) \) between the transmitted watermark message \( m \) and received data \( r \) and given by (Ramkumar, M. and A. Akansu, 1988):

\[
I(r|m) = h(r) - h(r|m)
\]  

(5)

Where \( h(r) \) is the differential entropy of \( r \) and \( h(r|m) \) is the differential entropy of \( r \) conditioned on the transmitted message \( m \). The PDFs \( p_r(r) \) and \( p_r(r|m = m) \) are required for the computation of \( h(r) \) and \( h(r|m) \). \( I(r|m) \) can be achieved only for an infinite number of data elements. For a finite number of data elements, a non-zero word error probability \( p_w \) or a bit error probability \( p_b \) are unavoidable.

The channel capacity \( C \) of a specific communication channel is defined as the maximum mutual information \( I(r, m) \) over all transmissions schemes with a transmission power constrained to a fixed value (Ramkumar, M. and A. Akansu, 1988). The watermark capacity \( C \) is defined correspondingly with a slight modification specific for watermarking scenarios. The capacity analysis provides a good method for comparing the performance limits of different communication scenarios, and thus is frequently employed in the existing literature. Since there is still no solution available for the general watermarking problem, digital watermarking is usually analyzed within certain constraints on the embedding and attacks.

**Watermark Detection:**

Watermark detection is defined as the decision whether the received data \( r \) is watermarked (\( H_1 \)) or not watermarked (\( H_0 \)) (Eggers, J. and B. Girod, 2002). In general, both hypotheses cannot be separated perfectly. Thus, we define the probability \( p_{fp} \) (false positive) as the case of accepting \( H_1 \) when \( H_0 \) is true and the probability \( p_{fn} \) of accepting \( H_0 \) when \( H_1 \) is true (false negative).

In this paper, the watermark detection is based on watermarking schemes that have been designed for reliable communication of a binary watermark message \( b \). A sub-vector \( f \) of length \( L_f \) of the watermark message \( b \) is used for a validity verification of a received watermark message \( b^\cdot \).

Two simple watermark detection methods using the verification bit vector \( f \) are discussed:

**Hard Decision Decoding:**

The verification message \( f \) is encoded together with all remaining watermark message bits to obtain the encoded watermark message \( b_c \). During the watermark extraction, the message \( b^\cdot \) is as in the communication scenario. One fraction of \( b^\cdot \) is the decoded watermark verification message \( f^\cdot \) that must be equal to \( f \) for a valid watermark message \( b^\cdot \). Therefore, the hypothesis decision rule is given by:

\[
H_0; f \neq f
\]  

(6)

\[
H_1; f = f
\]  

(7)

The false positive probability \( p_{fp} \) can be calculated based on the assumption that \( P_r(f^\cdot = 0|H_0) = P_r(f^\cdot = 0|H_1) = 0.5 \). The probability \( p_{fp} = 0.5 \) is obtained for \( L_f \) independent bits \( f \) and depends only on the number \( L_f \) of verification bits. The false negative probability depends on the bit error probability \( p_b \) and the number of verification bits \( L_f \) in the expression (h) statistically

\[
p_{fn} = 1 - (1 - p_b)^{L_f}
\]  

(8)

Independent received verification bits \( f^\cdot \) are assumed. In practice, the interleaving of all bits in \( b \) before error correction encoding is useful to ensure the validity of those assumptions. A generalization of the decision rule given above is to accept \( H_1 \) if the Hamming distance (Ramkumar, M. and A. Akansu, 1988), \( d_H(f^\cdot , f) \) is lower than a certain threshold. In that case, the threshold could be designed to find a better trade-off between \( p_{fp} \) and \( p_{fn} \).
**Soft Decision Decoding:**

Detection based on a hard decision decoding is very simple. However, if the accurate statistical models of the introduced attacks are known, soft decision decoding gives potentially a better detection performance. The verification message \( f \) is equal to the first \( L_f \) bits of \( b \) and error correction coding of \( b \) is such that the first \( L_f \) bits of the coded watermark message \( b_c \) are independent of the remaining watermark message bits. Without a loss of generality, we can assume:

\[
(b_c, 0, \ldots, b_c, L_{f-1}) = f = 0
\]  

(9)

Let \( L_f \) denote the set of the indices of all data elements with embedded coded verification bits. We assume that the PDFs \( p_r(r_i | H_0) \) and \( p_r(r_i | H_1) \) for receiving \( r_i \) depending on hypothesis \( H_0 \) or \( H_1 \), respectively, are known. Bayes’ solution to the hypothesis testing problem can be applied, which is:

\[
\frac{P_r(r_f | H_1)}{P_r(r_f | H_0)} > T \Rightarrow \text{accept } H_1, \text{else } \Rightarrow \text{accept } H_0
\]  

(10)

Where \( T \) is the decision threshold. \( T \) is a constant depending on the a priori probabilities for \( H_1 \) and \( H_0 \) and the cost connected with different decision errors. For \( T = 1 \), the decision rule above forms a maximum-likelihood (ML) detector. For equal a priori probabilities, the decision error probability is \( p_e = 1/2 (p_f + p_f) \). Assuming equal a priori probabilities and equal costs for both hypotheses, the above decision rule can be reformulated so that \( H_1 \) is accepted if:

\[
P_r = \frac{P_r(r_f | H_1)}{P_r(r_f | H_1) + P_r(r_f | H_0)} > 0.5
\]  

(11)

Where \( p_r \in [0, 1] \) denotes the reliability that a received watermark message \( b^r \) is a valid watermark message. For decision above, \( p_r \) and \( p_r \) depend directly on the PDFs \( p_r (r_i | H_0) \) and \( p_r (r_i | H_1) \).

**Exploiting Side Information During Watermark Embedding:**

In most blind watermarking schemes, as in a blind spread spectrum watermarking, the host signal is considered as interfering noise during the watermark extraction. Nevertheless, recently it has been realized that a blind watermarking can be modeled as communication with side information at the encoder. The embedding process exploiting side information of the host signal is separated into two parts: first, an appropriate watermark sequence \( w \) representing the watermark message \( m \) is selected, and, second, \( w \) is added to the host signal \( x \). The MSE distortion measure is used so that:

\[
d_{emb} = \frac{1}{L_x} E \left\{ \| s - x \|^2 \right\} = \frac{1}{L_x} E \left\{ \| w \|^2 \right\}
\]  

(12)

The mapping of \( m \) onto sequence \( w \), also of length \( L_w \), is determined by \( x \) and the by codebook \( W^x (K) \), which is encrypted with the watermark key \( K \). Secrecy is obtained by a pseudo-random selection of all entries in \( W^x (K) \).

The assumption is that the watermark sequences \( w \) are zero mean and IID. The embedding distortion \( d_{emb} \) is then equal to the variance \( \sigma_w \) of the watermark elements \( w_n \). The AWGN attack is independent of the characteristics of the original and watermarked signal so that attack distortion is \( d_{att} = d_{emb} + \sigma_w = \sigma_w + \sigma_x \). It should be noted that a blind spread spectrum watermarking also fits into the given model. For the spread spectrum watermarking, the codebook \( W^x (K) \) contains all combinations of possible messages \( m \) and spreading sequences derived from \( K \), which is a finite number of sequences. Furthermore, the performance limit of an optimal non-blind watermarking scheme can also be considered as the ultimate performance limit of blind watermarking (Yu, H., 2001).

**Capacity of the Data-hiding Channel:**

First we consider a simple data-hiding channel shown in Figure 3. Here, \( X \sim (f_x(x), \sigma_x) \) is the message to be embedded, \( Z \sim (f_z(x), \sigma_z) \) is the additive noise channel and \( Y \sim (f_y(x), \sigma_y) \) is the received signal at the
output of the channel. We also assume $X$ and $Z$ are independent, implying that $\sigma_y = \sigma_z + \sigma_x$. The channel capacity is given by:

$$
C = \max_{f_X(x)} I(X,Y) = \max_{f_X(x)} h(Y) - h(Y | X) = \max_{f_X(x)} h(Y)^2 - h(Z)[\text{bits}]
$$

(13)

$I(X,Y)$ is the mutual information between $X$ and $Y$. For a given statistics $f_z(z)$ and $\sigma_x$, the entropy of $Y$ should be maximized, $h(Y) = - \int f_Y(y) \log_2 (f_Y(y)) dy$ [bits], using a suitable distribution $f_X(x)$ of the message $X$. For a given $\sigma_x$, the maximum value of $h(Y)$ is $1/2 \log_2 (\frac{2\pi e}{\sigma_y})$ bits when $Y$ has a normal distribution. For instance, the maximum value of $h(Y)$ is achievable if both $f_z(z)$ and $f_X(x)$ are normally distributed. However, for an arbitrary distribution $f_z(z)$ and a fixed $\sigma_x$, the maximum achievable value of $h(Y)$ is not immediately obvious. This is because $Z$ is usually altered in such a manner that the amount of information in $Z$ is not altered, but the statistics of $Z$ is changed to Gaussian distributed $Z_g$. Therefore, for the purpose of calculating the channel capacity, we can replace $f_z(z)$ by $N(0, \sigma_{zg})$ and $h(Z) = h(Z_g) = \frac{1}{2} \log_2 (\frac{2\pi e}{\sigma_{zg}})$ and we get:

$$
C = \max_{f_X(x)} h(Y) - h(Z_g)[\text{bits}] = \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_x^2}{\sigma_{zg}^2} \right)[\text{bits}]
$$

(14)

The general data-hiding channel is usually decomposed into multiple channels, as hiding process is performed in a transform domain (Ramkumar, M. and A. Akansu, 2001). The decomposition is performed by the forward and inverse transform, as depicted in Figure 4. Signal decomposition into L bands results in L parallel channels with two noise sources in each channel. Let $\sigma_{ij}$, $j = 1, \ldots, L$ be the variances of the coefficients of each band of the decomposition. Let the corresponding Gaussian variances be $\sigma_{ijg}$. If $\sigma_{ijg}$ is the variance of the processing noise in the $j$th channel, the total capacity of the L parallel channels is given by:

$$
C_h = \frac{N^2}{2L} \sum_{j=1}^{L} \log_2 \left( 1 + \frac{T_j}{\sigma_{ijg}^2 + \sigma_{ij}^2} \right)[\text{bits}]
$$

(15)

For a sequence of $N$ samples. In the equation (o), $T_j$ is the masking threshold of band $j$, in other words, the maximum power of the embedded message permitted in band $j$. In the case of no-processing noise (or if the processing noise is negligible), and we assume that all the channel have the same probability distribution function (such that $K \sigma_{ij} = K \sigma_{ijg}$), the channel capacity is given by:

$$
C_h = \frac{N^2}{2L} \sum_{j=1}^{L} \log_2 \left( 1 + \frac{1}{\sigma_{ijg}^2} \right) \approx \frac{N}{2L} \log_2 \left( 1 + \frac{1}{\sigma_{ijg}^2} \right)[\text{bits}]
$$

(16)

It is clear that the minimum channel capacity is obtained when $\sigma_{ijg} = \sigma_{ij}$ or when no decomposition is employed. A transform with a good energy compaction or high gain of transform coding (GTC) (Mallat, S., 2001) would result in more imbalances of the coefficient variances, resulting in an increased channel capacity. Therefore, a wavelet decomposition or discrete cosine transforms (DCT) are good decompositions for low processing noise scenarios. The term processing noise here refers to equivalent additive noise which accounts for the reduction in correlation between the transform coefficients of the original signal and the transform.
coefficients of the audio signal obtained after MPEG compression, noise addition, low pass filtering, etc. On the other hand, the reduction in capacity with an increase of processing noise tends to be lower for transforms which are not used in compression methods, like DFT. While severe MPEG compression is certain to remove almost all high frequency components of DCT coefficients, it will not affect the high frequency DFT at the same extent. A signal decomposition with a low GTC is generally more immune to processing noise than decomposition with a high GTC and should predominantly be used in applications demanding robust watermarks. Therefore, signal decompositions with a high GTC, like the wavelet transform or DCT, are more suitable for high data rate steganography applications, where processing noise variance is low, because no intentional attacks are expected.

**Proposed High Data Rate Algorithm in Wavelet Domain:**

Using results from the information theory basis given above, we designed a novel audio steganography method with a high data rate of embedded information. The application scenario was to embed a MPEG compressed video sequence (high data rate requirement) into the host audio signal (mono signal, sampled at 44100 Hz). One example of the practical implementation of the algorithm was the hiding of the artist’s video clip in the artist’s audio track (CD format). If the watermarked music clips are, e.g., compressed to the mp3 format, the embedded video clip can not be extracted.

Therefore, no attacks or unintentional signal manipulations were expected, because it is the interest of the end user to obtain both multimedia files at the high quality data rate. The implemented method is a case of a fragile watermarking, as any distortion of the host audio signal leads to a severe quality loss of the embedded video clip.

Due to a low processing noise, the optimal selections of the signal decomposition algorithm are the wavelet decomposition and DCT. The wavelet domain is more suitable for frequency analysis because of its multi-resolutional properties that provide access both to the most significant parts and details of signal’s spectrum. Therefore, we are able to make easily the trade-off between the amount of the embedded information and perceptual distortion caused by information hiding, by handling sub-bands with different levels of power and perceptual significance.

Data hiding in the LSBs of the wavelet coefficients is practicable due to the near perfect reconstruction properties of the filterbank. The Discrete Wavelet Transform (DWT) decomposes the signal into low-pass and high pass components sub-sampled by two, whereas the inverse transform performs the reconstruction. We decided to make use of the simplest quadrature mirror filter-Haar filter. The Haar basis is obtained with a multiresolution of piecewise constant functions (Li, X. and H. Yu, 2000). The scaling function is equal to one. As the equivalent filter has two non-zero coefficients equal to $2^{-1/2}$ at $n = 0$ and $n = 1$ Haar wavelet is defined as:

$$
\psi(t) = \begin{cases} 
-1 & \text{if } 0 \leq t < 1/2; \\
1 & \text{if } 1/2 \leq t < 1; \\
0 & \text{otherwise}
\end{cases}
$$

(17)
The Haar wavelet has the shortest support among all orthogonal wavelets, and it is the only quadrature mirror filter that has a finite impulse response (Li, X. and H. Yu, 2000). FIR filters can be designed to be linear phase filters, which are important from the point of view of the perceptual transparency, as the linear phase filters delay the input signal, but do not distort its phase. In addition, the Haar filter is computationally simple to implement, as on most DSP processors, the FIR calculation can be done by looping a single instruction. This property gives the opportunity for real-time applications of the proposed algorithm. FIR filters have also desirable numeric properties. In practice, all DSP filters must be implemented using a finite precision arithmetic and a limited number of bits. As FIR filters have no feedback, they can usually be implemented using fewer bits, and the designer has fewer practical problems to solve related to non-ideal arithmetic, in comparison with IIR filters (Li, X. and H. Yu, 2000).

Signal decomposition into the low-pass and high-pass part of the spectrum is performed in five successive steps. After sub-band decomposition of 512 samples of host audio, using the Haar filter and decomposition depth of five steps, the algorithm produces 512 wavelet coefficients. All 512 wavelet coefficients are then scaled using the maximum value inside the given sub-band and converted to binary arrays in the two’s complement. A fixed number of the LSBs are thereupon replaced with bits of information that should be hidden inside the host audio. Coefficients are then converted and scaled back to the original order of magnitude and an inverse transformation is performed. The details of the decomposition of the signal and subsequent data embedding are given in Figure 5.

![Fig. 5: Signal decomposition prior to LSB embedding.](image)

The scheme was implemented using the integer wavelet transform (IWT) as well, in that case, there is no need for transforming coefficients (real values) into the integer format used for LSB embedding because IWT returns integers and would allow implementation on software with a less precise calculation than the Matlab© 16 bit floating point system.

The experimental results presented in the Paper IV are given for the case when wavelet coefficients of each of 32 sub-bands are modified in order to hide information. This is far from the optimal data hiding concept, as it has already been shown that the modification of the first four blocks of sub-band coefficients causes the largest degradation of perceptual quality of host audio (Boney, L., 1996; Gordy, J. and L.B. 2000; Laftsidis, C., 2003; Chou, J., 2001). Nevertheless, we tried to make a balanced comparison between the proposed algorithm and the time domain LSB coding, for the case when we use the same embedding method and add noise to the host audio in all parts of audio spectrum. Some other simple solutions that would add to the performance of the proposed data hiding algorithm because the randomizing of input data and removal of the DC bias caused by LSB replacement is not used during the tests for the same reason. During the subjective quality experiments, evaluation started with audio excerpts with three replaced LSBs for time domain and seven LSBs in wavelet domain because embedding to lower LSBs did not cause any noticeable perceptual distortion.

The subjective experiments showed that the sub-band information hiding scheme has a large advantage over the classic LSB algorithm. The wavelet domain algorithm produces stego objects perceptually hardly
discriminated from the original audio clip even when 8 LSBs of coefficients are modified, providing up to 5 bits per sample (220.5 kbps) higher data rate in comparison to time domain LSB algorithm.

The achieved bit rate of hidden information is clearly above the bit rate obtained by other developed audio steganography schemes (Cedric, T., 2000; Cvejic, N. and T. Seppanen, 2003). In addition, the scheme can easily be modified to be more robust against processing noise (achievable bit rate would be decreased though) and it was used as a basis for the development of a robust audio watermarking technique in wavelet domain (Cvejic, N. and T. Seppanen, 2003).

Conclusions:
Robust digital audio watermarking algorithms and high capacity steganography methods for audio are studied in this paper. The main results of this work are the development of novel audio watermarking algorithms, with the state-of-the-art performance and an acceptable increase in the computational complexity. The algorithms’ performance is validated in the presence of the standard watermarking attacks. The main point of the “magic triangle” concept is that if the perceptual transparency parameter is fixed, the design of a watermark system cannot obtain a high robustness and watermark data rate at the same time.

The details and experimental results for the modified time domain LSB steganography algorithm were discussed. The results of subjective tests showed that the perceptual quality of watermarked audio, when embedding is done by the proposed algorithm, is higher in comparison with the standard LSB embedding. The tests confirmed that the described algorithm succeeds in increasing the bit rate of the hidden data for one third without affecting the perceptual transparency of the resulting audio signal. However, the simple LSB coding method in time domain is able to inaudibly embed only 3-4 bits per sample, which is far from a theoretically achievable rate, mostly due to a poor shaping of the noise introduced by embedding and operation in time domain. Therefore, perceptual entropy and information theoretic assessment of the achievable data rates of a data hiding channel was necessary to develop a scheme that could obtain higher data rates.

A high bit rate algorithm in wavelet domain was developed based on these findings. The wavelet domain was chosen for data hiding due to its low processing noise and suitability for frequency analysis, because of its multi-resolutional properties that provide access both to the most significant parts and details of signal’s spectrum. The experiments showed that the wavelet information hiding scheme has a large advantage over the time domain LSB algorithm. The wavelet domain algorithm produces stego objects perceptually hardly discriminated from the original audio clip even when 8 LSBs of coefficients are modified, providing up to 5 bits per sample higher data rate in comparison with the time domain LSB algorithm.

A spread spectrum audio watermarking algorithm in time domain is presented. The overall watermark detection robustness of the algorithm is comparable to other state-of-the-art algorithms, specifically in the presence of mp3 compression, re-sampling and low pass filtering. On the other hand, the algorithm uses computationally low demanding embedding and detection methods and a simple perceptual model for describing two masking properties of the HAS. One of the malicious attacks on this scheme is the de-synchronization of the correlation calculation by time-scale modifications, such as the stretching of the audio sequence or insertion or deletion of samples. In that case, the watermark detection scheme does not properly determine the value of the embedded watermark, resulting in a high increase of the bit error rate.

A re-synchronization algorithm that is able to provide correct watermark detection even in the presence of these attacks, while maintaining a perceptual transparency by a perceptual noise shaping is presented subsequently. The consequence of an improved watermark decoding is a decreased bit rate of the embedded watermark; however the bit rate is still within an acceptable range for most copyright applications.

The possibility of improving the robustness of watermark detection and increasing the resistance to attacks was studied. An audio de-correlation algorithm for the spread-spectrum watermarking that uses least squares Savitzky-Golay smoothing filters is proposed. The test results showed a significant improvement in the detection performance of the described method, compared to the standard watermark detection, especially if a watermarked audio sequence is attacked with mp3 compression or low pass filtering attacks.

In order to further improve the detection robustness and decrease the bit error rate, channel coding was employed, because it has a property to reduce BER for a given watermark bit rate in comparison with the regular detection or equivalently increase an available watermark bit rate for a given BER. The simulations showed that a channel coding maintains a reliable watermark bit rate for a fixed BER, even after severe attacks.

The particular watermark channel model that was studied was a watermark channel model in the presence of MPEG compression. We showed that a far more appropriate model for the watermark channel in the presence of mp3 coding is the Rayleigh frequency-selective fading model, because it describes more precisely
the distortions that appear. The experimental results suggest that the noise introduced by mp3 compression can hardly be modeled as AWGN and that BER curves obtained by the Rayleigh fading channel model have steepness and values more close to the experimentally derived ones. The results confirmed that a far better watermark channel modeling is obtained by the proposed model than with the usual AWGN watermark channel model.

Using the available theoretical background, we developed a novel audio watermarking scheme that uses the attack characterization in order to obtain a high robustness against standard watermark attacks. The overall algorithm obtained high detection robustness, while decreasing the computational complexity and increasing the perceptual transparency of the watermarked signal. At the end, it was shown that the attack characterization algorithm that was proposed can be successfully used in other schemes as well. The detection performance of the system using an attack characterization and the ISS modulation is significantly higher compared to the method using the standard SS modulation uses the statistical properties of ISS modulation while maintaining a blind detection during the watermark extraction.

REFERENCES


