Survey on Different Level of Audio Watermarking Techniques

Shweta Sharma
Student, CS Department
Rajasthan Collage of
Engineering for Women,
Jaipur, India

Jitendra Rajpurohit
Student, CS Department
Poornima College of
Engineering, Jaipur, India

Sunil Dhankar
HOD, CS Department
Rajasthan Collage of
Engineering for Women,
Jaipur, India

ABSTRACT

Audio Watermarking is useful technique for audio systems. This technique can work on different domains like frequency and time. By using the different scheme of watermarking at different levels of audio, it can be secure from many types of attacks. This paper shows some techniques which can be used to secure the audio system from attacks and survey on various transformation techniques for embedding or extracting watermark.

General Terms

Audio Watermarking, Attacks, Security, Compression, Embed and Extraction.

Keywords

Discrete Cosine Transformation(DCT), Inverse Discrete Cosine Transformation(IDCT), Chaos, Geometric invariants.

1. INTRODUCTION

Audio watermarking is challenging than an image watermarking technique due to wider dynamic range of the HAS in comparison with human visual system (HVS) [1]. Human ear can perceive the power range greater than 109:1 and range frequencies of 103:1 [4]. In addition, human ear can hear the low ambient Gaussian noise in the order of 70dB [4]. However, there are many other useful features such as the louder sounds mask the corresponding slow sounds. This feature can be used to embed additional information like watermark. Further, HAS is insensitive to a constant relative phase shift in a stationary audio signal and some spectral distortions are interpreted as natural, perceptually non-annoying ones [2]. Two properties of the HAS dominantly used in watermarking algorithms are frequency (simultaneous) masking and temporal masking [3].

Frequency masking: Frequency (simultaneous) masking is a frequency domain observable fact where low levels signal (the maskee) can be made inaudible (masked) by a simultaneously appearing stronger signal (the masker), if the masker and maskee are close enough to each other in frequency [5]. A masking threshold can be found and is the level below which the audio signal is not audible. Thus, frequency domain is a good region to check for the possible areas that have imperceptibility.

Temporal masking: In frequency masking, two phenomena of the HAS in the time domain also play an important role in human auditory perception. Those are pre-masking and post-masking in time [5]. Temporal masking is used in those applications where the robustness is not of primary consideration.

2. PROBLEMS AND ATTACKS ON AUDIO SIGNALS

As discussed earlier the main requirements of an efficient watermarking technique are the robustness and inaudibility. There is a substitution between these two requirements; however, by testing the algorithm with the signal processing attacks the gap can be made minimal. Every application has its explicit requirements and provides an option to choose high robustness compensating with the quality of the signal and vice-versa. Without any transformations and attacks every watermarking technique performs efficiently. Many common types of processes for audio signal are used when transmitted through a medium [13].

Dynamics: The amplitude variation and reduction provide the dynamics of the attacks. Limiting, extension and compressions are some sort of more complex applications which are the non-linear modifications. Some of these types of attacks are re-quantization [12].

Filtering: Filtering is a common practice used to amplify or attenuate some part of any signal. The basic low pass and high pass filters can be used to attain these types of attacks.

Ambience: In some situations the audio signal gets deferred or there are situations where in people record signal from a source and claim that the track is theirs. Those situations can be simulated in a room, which has great importance to check the performance of an audio signal.

Conversion and lossy compression: Audio generation is done at a particular sampling frequency and bit rate; however, the created audio track will undergo so many different types of compression and conversion techniques. Some of the most common compression techniques are audio compression techniques based on psychoacoustic effect (MPEG and Advanced Audio Codec (AAC)). In addition to that, it is common process that the original audio signal will change its sampling frequencies like from 128Kbps to 64Kpbs or 48 Kbps.

Noise: Whenever the signal is transmitted there is always noise is present in the signal. Hence, watermarking algorithm has to make the technique robust against the noise attacks. It is commonly used to check the algorithm for this type of noise attacks by adding the host signal by an additive white Gaussian noise (AWGN) to check its robustness.

Time stretch and pitch shift: These attacks change either the length of the signal without changing its pitch and vice versa. These are some de-synchronization attacks which are quite common in the data transmission. Jittering is one type of such attack.

3. AUDIO WATERMARKING TECHNIQUES – OVERVIEW

An audio watermarking technique can be classified into two groups based on the domain of operation. One is **time domain** technique and the other is **transformation based method**. The time domain techniques include methods where the embedding is performed without any transformation. Watermarking is employed on the original samples of the audio signal. Time domain watermarking technique is the example of **least significant bit** (LSB) method. In LSB method the watermark is embedded into the least significant bits of the host signal. As against these techniques, the transformation based watermarking methods perform watermarking in the transformation domain. Few transformation techniques that can be used are discrete cosine transform and discrete wavelet transform. In transformation based approaches the embedding is done on the samples of the host signal after they are transformed. Using of transformation based techniques provides additional information about the signal [11].

In all-purpose, the time domain techniques provide least robustness as a simple low pass filtering can remove the watermark [1]. Hence time domain techniques are not advisable for the applications such as copyright protection and airline traffic monitoring; however, it can be used in applications like proving ownership and medical applications.

Watermarking techniques can be distinguished as visible or we can say it as non-blind watermarking and blind watermarking. In the following, we present typical watermarking strategies such as LSB coding, spread spectrum technique, patchwork technique, and quantization index modulation (QIM).

3.1 LSB Coding

This technique is one of the common techniques in use in signal processing applications. It is based on the replacement of the LSB of the carrier signal with the bit pattern from the watermark noise [16]. The robustness depends on the number of bits that are being replaced in the host signal. This type of technique is usually used in image watermarking because each pixel is represented as an integer hence it will be easy to replace the bits. The audio signal has real values as samples, if transformed to an integer will degrade the quality of the signal to a great extent (see Fig 1).

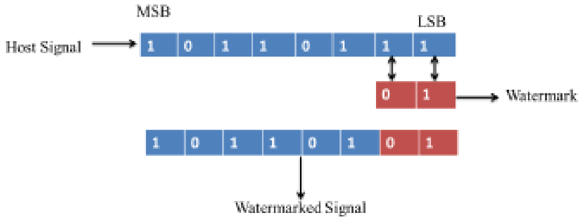

Fig 1: LSB embedding

3.2 Spread Spectrum Technique

These techniques are derived from the concepts used in spread spectrum communication [15]. The basic approach is that a narrow band signal is transmitted over the large bandwidth signal which makes them undetectable as the energy of the signal is overlapped. In the similar way the watermark is spread over multiple frequency bins so that the energy in any one bin is very small and certainly undetectable [14].

In spread spectrum technique, the original signal is first transformed to another domain using domain transformation techniques [13]. The embedding technique can use any type of approach for example quantization. Zhou *et al.* proposed an algorithm embedding watermark in 0th DCT coefficient and 4th DCT coefficients which are obtained by applying DCT on the original signal [5]. Both embedding and extraction procedure can be interpreted using Figure 2. The original signal is transformed into frequency domain using DCT. Then watermark is embedded to the sample values in that domain. Reverse procedure is followed to obtain the watermarked signal (see Fig 2).

Embedded signal will go through some attacks, thus, noise is added to the signal. To extract the watermark the attacked signal is fed through extraction procedure. The procedure for extractions follows the same steps as that in embedding procedure as shown in Figure 2. The extraction process involves taking the attacked signal and applying DCT, framing the obtained components. And they obtained frames are used to obtain the watermark. Care is taken to replicate the procedure used for embedding process.

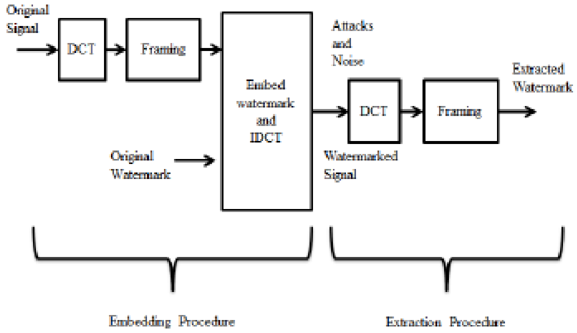

Fig 2: Example for spread spectrum technique

3.3 Patchwork Technique

The data to be watermarked is divided into two distinct subsets. One feature of the data is selected and customized in opposite directions in both subsets [13]. For an example let the original signal is divided into two parts *A* and *B*, then the part *A* is increased by a fraction *Δ* and the part *B* is decreased by some amount *Δ*. The samples separation is the secret key which is termed as watermarking key. Detection of watermark is done by following the statistical properties of the audio signal. Let N*A* and N*B* indicate the size(s) of the individual *A* and *B* parts and *Δ* be the total of the change made to the host signal. Suppose that *a*[*i*] and *b*[*i*] represent the sample values at **i**$^{\text{th}}$ position in blocks *A* and *B*. The difference of the sample values can be written as [23]:

$$S = \frac{1}{N_A}\sum_{N_A} a[i] - \frac{1}{N_B}\sum_{N_B} b[i]$$

$$= \frac{1}{N}\sum_{N}(a[i]-b[i]); \quad N_A = N_B = N.$$

The expectation of the difference is used to extract the watermark which is expressed as follows [4].

$$E\{S\} = \begin{cases} 2\Delta; \text{for watermarked data} \\ 0; \text{for unwatermarked data} \end{cases}.$$

4. TRANSFORMATION TECHNIQUES

Here we talk about the background about discrete cosine transform (DCT) and discrete wavelet transforms (DWT). The contents also presents different DWT types such as orthogonal, bi-orthogonal and frame based filters.

4.1 Discrete Cosine Transform

The discrete cosine transform is a technique for converting a signal into elementary frequency components [17]. The DCT can be employed on both one-dimensional and two-dimensional signals like audio and image, respectively. The discrete cosine transform is the spectral transformation, which has the properties of Discrete Fourier Transformation [17]. DCT uses only cosine functions of various wave numbers as basic functions and operates on real-valued signals and spectral coefficients. DCT of a 1-dimensional (1-d) sequence and the reconstruction of original signal from its DCT coefficients termed as **inverse discrete cosine transform** (IDCT) can be computed using equations [17]. In the following, $f_{dct}(x)$ is original sequence while $C_{dct}(u)$ denotes the DCT coefficients of the sequence.

$$C_{dct}(u) = \alpha(u)\sum_{x=1}^{N_{1t}-1} f_{dct}(x)\cos\left[\frac{\pi(2x+1)u}{2N_{1t}}\right],\ \mathit{for}\, u = 0,1,2,...,N_{1t}-1$$

$$f_{dct}(x,y) = \sum_{u=1}^{N_{1t-1}} \alpha(u)\,C_{dct}(u)\cos\left[\frac{\pi(2x+1)u}{2N_{1t}}\right],\ \mathit{for}\ x = 0,1,2,...,N_{1t}-1$$

$$\mathit{where}\ \alpha(u) = \begin{cases} \sqrt{\dfrac{1}{N_{1t}}} & \text{for}\, u = 0 \\ \sqrt{\dfrac{2}{N_{1t}}} & \text{for}\, u \neq 0 \end{cases} .$$

From the equation for $C_{dct}(u)$ it can be inferred that for u = 0, the component is the average of the signal also termed as dc coefficient in literature [38]. And all the other transformation coefficients are called as ac coefficients. Some of the important applications of DCT are image compression and signal compression.

The most useful applications of two-dimensional (2-d) DCT are the image compression and encryption [17]. The 1-d DCT equations, discussed above, can be used to find the 2-d DCT by considering every row as an individual 1-d signal. Thus, DCT coefficients of an M×N two dimensional signals Cdct2(u, v) and their reconstruction f_{dct2}(x, y)can be calculated by the equations below.

$$C_{dct2}(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{M_{2t-1}}\sum_{y=0}^{N_{2t-1}} f_{dct2}(x,y)\cos\left[\frac{\pi(2x+1)u}{2M_{2t}}\right]\cos\left[\frac{\pi(2y+1)v}{2N_{2t}}\right]$$

$$f_{dct2}(x,y) = \sum_{x=0}^{M_{2t-1}}\sum_{y=0}^{N_{2t-1}} \alpha(u)\,\alpha(v)\,C_{dct2}(u,v)\cos\left[\frac{\pi(2x+1)u}{2M_{2t}}\right]\cos\left[\frac{\pi(2y+1)v}{2N_{2t}}\right]$$

$$\mathit{where}\ u\ \&\ x = 0,1,2,...,M_{2t-1}\quad \mathit{and}\quad \mathit{where}\ v\ \&\ y = 0,1,2,...,N_{2t-1}$$

$$\alpha(u) = \begin{cases} \sqrt{\dfrac{1}{N_{2t}}} & \text{for}\, u = 0 \\ \sqrt{\dfrac{2}{N_{2t}}} & \text{for}\, u \neq 0 \end{cases} \quad \& \quad \alpha(v) = \begin{cases} \sqrt{\dfrac{1}{N_{2t}}} & \text{for}\, v = 0 \\ \sqrt{\dfrac{2}{N_{2t}}} & \text{for}\, v \neq 0 \end{cases} .$$

4.2 Spread spectrum audio watermarking in time domain

One of the first audio watermarking algorithms is a time-domain spread spectrum algorithm. It embeds a spread-spectrum-based watermark into an uncompressed, raw audio by slightly modifying the values of samples of the host audio in time domain. The main motivation was the development of an algorithm with a low computational complexity and with an embedding and extraction of watermarks in time domain. One of the most robust methods already developed for audio watermarking was a time domain algorithm. It would definitely be hard to prove mathematically that watermarking in time domain gives smaller computational complexity in comparison with other, non-temporal algorithms because it is hard to compare complexity with each developed watermarking scheme. However, time domain algorithms have at least a lower implementation complexity and a smaller number of blocks in embedding and extraction algorithms

5. COMMUNICATIONS MODEL OF THE WATERMARKING SYSTEMS

In order to describe the link between watermarking and standard data communications, the traditional model of a data communications system is often used to model watermarking systems. The basic components of a data communications system, related to the watermarking process, are highlighted. One of the most important parts of the communications models of the watermarking systems is the communications channel, because a number of classes of the communications channels have been used as a model for distortions imposed by watermarking attacks [18]. The other important issue isthe security of the embedded watermark bits, because the design of a watermark system has to take into account access that an adversary can have to that channel.

5.1 Components of the communications model

The main elements of the traditional data communications model are depicted in Figure 3. The main objective is to transmit a message m across a communications channel. The channel encoder usually encodes this message in order to prepare it for transmission over the channel. The channel encoder is a function that maps each possible message into a code word drawn from a set of signal that can be transmitted over the communications channel. The code word mapped by the channel encoder is denoted as x. It is common as we deal with digital data and signals that the encoder consists of a source coder and a modulator. The source coder removes the redundancy from the input message and maps a message into a sequence of symbols drawn from some alphabet [19]. The duty of the modulator is to convert a sequence of symbols from the source coder into a signal suitable for transmission through a physical communications channel. It can use different modulation techniques such as amplitude, phase or frequency modulation.

The definite form of the channel encoder's output depends on the type of the transmission channel used in a particular model, but it is usually described as a sequence of real values, quantized to some arbitrary precision. In addition, we assume that the range of values of the channel encoder is limited in some way, usually by a power or amplitude constraint [20].

The signal x is then sent over the communications channel, which is assumed to be noisy. The result of the existence of noise is that the received signal, conventionally denoted as y, is generally different from x. The extent of the change depends of the level of the noise present in the channel and is modeled here as additive noise. In other words, the transmission channel is modeled as adding a random noise n to the encoder's output x. At the receiver part of the system, the received signal, y, is forwarded, as the input signal, to the channel decoder which inverts the encoding process and attempts to correct for errors caused by the presence of noise. This is a function that maps transmitted signals into messages m_r. The decoding process is typically a many-to-one function, so that correct decoding is possible even using noisy coded words [21, 22]. If the channel code is well matched to a given channel model, the probability that the decoded message contains an error is negligibly small.

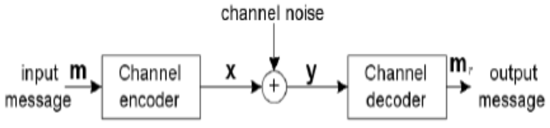

Fig 3: Standard model of a communications system

5.2 Models of communications channels

During the modeling of a communications system given in Figure 1, the parameters of the transmission channel are usually predetermined. That is, the function that is used for the modeling of the transmission channel cannot be modified during the transmission. The channel is generally characterized using a conditional probability distribution $P_{Y|X}(y)$, which gives the probability of obtaining y as the received signal if signal x was transmitted over the transmission channel.

Diverse communications channels can be classified in relation to the type of the noise function they apply to the signal and the way the distortion is introduced. The model from the Figure 1 is, as already mentioned above, an additive noise channel in which signals are distorted by the addition of noise signal n.

$$y = x + n$$

The noise signal is usually modeled as independent of the signal x. The simplest and most important channel for analysis is a Gaussian channel where each element of the noise signal, *n(i),* is drawn independently from a normal distribution with zero mean and a variance σ^2_n. The variance models the level of distortion of the signal introduced by channel noise and zero mean distribution means that channel noise does not have an impact on the DC component of the transmitted signal. Despite being simple, this model is the most frequently used one in the watermark literature and it was extensively used in our papers as well.

However, several non-additive communications channel models are also important. One of the frequently used models is the fading channel model [6] which causes the variation of the transmitted signal's power during the transmission. Generally, this variation can be modeled as a scaling of the signal

$$y = v(t)x$$

where 0 < v(t) < 1 is an unknown parameter that vary slowly during the transmission or with each use of the channel. Such a channel might also include an additive noise component, rendering

$$y = v(t)x + n$$

There are only a small number of watermark papers that use a fading channel model for the description of the channel noise.

5.3 Secure data communications

An important issue in watermarking is the security of the embedded watermark bits because the design of a watermark system has to take into account access that an adversary can have to the communications channel. In particular, we are interested in applications that demand security against passive and active adversaries. In the case of passive attacks, an adversary monitors the transmission channel and attempts to illegally read the message.

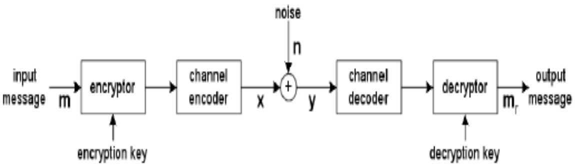

Fig 4: A model of a communications channel with encryption

In the active attack case, the adversary actively tries either to disable communication or transmit unauthorized messages. There are two main methods of defense against attacks, first, cryptography and second, spread spectrum communications. Prior to communication, cryptography is used to encrypt a message using a secret key and after that the encrypted message is transmitted. On the receiver side, the encrypted message is received and then decrypted using the same or a related key to reveal the message. The block scheme is given in Figure 4. Cryptography introduces two advantages in a data communications system. The first is to prevent passive attacks in the form of an unauthorized reading of the message and the second is to prevent active attacks in the form of illicit writing. However, cryptography does not necessary prevent the adversary from knowing that a message is being transmitted. In addition, cryptography is helpless if adversary intents to distort or remove a message before it is delivered to receiver.

Signal congestion (the premeditated effort by an adversary to inhibit communication between transmitter and receiver) was a great problem for military communications and hasled to the development of the spread spectrum communication. In those systems, the modulation is performed according to a secret code that spreads the signal across a wider bandwidth than is regularly required. The code can be modeled as a form of the key used in the channel coder and decoder, as depicted in Figure 5. One of the examples of the spread spectrum communications is the frequency hopping method, one of the earliest and simplest spread spectrum techniques. In a frequency-hopping system, the transmitter broadcasts a message by first transmitting a part of the message bit stream on one frequency, the next fraction of the bit stream on the another frequency, and so on. A secret key that is known at the receiver as well as on the transmitter side controls the order of frequencies used for frequency hopping. Without a key, an adversary could monitor the transmission. The disruption of the transmission is also very difficult, because it could be done only by introducing noise at all possible frequencies, which would require too much power.

The cryptography and SS communications are complementary. The SS guarantees the delivery of signals, while the cryptography guarantees the secrecy of messages. Thus, it is common that these two technologies are combined in watermarking applications.

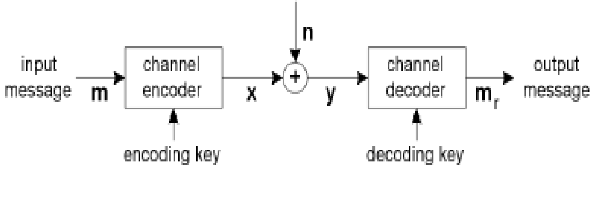

Fig 5: A model of a communications channel using spread spectrum key-based coding

5.4 Communication-based models of watermarking

The essential process in each watermarking scheme can be modeled as a form of communication where a message is transmitted from watermark embedder to the watermark receiver [22]. Therefore, it is natural to place watermarking into the framework of the traditional communications system. In Figures 4 and 5, two ways of mapping a watermarking system into communications framework are given. Figure 4 shows a watermarking system with an informed detection and Figure 5 a system that uses a blind detector.

In the watermarking-communications mapping, the process of watermarking is seen as a transmission channel through with the watermark message is being sent, with the host signal being a part of that channel. The embedding method consists of two basic steps, regardless of the detection method used (informed or blind detection).

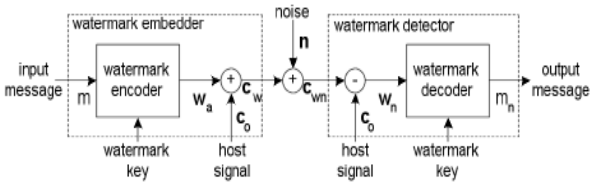

Fig 6: Watermarking system with informed detection-equivalent communications model

In the first step, the message to be transmitted is mapped into an added pattern, w_a, of the same type and dimension of the host signal c_o (two dimensional patterns for images and videos and one dimensional patterns for audio). The mapping is usually performed using a secret watermark key [8]. The calculation of the optimal added pattern w_a is typically performed in several steps, and it starts with one or more reference patterns w_{r0}, w_{r1}… which are predefined patterns, dependent on a watermark key. The reference patterns are subsequently combined to construct a pattern that encodes the message, which is referred to as a message pattern. The message pattern is the perceptually weighted in order obtain the added pattern w_a. After that, w_a is added to the host signal c_o, to construct the watermarked signal c_w. If the watermark embedding process does not use information about the host signal, it is called the blind watermark embedding; otherwise the process is referred to as an informed watermark embedding. After the added pattern is embedded, the watermarked work is usually distorted during watermark attacks. We model the distortions of the watermarked signal as added noise, as in the data communications model. The types of attacks may include compression and decompression, broadcast over analogue channels, low pass filtering, dynamic compression, etc. However, the additive noise modeling is a simplified representation of the introduced distortions because all these types of distortions are non-stationary signal-adaptive processes.

If an informed watermark detector is used, the watermark detection is performed in two steps. In the first step, the un-watermarked host signal may be subtracted from the received signal c_{wn} in order to obtain a received noisy added watermark pattern w_n. It is subsequently decoded by a watermark decoder, using the same watermark key used during the embedding process. Because the addition of the host signal in the embedder is exactly canceled by its subtraction in the detector, the only difference between w_a and w_n is caused by the added channel noise. Therefore, the addition of the host signal can be neglected, making watermark embedding, channel noise addition and watermark extraction equivalent to the data communications system given in Figure 6. In more advanced, informed detection systems, the entire un-watermarked host signal is not needed [7]. Instead, some function of c_o, usually a data reducing function, is used by the watermark detector to nullify "noise" effects represented by the addition the host signal in the embedder.

In a blind watermark detector, the un-watermarked host signal is unknown, and cannot be removed before a watermark extraction. Under these conditions, the analogy with Figure 6 can be made, where the added watermark is corrupted by the combination of impacts of the cover work and the noise signal. The received watermarked signal c_{wn}, is now viewed as a corrupted version of the added pattern w_a and the entire watermarked detector is viewed as the channel decoder.

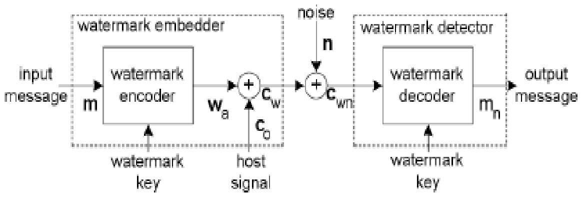

Fig 7: A watermarking system with blind detection-equivalent communications model

In application that require robustness of the embedded watermark, e.g. a transaction tracking and copy control, the likelihood that the embedded message is identical to the extracted one, must be maximized, like in the traditional data communications systems. However, in the authentication watermarking systems, the goal is not to communicate a message, but to discover whether and how a host signal has been modified since watermark was embedded. Therefore, models from Figures 4 and 5 are not typically used to describe authentication systems.

5.5. Spread spectrum watermarking algorithm in time domain

The basic audio watermarking algorithm that we developed is a time domain spread spectrum algorithm. It embeds a SS-based watermark into uncompressed, raw audio by slightly modifying the values of samples of the host audio in time domain. The procedure uses the virtues of the spread-spectrum communications given above, as well as temporal masking property of the HAS and the basic information about the spectrum of the host audio. Figure 8 gives a general overview of the proposed watermark embedding algorithm. A simple trade-off between the watermark data rate and the robustness of the embedded watermark is possible, because the m-sequence length is decreased, the algorithm is able to embed a higher data rate watermark, but with less robustness against common watermark attacks, because low pass filtering or MPEG compression. For example, with the spreading sequence block length of 1023 samples, a watermark data rate of 43.10 bps is obtained.

The host audio sequence is initially analyzed in time domain, in order to determine the just noticeable distortion threshold, using the time domain masking property of the HAS. The goal is to place the watermark inside the host audio without causing perceptual quality degradation in the process, while maximizing the amplitude values of the watermark sequence samples in order to increase algorithm's robustness in the presence of attacks [9]. In the next step, a simple frequency analysis of the host audio is implemented as a common zero crossings counter in the basic block interval. The counting process derives information of the presence of the higher frequencies within the spectrum. If the presence of high frequency content is emphasized in a block, the power of the embedded watermark sequence can be greater as well, without affecting the overall subjective quality of the watermarked audio. The embedding algorithm obtains coefficient b(n) from the frequency analysis block, with higher values in the blocks in which the host audio has a significant high-frequency content. At the output of the watermark embedding process, the perceptually weighted spreading sequence is added to the host audio sequence resulting in:

$$Y^*(n) = x(n) + a(n)b(n)w(n)$$

where a(n) and b(n) are coefficients obtained from temporal and frequency analysis blocks, respectively, x(n) is the host audio sequence and w(n) is the watermark sequence spread in time.

Figure 9 gives an overview of the watermark detection algorithm. The corner stone of the detection process is, as in all spread spectrum systems, a cross-correlation calculation, in this case mean-removed cross-correlation between the watermarked audio signal and the equalized m-sequence. Before the watermarked signal is segmented into blocks and cross-correlation with the m-sequence is calculated, the detection algorithm filters it with the equalization filter. The equalization filter is a high pass filter that filters out strong low pass components, increase correlation value and enhance detection results. The drawback is that it is a fixed coefficient filter, not adaptive to the local properties of the watermarked audio. The values from the correlation calculation block are forwarded to the detection/sampling block, which samples the output of the correlated in order to obtain values for the threshold/decision block.

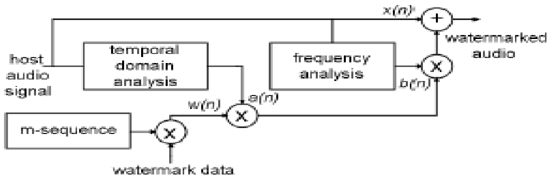

Fig 8: A proposed watermark embedding scheme

The threshold/decision block provides the majority vote decision regarding the value of the embedded bit, depending on the sign of the correlation value.

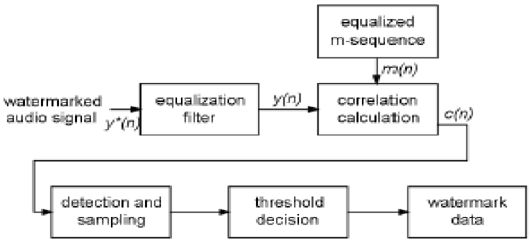

Fig 9: A watermark detection scheme

The correlation method demands alignment between the blocks of the equalized m-sequence and watermarked audio blocks for reliable watermark detection. One of the malicious attacks on this scheme is the de-synchronization of the correlation calculation procedure by time-scale modifications, such as the stretching of the audio sequence (without affecting the pitch) or the insertion/deletion of samples. In that case, the watermark detection system does not properly conclude the value of the embedded watermark, resulting in a high increase of the bit error rate. A resynchronization algorithm is able to provide a low bit error rate during the watermark decoding even in the presence of these attacks.

The algorithm obtained a high detection performance [6, 7, 8, and 9] in the cases of band equalization, all-pass filtering, amplitude compression, echo addition and noise addition attacks. After re sampling and mp3 compression attacks, the bit error rate is higher than in the case of other attacks, but the detection robustness was still equal to the other state-of-the-art algorithms. The reason for a poorer detection performance in the presence of a down sampling attack is that half of the spreading sequence power is lost after down sampling and strong low frequency components of the host audio remain unaffected by the attack. On the other hand, mp3 compression crops the high frequency spectrum of the watermarked audio and smoothens out audio waveform, destroying small modifications introduced by the watermark embedding algorithm.

The overall watermark detection robustness of the algorithm is comparable with other state-of-the art algorithms [9, 10], specifically in the presence of the most malicious attacks for SS watermarking algorithms (mp3 compression, re sampling, 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. Thus, a successful compromise between the computational complexity and the detection performance of the algorithm is obtained.

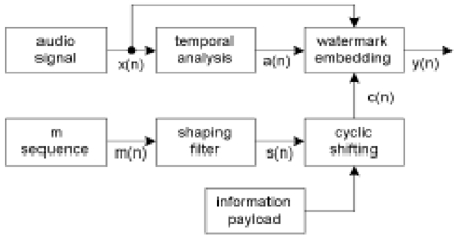

Fig 10: The improved watermark embedding algorithm

5.6 Increasing detection robustness with perceptual weighting and redundant embedding

After the development of the basic audio watermarking algorithm for digital audio, described, we improved the performance of the given method by utilizing more of the HAS properties and using a redundant embedding during watermark insertion.

The basic idea is that the spectrum of the m-sequence is shaped in accordance to the HAS in order to make the watermark even more imperceptible. An integration function is added jointly with a synchronization scheme in the receiver to obtain a higher robustness against attacks. For handling time scaling attacks, a multiple chip embedding is used. With these enhancements, a considerably lower demand for computational power is attained and better time-scaling resistance than with our earlier algorithm.

Figure 10 gives a general overview of the watermark embedding algorithm. Prior to further processing, the m-sequence is filtered in order to adjust it to masking thresholds of the HAS in the frequency domain. The frequency characteristic of the filter is the approximation of the threshold in quiet curve of the HAS. Despite the simplicity of the shaping process of the m-sequence in frequency domain, the result is an in audible watermark as the largest amounts of the shaped watermark's power are concentrated in the frequency sub-bands with a lower HAS sensitivity. A significant number of computational operations needed for the frequency analysis of audio, which have to be run in order to derive global masking thresholds in a predefined time window, are skipped, making this scheme appreciably faster. Although standard frequency analyses have more accurate data about the audio spectrum, the simulation tests done with selected audio clips show high level of similarity with the frequency masking thresholds derived from the masking model defined in ISO-MPEG Audio Psychoacoustic Model.

A cyclic shifted version c(n) of the shaped sequence s(n) is used to achieve a multi-bit payload. Every possible shift is associated with different information content and watermark bit rate is directly proportional to the length of the m-sequence. Therefore, a simple trade-off between the embedded data size and robustness of the algorithm is obtained. The host audio sequence is also analyzed in the time domain, where a minimum or a maximum is determined in the block of audio signal that has the length of 7.6 ms. As the result of this analysis, the watermark samples are weighted by the coefficient a(n) in order to be adjusted to psycho-acoustic perceptual thresholds in time domain.

Therefore, the watermark signal is embedded into a host audio using three time-aligned processes. In the first stage, the m-sequence has been filtered with the shaping filter, where a colored-noise sequence s(n) is the output. Samples of the s(n) sequence are then cyclically shifted, where the shift value is dependent of the input information payload. At the output of the watermark embedding scheme, the shifted version of s(n), sequence c(n) is being weighted and added to the original audio signal:

$$y(n) = x(n) + a(n)c(n)$$

where x(n) denotes input audio signal and a(n) are coefficients from the temporal analysis block. The addition of the c(n) sequence in the embedding process is done redundantly in order to make the system resistant to time scaling attacks that tend to desynchronize the extraction process.

The detection process is again performed using the mean removed cross-correlation between the watermarked audio signal and the equalized m-sequence. Before the start of the integration process, which determines the peak and the embedded bit, the block power normalization part normalizes the energies of the output blocks from correlation calculations.

The integration block sum the normalized output block from correlation detection and determines the peak and its position. The detection reliability depends strongly on the number of accumulated frames. In general, the trade-off is made between the time of integration and the amount of hidden data.

The extraction scheme uses redundancy in the watermark chip pattern, similar to the one described in [5]. The basic idea is to spread each chip of the shaped m-sequence onto R consecutive samples of watermarked audio. It has been proved that the correlation is correctly calculated even if a linear shift of bR=2c samples across the temporal or frequency domain is induced. However, there is a trade-off between the robustness of the algorithm and computational complexity, which is significantly increased by performing multiple correlation tests.The test results showed that if attacks are performed by mp3 and AAC compression and time-scaling, the bit error rate is higher than in the case of other attacks, but the detection performance is still within the range of the state-of-the-art algorithms [9, 10]. The reason for poorer extraction capabilities after mp3 and AAC coding is that these compression techniques crop high frequency spectrum of the watermarked audio, where most of the watermark energy is situated. Time scaling is one of the most malicious attacks on the block-based watermarking algorithms, but the redundant spread sequence embedding solution reduced decoding BER in the presence of these attacks to an acceptable level.

The penalty for an improved watermark decoding is a decreased bit rate of the embedded watermark. However, the bit rate is still within an acceptable range for copyright applications. Now we will see other important strategies in this field.

5.7 Improved Audio Watermarking Using DWT-SVD

Digital audio watermarking involves the concealment of data within a discrete audio file. Applications for this technology are numerous. Intellectual property protection is currently the main driving force behind research in this area. In this paper they present an efficient audio watermarking algorithm in the frequency domain by embedding an inaudible audio water mark. Comparison of two different algorithms i.e. Discrete Cosine Transform (DCT)-Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT)-SVD is presented here. After experimentation results show that the watermarked audio has good imperceptibility and is robust against different kinds of attacks, such as noise adding, re-sampling, cropping.

It is verified that the DWT-SVD technique is robust for most of the attacks rather than the DCT-SVD. By means of combining the two transforms DWT-DCT along with SVD, inaudibility and different levels of robustness can also be achieved.

5.8 Robust Audio Watermarking Based Non-Uniform DCT

By targeting at the security attack in digital audio watermarking, based non-uniform DCT audio watermarking algorithm was projected. The watermark of algorithm was generated by chaos by means of key. The watermark was embedded in the imply of non-uniform DCT (NDCT) coefficients by quantifying the coefficients mean. NDCT frequency sample positions were generated by chaos which was restricted by key and security of watermark structure is better. Simulating experiments outcome demonstrated: the proposed algorithm was tremendously robust to MP3 lossy compression, filtering, re-sample, quantization etc. signal process manipulations, the algorithm become accustomed to copyright protect.

This paper proposed an algorithm based on NDCT, watermark is generated by Henon chaos, and Logistic Map restricted NDCT matrix, the algorithm is security. Experiment results demonstrated the algorithm is robust and imperceptible.

5.9 Geometric Invariant Audio Watermarking Based on an LCM Feature

The expansion of a geometric invariant audio watermarking scheme without degrading acoustical quality is challenging work. This paper proposes a multi-bit spread-spectrum audio watermarking method based on a geometric invariant log coordinate mapping (LCM) feature. The LCM feature is incredibly robust to audio geometric distortions. The watermark is embedded in the LCM feature, but it is really embedded in the Fourier coefficients which are mapped to the characteristic via LCM, so the embedding is actually performed in the DFT domain without interpolation, thus eliminating completely the severe distortion resulted from the non-uniform interpolation mapping. The watermarked audio achieves high auditory quality in both objective and subjective quality assessments. A mixed correlation between the LCM feature and a key-generated PN tracking sequence is proposed to align the log-coordinate mapping, thus synchronizing the watermark efficiently with only one FFT and one IFFT. Both the theoretical analysis and experimental results show that the proposed audio watermarking scheme is not only resilient against common signal processing operations, including low-pass filtering, MP3 recompression, echo addition, volume change and normalization.

They have developed an audio watermark embedding strategy, which is actually performed in DFT domain without interpolation, to avoid completely the severe distortion due to the non-uniform interpolation mapping while achieving the effect of embedding in LCM domain. The watermark is embedded in the LCM feature but is actually embedded in the Fourier coefficients which are mapped to the feature via the LCM. The watermarked that the proposed audio watermarking scheme is not only resilient against common signal processing operations, including low-pass filtering, MP3 recompression, echo addition, volume change, normalization, test functions in show that the LCM feature is very robust to kinds of audio geometric distortions such as cropping, resample TSM, pitch-invariant TSM, tempo-invariant pitch shifting, and DA/AD conversion.

6. CONCLUSIONS

The level of watermarking increases robustness of the secret information. The watermarks are embedded into non overlapping DCT coefficients of the audio signal which are randomly selected and very hard to detect even with the blind detection. The audio watermarking is relatively new and has wide scope for research. For future, a new algorithm will proposed that taking features of Human Auditory System and the signal processing theories. Proposed algorithm is based on DCT domain while considering the more active components of the signal.

7. REFERENCES

[1] Hong, Z., Wu M., Wang, Z. & Liu, K. 2003, Non linear collusion attacks on independent fingerprints for multimedia. In: Proc. IEEE Computer Society Conference on Multimedia and Expo, Baltimore, MD, p 613–616.

[2] Goresky, M., Klapper, A. M., Fibonacci and Galois 2002, Representations of Feedback-With-Carry Shift Registers, IEEE Transaction on Information Theory, Vol. 48, No. 11, pp. 2826-2836.

[3] Shoemaker, C. 2002, Hidden bits: A survey of techniques for digital watermarking, Independent study, EER 290, spring.

[4] Bloom, J., Cox, I., Kalker, T., Linnartz, J., Miller, M., and Traw, C. 1999, Copy protection for DVD video, Proceedings of the IEEE, vo. 7, Issue 87, pp. 1267- 1276.

[5] Zwicker, E., and Fast, H., Psychoacoustics 1999. Springer Verlag, Berlin, Germany.

[6] Hernandez, J., Rodriguez, J. & Perez-Gonzalez, F. 2001 Improving the performance of spatial watermarking of images using channel coding. Signal Processing 80(7): p 1261–1279.

[7] Polikar, R., Home page - Dr. RobiPolikar, Jan 2001. [Online]. Available: http://users.rowan.edu/~polikar /WAVELETS/WTtutorial.html. Accessed: July 21, 2010.

[8] Kennedy, J., IFPI Digital Music Report 2010. [Online]. Available: http://www.ifpi.org/content/library/DMR2010.pdf. Accessed: Aug 8, 2010.

[9] Khayam, S. A. 2003, The Discrete Cosine Transform (DCT): Theory and Application, Information Theory and Coding, Seminar 1 – The Discrete Cosine Transform: Theory and Application, March 10.

[10] Salomonsen, K. 1997, Design and Implementation of an MPEG/Audio Layer III Bit stream Processor, Master´s thesis, Aalborg University, Denmark.

[11] Voyatzis, G. and Pitas, I. 1996, Applications of toral automorphisms in image watermarking, Proceedings of International Conference on Image Processing, vol. 1, pp. 237– 240.

[12] SDMI 2000. Call for Proposals for Phase II Screening Technology Version 1.0, [Online]. Available: http://www.sdmi.org/download/FRWG00033102-AMD1.htm [Accessed July 15. 2010].

[13] Steinebach, M., Petitcolas, F., Raynal, F., Dittmann, J., Fontaine, C., Seibel, S., et al., Stirmark benchmark: Audio watermarking attacks, Proceedings of the International Conference on Information Technology: Coding and Computing, pp. 49-54, 2001, Las Vegas, Nevada.

[14] Wang, X. and Zhao, H. 2005, A Blind Audio Watermarking Robust Against Synchronization Attacks, CIS 2005, Part II, LNAI 3802, pp. 617-622.

[15] Katzenbeisser, S. and Petitcolas, F.A.P. 2000, Information hiding techniques for steganography and digital watermarking, Artech House Publishers.

[16] Arnold, M., Schmucker, M. and Wolthusen, S. D. 2003, Techniques and Applications of Digital Watermarking and Content Protection. Boston, London: Artech House, INC.

[17] Kumar, M. N. 2004, Watermarking Using Decimal Sequences, M.S. thesis, Louisiana State University, Baton Rouge, LA, USA.

[18] Cvejic, N. & Seppanen, T. 2003, Robust audio watermarking in wavelet domain using frequency hopping and modified patchwork method. In: Proc. International Symposium on Image and Signal Processing and Analysis, Rome, Italy, p 251–255.

[19] Petitcolas, F. 2000, Watermarking schemes evaluation, IEEE Signal Processing Magazine 17(5): p 58–64.

[20] Steinebach, M., Petitcolas, F., Raynal, F., Dittmann, J., Fontaine, C., Seibel, S., Fates, N. & Ferri, L. 2001 Stirmark benchmark: Audio watermarking attacks, In: Proc. International Conference on Information Technology: Coding and Computing, Las Vegas, NV, p 49–54.

[21] Voloshynovski, S., Pereira, S., Iquise, V. & Pun, T. 2001, Attack modelling: towards a secondgeneration watermarking benchmark, Signal Processing 81(6): p 1177–1214.

[22] Miller, M, Dorr, G. & Cox, I. 2002, Dirty-paper trellis codes for watermarking, In: Proc. IEEE International Conference on Image Processing, Rochester, NY, p 129–132.

[23] Petitcolas, F.A.P. 2000, Watermarking schemes evaluation, IEEE Signal Processing Magazine [Online], Volume 17, Issue 5, pp.58-64.