DIGITAL AUDIO WATERMARKING: SURVEY

MIKDAM A. T. ALSALAMI* and MARWAN M. AL-AKAIDI**

* *Computer Science Dept. – Zarka Private University, Jordan*
** *School of Engineering and Technology - De Montfort University, UK*
email: mma@dmu.ac.uk

Abstract: Digital audio watermarking is a technique for embedding additional data along with audio signal. Embedded data is used for copyright owner identification. A number of audio watermarking techniques are proposed. These techniques exploit different ways in order to embed a robust watermark and to maintain the original audio signal fidelity. This paper makes a tutorial in general digital watermarking principles and focus on describing digital audio watermarking techniques. These techniques are classified according to the domain where the watermark is embedded.

Keywords: Digital watermarking, audio, copyright protection.

1. INTRODUCTION

As digital multimedia works (video, audio and images) become available for retransmission, reproduction, and publishing over the Internet, a real need for protection against unauthorized copy and distribution is increased. These concerns motivate researchers to find ways to forbid copyright violation. The most promising solution for this challenging problem seems to lie in information hiding techniques. Information hiding is the process of embedding a message into digital media. The embedded message should be imperceptible; in addition to that the fidelity of digital media must be maintained.

Information hiding is unlike cryptography. In cryptographic techniques significant information is encrypted so that only the key holder has access to that information, once the information is decrypted the security is lost. In information hiding, message is embedded into digital media, which can be distributed and used normally. Information hiding doesn't limit the use of digital data.

Information hiding can be classified into two types of techniques: Steganography and Watermarking. The main purpose of steganography is to hide the fact of communication. The sender embeds a secret message into digital media (e.g. image) where only the receiver can extract this message. The warden of communication channel will notice the transmitted media, but he/she will never perceive the buried secret message inside this media. Figure 1.1 illustrates a simple steganographic system. In this system the message *m* is embedded into the Cover-object *C* (could be image, audio or video) to produce the Stego-object *S* that should has the same fidelity of *C*. The Cover-Object is only used for the Stego-object generation and is then discarded. The embedding operation is parameterized by the key *k* that is known for both ends of communication: sender and receiver. On receiver side the buried message is extracted from Stego-object in detection process. Embedding message should be perceptually and statistically undetectable for the warden. An ideal steganographic system would embed a large amount of information perfectly securely with no visible degradation to the cover-object.

Watermarking is very similar to steganography in that both seek to hide information in the Cover-object. However steganography is related to secret point-to-point communication between two parties. Thus, steganography techniques are usually having a limited robustness and protect for the embedded information against modifications that may occur during transmission, like format conversion, compression or A/D conversion. On the other hand, watermarking rather than steganography principles

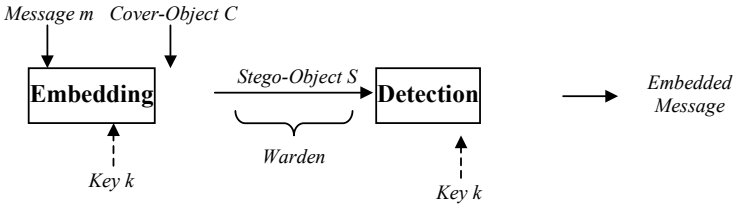

Figure 1.1 Steganographic System

is used whenever the media is available to parties who know the existence of the embedded information and may have interest removing it. Thus, watermarking adds additional requirements of robustness. An ideal watermarking system would embed information that could not be removed or altered without making significant perceptual distortion to the media. A popular application of watermarking is to give a proof of ownership of digital data by embedding copyright statements.

This paper is organized as follow. Section 2 describes the modules of watermarking systems and the function of each module. Sections 3 and 4 are to explain the applications and requirements of digital watermarking. Section 5 covers digital audio watermarking techniques through subsections. Finally, conclusions and general work frame for audio signal are presented.

2. WATERMARKING SYSTEM MODULES

A watermarking system consists of three modules that are watermark signal generation module, watermark embedding module and watermark detection module. Watermark signal is generated by using a non-invertible function that takes, as an input, a watermark key. In some systems the host signal (cover-object) is taken into account when watermark is generated. This will help watermark generator in producing an imperceptible signal-dependent watermark.

Watermark embedding is performed in time domain or in transform domain (DFT, DCT, DWT, …etc) using a suitable embedding rule (e.g. addition or multiplication). Finally, watermark is detection is performed by some sort of correlation detector or statistical hypothesis testing, with or without resorting to the original signal.

3. DIGITAL WATERMARKING APPLICATIONS

The requirements that watermarking system has to comply with are always based on the application. Thus, before we review the requirements and design considerations, we will present the applications of watermarking [Cox et al, 2002; Katzenbeisser and Petitcolas, 2000]:

3.1 Copyright protection

Copyright protection is the most important application of watermarking. The objective is to embed information identifies the copyright owner of the digital media, in order to prevent other parties from claiming the copyright. This application requires a high level of robustness to ensure that embedded watermark cannot be removed without causing a significant distortion in digital media. Additional requirements beside the robustness have to be considered. For example, the watermark must be unambiguous and still resolve rightful ownership if other parties embed additional watermarks.

3.2 Fingerprinting

The objective of this application is to convey information about the legal recipient rather than the source of digital media, in order to identify single distributed copies of digital work. It is very similar to the serial number of software product. In this application a different watermark embedded into each distributed copy. In contrast the first application where only a single watermark is embedded into all copies of digital media. As well as copyright protection application of watermarking, fingerprinting requires high robustness.

3.3 Content Authentication

The objective of this application is to detect modification of data. This can be achieved with so-called fragile watermark that have a low robustness to certain modification (e.g. Compression).

3.4 Copy Protection

This application tries to find a mechanism to disallow unauthorized copy of digital media. Copy protection is very difficult in open systems; in closed system, however, it is feasible. In such systems it is possible to use watermarks to indicate the copy status of the digital media (e.g. copy once or never copy). On the other side, copy software or device must be able to detect the watermark and allow or disallow the requested operation according to the copy status of the digital media being copied.

3.5 Broadcast Monitoring

Producers of advertisements or audio and video works want to make sure that their works are broadcasted on the time they purchase from broadcasters. The low-tech method of broadcast monitoring is to have human observers watch the broadcasting channels and record what they see or hear. This method is costly and error prone. The solution is to replace the human monitoring with automated monitoring. One method of automated broadcast monitoring is to use the watermarking techniques. With watermarking we can embed an identification code in the work being broadcasted. A computer-base monitoring system can detect the embedded watermark, to ensure that they receive all of the airtime they purchase from the broadcasters.

4. PROPERTIES OF DIGITAL WATERMARKING

Watermarking systems can be characterized by a number of properties [Cox et al, 2002; Katzenbeisser and Petitcolas, 2000]. The relative importance of each property depends on the requirements of the system application.The properties being discussed in this section are

associated with watermark embedder, watermark detector, or both.

4.1 Embedding Effectiveness

The effectiveness of a watermarking system is the probability that the output of the embedder will be watermarked. The cover work is said to be watermarked when input to a detector result in positive detection. The effectiveness of a watermarking system may be determined analytically or empirically by embedding a watermark in a large number of cover works and detect the watermark. The percentage of cover works that result in positive detection will be the probability of effectiveness.

4.2 Fidelity

In general, the fidelity of a watermark system refers to the perceptual similarity between the original and the watermarked version of the cover work. However, watermarked work may be degraded in the transmission process prior to its being perceived by a person, a different definition of fidelity may be more appropriate. We may define watermarking system fidelity as a perceptual similarity between the unwatermarked and watermarked works at the point at which they are presented to a viewer.

4.3 Data Payload

Data payload refers to the number of bits a watermark embeds in a unit of time or works. For audio, data payload refers to the number of embedded bits per second that are transmitted. Different applications require different data payload. For example, Copy control applications may require a few bits embedded in cover works.

4.4 Blind or Informed Detector

We refer to the detector that requires the original, unwatermarked work as an informed detector. Informed detectors may require information derived from the original work rather than original work itself. Conversely, detectors that do not require the original work are referred to as blind detectors. Informed detector has a good performance in watermark extraction. However, this will result in a huge number of original works have to be stored.

4.5 False Positive Rate

A false positive is the detection of a watermark in a cover work that does not actually contain one. When we talk of a false positive rate, we refer to the number of false positives we expect to occur in a given number of runs of the detector.

4.6 Robustness, Security and Cost

Robustness refers to the ability to detect the watermark after common signal processing operations. Audio watermarking needs to be robust to temporal filtering, A/D conversion, time scaling, etc. not all applications of watermarking require all the forms of robustness. This depends on the nature of application of watermarking system.

The security of a watermark refers to its ability to resist hostile attacks. Hostile attack is the process specifically intended to thwart the watermark's purpose. The types of attacks can fall in three categories: unauthorized removal, unauthorized embedding, and unauthorized detection.

The Cost of watermarking system refers to the speed with which embedding and detection must be performed and the number of embedders and detectors that must be deployed. Other issues include the whether the detector and embedder are to be implemented as hardware device or as software application or plug-ins.

5. DIGITAL AUDIO WATERMARKING

Watermarking digital media has received a great interest in the literature and research community. Most watermarking schemes focus on image and video watermarking. A few audio watermarking techniques have been reported. Digital audio watermarking is the process of embedding a watermark signal into audio signal. Audio watermarking is a difficult process because of the sensitivity of Human Auditory System (HAS).

The requirements mentioned earlier are common to both image and audio watermarking techniques. Despite their similarities, audio and still image watermarking systems exhibit significant differences. First of all, the fact that images are two-dimensional signals provides attackers with more ways of introducing distortions that might affects watermark integrity e.g. scaling, rotation or removal of rows/columns. Audio watermarking methods need not to deal with such attacks, as audio is a one-dimensional signal. Due to the difference between HAS and Human Visual System (HVS), different masking principles should taken into account in each case.

Digital audio watermarking techniques can be classified according to the domain where the watermarking takes place. The following sections will discuss audio watermarking techniques and classify them to four categories.

5.1 Frequency Domain Audio Watermarking

Audio watermarking techniques, that work in frequency domain, take the advantage of audio masking characteristics of HAS to embed an inaudible watermark signal in digital audio. Transforming audio signal from time domain to frequency domain enables watermarking system to embed the watermark into perceptually significant components. This will provide the system with a high level of robustness [Cox et al, 1997], because of that any attempt to remove the watermark will

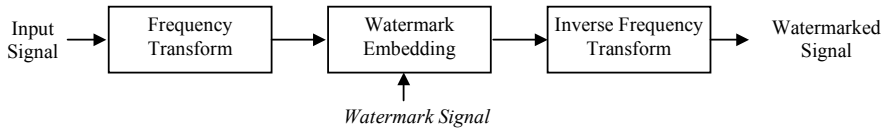

Figure 5.1
Watermarking in Frequency Domain

result in introducing a serious distortion in original audio signal fidelity.

The input signal is first transformed to frequency domain where the watermark is embedded, the resulting signal then goes through inverse frequency transform to get the watermarked signal as output as shown in Figure 5.1.

Watermark can be embedded into frequency domain components by mean of different methods, Cox and et al [Cox et al, 1997] proposed the use of spread spectrum technique in frequency domain. In spread spectrum communication, one transmits a narrowband signal over a much larger bandwidth such that the signal energy present in any single frequency is imperceptible. Similarly the watermark is spread over very many frequency components so that the energy of any component is very small and certainly undetectable. In this method the frequency domain of cover signal is viewed as a communication channel and the watermark is viewed as a signal that is transmitted through it. Attacks and unintentional signal distortions are thus treated as noise that the transmitted signal must be immune to. They claim that in order for the watermark to be robust, watermark must be placed in perceptually significant regions of the cover signal despite the risk of potential fidelity distortion. Conversely if the watermark is placed in perceptually insignificant regions, it is easily removed, either intentionally or unintentionally by, for example, signals compression techniques that implicitly recognize that perceptually weak components of a signal need not be represented.

Suppose that the watermark W consists of a sequence of real numbers, $W = w_1, w_2, \ldots, w_n$. In order for W to be embedded into a cover signal, S, a sequence of values, $V = v_1, v_2, \ldots, v_n$, is extracted from frequency spectrum of S, the watermark W will be embedded into V to obtain $V' = v'_1, v'_2, \ldots, v'_n$. V' is then inserted back to S in place of V to obtain a watermarked signal S'. Only copyright owner knows the locations of V sequence values in frequency spectrum of S. This will ensure the security of the watermark. S' maybe altered, by intentional or unintentional attacks, to produce S^*. Given S and S^*, a possibly corrupted watermark W^* is extracted and compared to W. W^* is extracted by first extracting V^* from S^* and then generating W^*. Figure 5.2 depicts watermark embedding and extraction.

There are three natural formulae for computing V':

$$v'_i = v_i + \alpha w_i$$

$$v'_i = v_i (1 + \alpha w_i)$$

$$v'_i = v_i (e^{\alpha w_i})$$

α is scaling parameter (controls robustness and fidelity).

There are a number of ways that one can use to evaluate the similarity between two watermarks. A traditional correlation measure can be used, for example. Similarity of W and W^* can be measured by:

$$\text{sim}(W, W^*) = \frac{W^* . W}{\sqrt{W^* . W^*}}$$

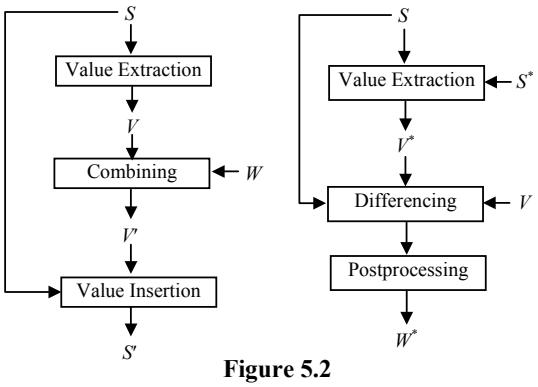

Figure 5.2
Watermark Embedding and Extraction

Where $X.Y = \sum_{i=1}^{n} x_i . y_i$

Another audio watermarking technique uses statistical algorithm works in Fourier domain [Arnold, 2000; Arnold, 2001]. This method is based on the patchwork algorithm [Bender et al, 1996] and doesn't need the original audio in detection process.

Audio signal is broken into frames; each frame is used to embed one bit. Each frame is transformed to frequency domain using DFT. Assume that the transformed frame contains 2N values, and then the embedding process works as follows:

1. Map a secret key and the watermark to the seed of random-number generator. Start the generator to pseudorandomly select two intermixed subsets A ={a_i} i=1,…,M and B={b_i},i=1,…,M of equal size M ≤ N from the original set of audio signal frequency spectrum.

2. Alter the selected elements $a_i \in A$ and $b_i \in B$, i=1,…,M according to the following embedding function:

   $a_i' = a_i + \Delta a_i$ & $b_i' = b_i - \Delta b_i$

Δa_i and Δb_i are two patterns generated by the secret key. There are two patterns for 0 and another

two for 1. We have to select the correct patterns according to the value of the bit being embedding. The alterations of frequency domain coefficients have to be performed in a way that achieves inaudibility. Therefore, Δa_i and Δb_i are driven from psychoacoustics model. Thus, Δa_i and Δb_i are reshaped for each individual frame. For more information about psychoacoustics model see [Painter and Spanias, 2000].

In watermark detection process, hypothesis testing is used. We formulate test hypothesis, H_0, and alternative hypothesis, H_1, the appropriate test statistic z will be a function of the sets A and B with probability distribution function PDF $\varnothing(z)$ in the unwatermarked case and $\varnothing_m(z)$ in watermarked case.

H_0: the watermark is not embedded; z follows PDF $\varnothing(z)$.

H_1: the watermark is embedded; z follows PDF $\varnothing_m(z)$.

Two kind of error are incorporated in hypothesis testing:

$$\text{I}: \int_T^{+\infty} \phi(z)dz = P_I \quad \text{(Type I error)}$$

$$\text{II}: \int_{-\infty}^{T} \phi_m(z)dz = P_{II} \quad \text{(Type II error)}$$

Hypothesis testing is used in the detection to decide whether the watermark bit is embedded or not. The threshold T is used in the detection step. Detection procedure is as follows:

1. Map the secret key ad the watermark to the seed of random-number generator to generate the sunset C and D. C =A and D = B if a correct key is used.
2. Decide the probability of correct rejection 1 – PI according to the application and calculation the threshold T from error type I equation.
3. Calculate the sample mean E(z) = E(f(C,D)) and choose between two mutually exclusive propositions:

   H_0: E(z) ≤T the watermark bit is embedded.

   H_1: E(z)>T the watermark is not embedded.

Hypothesis testing depends on appropriate test statistic. Two test statistics can be used in watermark detection:

1. The first test statistic uses the function to measure the difference between population means of A and B:

$$z = f(A,B) = \frac{\overline{a'} - \overline{b'}}{\sigma_{\overline{a'}-\overline{b'}}}$$

Therefore the two mutually exclusive propositions become:

H0: $\varnothing(z) = N(0,1)$

H1: $\varnothing m\,(z) = N\,(zm, 1)$,

$$z_m = \frac{k(\overline{a} + \overline{b})}{\hat{\sigma}_{\overline{a'}-\overline{b'}}}$$

Where N (μ, σ^2) is the normal distribution with the mean μ and standard deviation σ, and

$$k = \sqrt{\frac{1}{1-(z_1 - P_I + z_1 - P_{II})^2 \varepsilon^2} - 1}$$

2. The second test statistic uses another function:

$$z = f(A,B) = \frac{\dfrac{\overline{a'} - \overline{b'}}{\sigma_{\overline{a'}-\overline{b'}}}}{\dfrac{1}{2}\dfrac{\overline{a'} + \overline{b'}}{\sigma_{\overline{a'}+\overline{b'}}}} = 2\frac{\overline{a'} - \overline{b'}}{\overline{a'} + \overline{b'}}$$

The threshold T must be computed and compared with the mean value calculated by one of the above statistics functions.

It is clear that the detection process doesn't require the original audio signal while it works to detect the statistics changes in the media to determine whether it is watermarked or not.

Further research has been achieved to improve the performance of above watermarking system, for more information see [Hong et al, 2002; Yeo and Kim, 2001]

5.2 Time Domain Audio Watermarking

In time domain watermarking techniques, watermark is directly embedded into audio signal. No domain transform is required in this process. Watermark signal is shaped before embedding operation to ensure its inaudibility (Figure 5.3). The available time domain watermarking techniques insert the watermark into audio signal by simply adding the watermark to the signal.

Embedding a watermark into time domain involves challenges related to fidelity and robustness. Shaping the watermark before embedding enables the system to maintain the original audio signal fidelity and renders the watermark inaudible. As for robustness, time domain watermarking systems use different techniques to improve the robustness of the watermark.

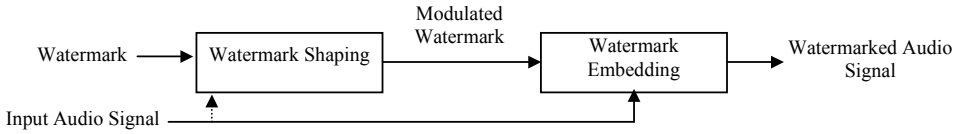

Figure 5.3
Time Domain Watermarking

Working in frequency domain enables watermarking system to embed a robust watermark, while it is possible to identify the most significant components of the cover signal. Also, masking characteristics of audio signal can be exploited, in order to reduce the distortion of embedded watermark.

In this section, two methods for audio watermarking in time domain are shown. The first one presented in [Bassia and Pitas, 1998; Bassia et al, 2001] and in which the watermark signal is modulated using the original audio signal and filtered by lowpass filter to reduce the distortion that might be result from embedding the watermark. The original audio signal is divided into segments and then each segment is watermarked separately by embedding the same watermark. Watermark signal, $w_i \in \{1, -1\}$, i=0,1,…,n-1 is generated by threshold a chaotic map in a way similar to the one described in [Bassia et al, 2001]. The seed (start point) of the chaotic sequence generator is the watermark key. Using the chaotic sequence generator is to ensure the security of the watermarking system i.e. the sequence generation mechanism cannot be reversed engineered.

Suppose that we have a segment of audio signal $S = s_1, s_2, \ldots, s_n$ then the watermarking process begin by modulating the watermark signal w_i by using audio signal S,

$$w_i' = \alpha \mid s_i \mid \oplus w_i \; i = 0,1,\ldots, n\text{-}1$$

Where $\oplus$ denotes a superposition law which can be multiplication, power law, etc, and α is a constant controls the amplitude of the watermark signal. The maximum allowable watermark amplitude is the limited by the maximum perceived signal distortion.

In next stage, w_i' is shaped using a lowpass Hamming filter of length (order) L:

$$w_i'' = \sum_{l=0}^{L-1} b_l w_{i-l}'$$

where b_l is the filter coefficients. This process results in inaudible watermark signal. Figure 5.4 [Bassia et al, 2001] shows the power spectral density (*PSD)* of two watermark signals, one is shaped and the other is not. It is clear that the unshaped watermark signal is audible while it has a PSD exceeds the power of the original signal in certain frequencies. The PSD of the shaped watermark signal lies underneath the original audio signal in the entire frequency range.

Finally the shaped watermark signal is embedded into audio signal:

$$y_i = s_i + w_i'' \quad i = 0,1,\ldots, n\text{-}1$$

It is obvious that the calculation of watermarked sample y_i is based on the neighbors of the sample s_i and the chaotic signal (watermark) w_i.

In detection stage, the received signal, Y, broken in the same way that original signal is broken. Consider the following sum:

$$C_k = \frac{1}{n} \sum_{i=0}^{n-1} y_{(i+)\bmod n} w_i$$

C_k is the correlation of W with Y, evaluated for all possible circular shift of Y. By substitution and

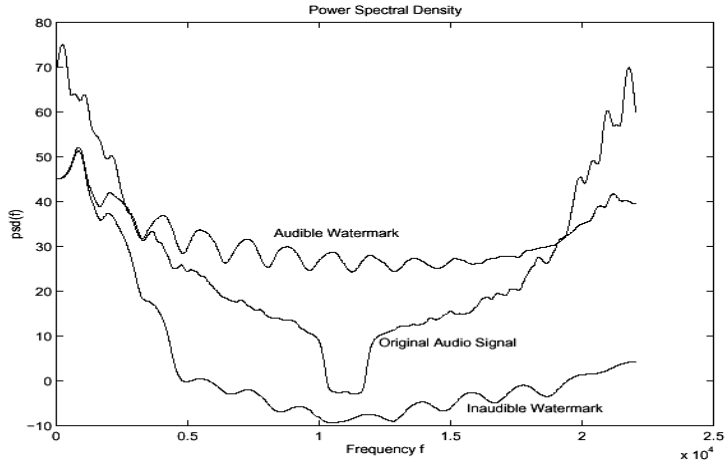

Figure 5.4
Power spectral density of two watermark and original signals

rearranging the above equation we get:

$$C_k = \frac{1}{n}\left(\sum_{i=0}^{n-1} s_{(i+k) \bmod n} w_i + \sum_{i=0}^{n-1} w''_{(i+k) \bmod n} w_i\right)$$

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

$$\sum_{i \in B} w_i = \Delta w$$

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

$$\sum_{i \in A} w_i = 0$$

So, C_k can be expressed as follows:

$$C_k = \frac{1}{n}\left(\sum_{i \in A} s_{(i+k) \bmod n} w_i + \sum_{i \in B} s_{(i+k) \bmod n} w_i + \sum_{i=0}^{n-1} w''_{(i+k) \bmod n} w_i\right)$$

Let us define the following terms:

$$T_{1,k} = \frac{1}{n}\left(\sum_{i \in A} s_{(i+k) \bmod n} w_i\right)$$

$$T_{2,k} = \frac{1}{n}\left(\sum_{i \in B} s_{(i+k) \bmod n} w_i\right)$$

$$T_{3,k} = \frac{1}{n}\left(\sum_{i=0}^{n-1} w''_{(i+k) \bmod n} w_i\right)$$

It can be easily shown that $E(T_{1,k}) = 0$, where E() denotes the expected value operator. For the term T_2, it is easy to show that:

$$T_{2,k} = sign(\Delta w)\frac{1}{n}\left(\sum_{i \in B} s_{(i+k) \bmod n} w_i\right) = \frac{\Delta w}{n}\frac{1}{N_B}\sum_{i \in B} s_{(i+k) \bmod n}$$

Therefore

$$E(T_{2,k}) = \frac{\Delta w}{n} m_S$$

If no watermark has been embedded in the signal, $T_3 = 0$and thus:

$$C_k \approx T_{2,k} = \frac{\Delta w}{n} m_S$$

On the other hand, if the signal is watermarked

$$C_k \approx T_{2,k} \approx \frac{\Delta w}{n} m_S + \frac{1}{n}\left(\sum_{i=0}^{n-1} w''_{(i+k) \bmod n} w_i\right)$$

For watermark detection we construct the ratio r_k:

$$r_k = \frac{C_k - T_{2,k}}{T_{3,k}}$$

The original signal S is required for evaluation of $T_{2,k}$ and $T_{3,k}$, but it can be replaced by Y without significant error.

The value of r_k is computed for every k = 0,1,…,n-1, for all segments. We compute the detection value of the audio segment j as $R_j = \sum_{i=0}^{n-1} r_i$, the final detection value is $R = \sum_{j=0}^{N_s-1} R_j$, where N_s is the number of segments in signal.

The decision about the existence of the watermark is made depending on a threshold value compared with R.

It is clear that this watermarking system is immune against time-shifting and cropping. The fact that C_k is computed for all possible circular shift of Y, ensures synchronization between Y and W will occur for certain value of k=0,1,…,n-1.

Another watermarking system uses the HAS masking effects to shape the watermark signal [Boney et al, 1996; Swanson et al, 1998]. Shaping operation is performed in frequency domain, but the shaped watermark is embedded into audio signal in time domain. Watermark is a noise-like sequence generated by using two keys x_1 and x_2. The first key x_1 is author dependent. The second key x_2 is computed from audio signal that the author wants to watermark. It is computed from the signal using a one-way hash function. The two keys are mapped to pseudorandom number generator to generate a noise-like sequence, watermark. Original audio signal is required in detection process to compute the second key x_2, and to extract the embedded watermark.

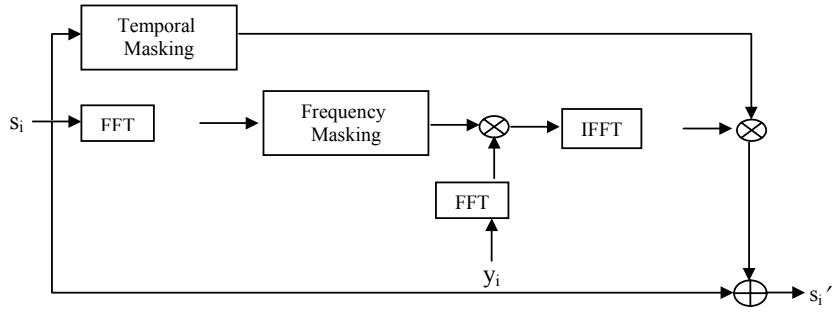

Figure 5.5
Audio Segment Watermarking Procedure

The watermarking process begins with dividing the audio signal into segments, and then each segment is watermarked separately. Suppose that you have a generated watermark y_i, and then the algorithm of watermarking an individual segment, s_i, works as follows:

1. Compute the power spectrum S_i of audio signal segment s_i as follows:

$$S_i = 10\log_{10}\left[\frac{1}{N}\left\|\sum_{n=0}^{N-1} s_n h(n)\exp(-j2\pi\frac{2i}{N}\right\|^2\right]$$

Where h(n) is a Hann window: $h(n) = \frac{\sqrt{8/3}}{2}\left[1-\cos\left(2\pi\frac{n}{N}\right)\right]$

N is the number of samples in one segment and j is $\sqrt{-1}$

2. Compute the frequency masking threshold M_i of the power spectrum S_i.
3. Use the mask M_i to weight the noise-like watermark, $P_i = M_i * Y_i$, where P_i is the weighted watermark and Y_i is the power spectrum of the watermark signal y_i.
4. Compute the inverse of FFT of the shaped watermark $p_i = IFFT(P_i)$.
5. Compute the temporal masking t_i of s_i.
6. Use the temporal masking t_i to further shape the frequency shaped watermark to create the final watermark $w_i = t_i * p_i$ of the audio segment.
7. Create the watermarked segment $s_i' = s_i + w_i$.

Figure 5.5 shows a diagram of watermark shaping and embedding.

In detection process, the original audio signal is known. Thus, second key can be computed and then the watermark signal can be reconstructed. Also the embedded possible distorted watermark can be extracted. Assume that r_i i = 0,1,…,N is a recovered piece of audio signal, then we can compute $x_i = r_i - s_i$. If r_i has a watermark then $x_i = w_i' + n_i$, where n_i is noise (intentionally or unintentionally added to the watermarked signal). Otherwise, $x_i = n_i$. Similarity between extracted watermark, x_i, and the reconstructed one can be measured by correlation as follows:

$$sim(x,w) = \frac{\sum_{i=0}^{N-1} x_i w_i}{\sum_{i=0}^{N-1} w_i w_i}$$

Then the value can be compared with a threshold T.

The recovered signal r_i is possible shifted. This leads to lose the synchronization between the extracted watermark and the reconstructed one. In such case we can assume that $r_i = s_{i+\tau} + x_i$, where x_i as mentioned before. τ is a n unknown delay, thus, a generalized likelihood ratio test must be performed to determine whether the audio signal is watermarked or not.

$$\frac{\max_\tau \exp(-\sum_{n=0}^{N-1}(r_n - (s_{n+\tau} + w_{n+\tau}))^2}{\max_\tau \exp(-\sum_{n=0}^{N-1}(r_n - (s_{n+\tau})^2)}$$

Then, this ratio is compared to a threshold.

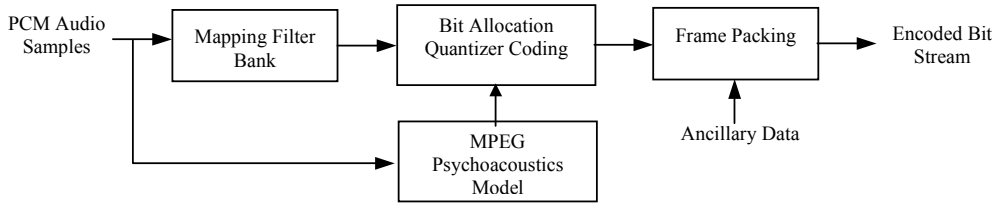

Figure 5.6
Structure of MPEG Audio Encoder

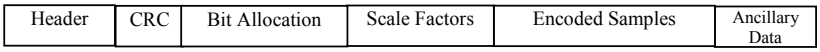

Figure 5.7
Frame Format of MPEG Audio

5.3 Compressed Domain Audio Watermarking

A number of techniques are proposed to embed a watermark signal into MPEG audio bit stream, rather than going through decoding/encoding process in order to apply watermarking scheme in uncompressed domain [Qiao and Nahrstedt, 1999; Neubauer and Herre, 2000a; Neubauer and Herre, 2000b; Neubauer and Herre, 1998]. Such systems are suitable for "pay audio" scenario, where the provider stores audio contents in compressed format. During download of music, the customer identifies himself/herself with his/her unique customer ID, which therefore is known to the provider during delivery. In order to embed the customer ID into the audio data using a watermarking technique, a scheme is needed that is capable of watermarking compressed audio on the fly during download.

MPEG audio compression is a lossy algorithm and uses the special nature of the HAS. It removes the perceptually irrelevant parts of the audio and makes the audio signal distortion inaudible to t human ear. For more information about MPEG audio Compression sees [Pan, 1995].

MPEG encoding process has the following steps:

1. Input audio samples pass through a mapping filter bank to divide the audio data into subbands (subsamples) of frequency.
2. At the same time, the input audio samples pass through MPEG psychoacoustics model, which creates a masking threshold of audio signal. Masking threshold is used by quantization and coding step to determine how to allocate bits to minimize the quantization noise audibility.
3. Finally, the quantized subband samples are packed into frames (coded stream).

Figure 5.6 shows the basic structure of an MPEG audio encoder.

Filter bank divides the input audio signal into 32 equal-width subbands, then the number of bits used in quantization is determine upon masking threshold to minimize the audibility of possible distortion maybe introduced by quantization.

The MPEG audio stream consists of frames. Frame is the smallest unit which can be decoded individually. Each frame contains audio data, header, CRC (Cyclic Redundancy Code), and ancillary data. In frame, each subband has three groups of samples with 12 samples per group. The encoder can use a different scale factor for each group. Scale factor is determine upon masking threshold and used in reconstruction of audio signal. The decoder multiplies the quantizer output to reconstruct the quantized subband sample. Figure 5.7 depicts the general format of MPEG frame.

MPEG audio decoding process is simple a reverse of the encoding process. The decoding takes the encoded bit stream as an input, unpacks the frames, reconstructs the frequency samples (subbands samples) using scale factors, and then inverses the mapping to re-create the audio signal samples. This process is depicted in Figure 5.8.

One audio watermarking technique [Qiao and Nahrstedt, 1999] embeds the watermark into scale factors of MPEG audio frames. In this technique, DES encryption algorithm is used in generating non-invertible watermark. Original data is applied into encryption algorithm to get the watermark as follows:

First, a key *KEY* is selected and for each MPEG audio frame a_j j=1,…,N (number of audio frames), we apply DES with *KEY* to it to get a random byte sequence *RBS*:

$$RBS = DES_{KEY}\ (\text{one audio frame } a_j)$$

Second, let RBS_i be i-th byte of random byte sequence and w_i be the i-th bit of the watermark bit stream, then the watermark can be created by:

$$w_i = \begin{cases} -1 & \text{if } RBS_i = \text{even number} \\ 1 & \text{otherwise} \end{cases}$$

Each scale factor takes 6 bits; therefore, we have as many as 63 levels of scale factors (indexed

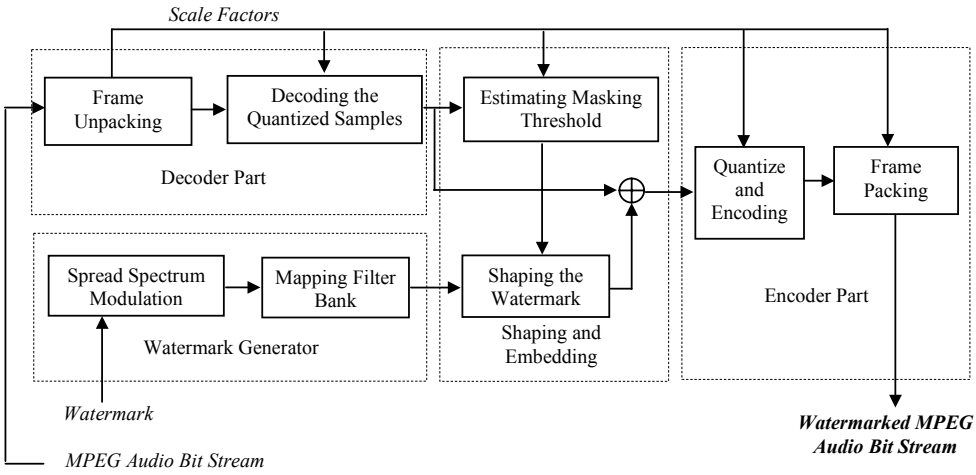

Figure 5.9
MPEG Audio Bit Stream Watermarking

from 0 to 62, 63 is not used by the standard). The level change of scale factor has an auditory effect that the sound becomes stronger when the scale factor level increases, and becomes weaker when the scale factor decreases. Increasing or decreasing scale factor by one level normally cannot be detected by listeners.

Let $ScaleFactor_i\ (index)$ be the i-th scale factor with the level indicated by *index* and SW_i be the i-th watermarked one. The watermarking procedure works as follows:

$$SW_i = \begin{cases} ScaleFactor_i(index) & \text{if } index + w_i = -1 \text{ or } 63 \\ ScaleFactor_i(inedex + w_i) & \text{otherwise} \end{cases}$$

This scheme has drawbacks. The first one is that the scheme doesn't has much data to watermark due to the few number of scale factors in audio frame. Also, the watermark scheme is not robust enough against attacker who is trying lower scale factors by 2 or 3 levels. On the other side, multiple watermarks cannot be applied. The reason is that when multiple watermarks are applied, certain scale factors would be increased by multiple levels and perceptible noise would be introduced.

Another watermarking scheme embeds the watermark into the encoded data. However, changing the all encoded samples shows a perceptible distortion. *Spacing Parameter sp* is introduced to solve this problem. sp is used in way like that every sp samples, we randomly select 1 or 2 samples to be watermarked. The watermark generation procedure will be modified to incorporate spacing parameter:

$$w_i = \begin{cases} -1 & \text{if } \mathrm{RBS}_i = 0 \pmod{sp} \\ 1 & \text{if } \mathrm{RBS}_i = 1 \pmod{sp} \\ 0 & \text{otherwise} \end{cases}$$

Let Sample_i be the i-th sample in audio frame and SW_i be the i-th watermarked sample. The watermarking will be:

$$SW_i = \begin{cases} Sample_i \text{ if every bit of } (Sample_i + w_i) \text{ is 1} \\ Sample_i + w_i \text{ otherwise} \end{cases}$$

Both watermarking schemes described above use the concept of spread spectrum watermarking, but through compressed domain.

The original MPEG audio is required in detection process and the watermark can simply extracted and verified.

Another technique [Neubauer and Herre, 2000a; Neubauer and Herre, 2000b; Neubauer and Herre, 1998] in MPEG audio stream watermarking is to partly decode the input bit stream, embed a perceptually hidden watermark in the frequency domain and finally quantize and code the signal again. Figure 5.9 illustrates a general structure of bit stream watermarking system.

This watermarking system consists of four parts. Each part has a specific function. We can see that this watermarking system has assembled parts of MPEG encoder and decoder, in addition to parts of frequency domain audio watermarking systems (watermark generation and watermark embedding). These parts have been modified in order to enable the system to embed the watermark in subbands samples.

The first part, decoder part, takes MPEG audio bit stream as an input and gives frequency subbands samples as output. This part supplies the other parts with scale factors that are necessary in masking threshold estimation and encoder process.

The second part, watermark generator, is used to convert the watermark to subband representation in order to be ready for embedding. The watermark can be any data provided by copyright owner. The

generated watermark is fed into watermark shaping and embedding part, which in turn, takes the decoded subbands samples and scale factors to estimate the masking threshold of the audio signal and use it in shaping the watermark. The last two parts have much similarity to the technique proposed in [Swanson et al, 1998].

The last part, encoder part, takes the watermarked subbands samples and scale factors. It decodes the samples using the original scale factors and then packs the resulting decoded samples. In order to avoid the possible distortion of requantization, the original scale factor is used and no need to recomputed new scale factors.

The embedded watermark can be detected in uncompressed domain as well as compressed domain. Original audio data is required to extract the watermark and then measure the similarity between the extracted watermark and the original one. The watermark detection in uncompressed domain can be achieved, exactly like the way presented in [Swanson et al, 1998], by using correlation measurement.

5.4 Wavelet Domain Audio Watermarking

Wavelet transform can be used to decompose a signal into two parts, high frequencies and low frequencies. The low frequencies part is decomposed again into two parts of high and low frequencies. The number of decompositions in this process is usually determined by application and length of original signal. The data obtained from the above decomposition are called the DWT coefficients. Moreover, the original signal can be reconstructed form these coefficients. This reconstruction is called the inverse DWT. The process of decomposition is depicted in Figure 5.10. For more information on Wavelet transform, see [Daubechies, 1992; Daubechies, 1988].

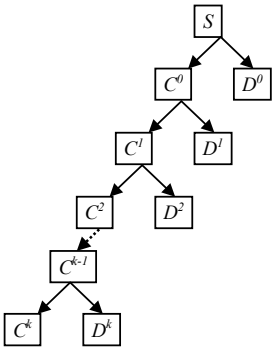

Figure 5.10
Wavelet Decomposition

A method of audio signal watermarking in wavelet domain uses patchwork algorithm [Kim et al, 2002]. In this method, a binary watermark w_i is embedded one bit in one data block. Watermark bits are locally repeated for the purpose of robustness. Also a number of bits are added in front of watermarks bits to locate the point where the watermark bit is embedded in watermarked signal. These bits are called synchronization bits. For example, with local redundancy rate 3 and synchronization bits 10101011, we change the original watermark as:

$$w_0w_1w_2... \rightarrow 10101011w_0w_0w_0w_1w_1w_1w_2w_2w_2....$$

Suppose that B is a block of audio signal being watermarked, we use DWT to have $D^0, D^1, D^2, ..., D^k, C^k$, for some integer k. then after patchwork algorithm is used to embed the watermark by

$$P_N = \sum_{i \in I} D_i^k - \sum_{j \in J} D_j^k$$

artificially modifying a patch value P_N as

Where I and J are two subset of indexes randomly generated. Proposed algorithm modifies P_N in a way that the modified P_N is deviation away

$$D_i^k \rightarrow D_i^k + \delta, \quad D_j^k \rightarrow D_j^k - \delta \quad \text{if} \quad w_n = 1$$
$$D_i^k \rightarrow D_i^k - \delta, \quad D_j^k \rightarrow D_j^k + \delta \quad \text{if} \quad w_n = 0$$

from expected. To be specific, we modify some wavelet coefficients in D^k as

For $i \in I$ and $j \in J$, w_n is a watermark bit being embedded and δ is a real number.

Different two subsets of indexes I and J are

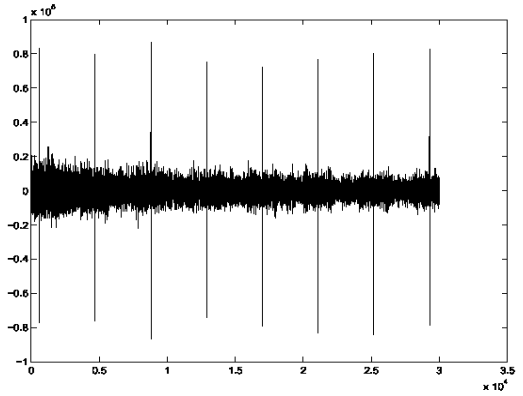

Figure 5.11
Δ^t_N for watermarked Audio Signal

used to embed the synchronization bits for security purpose.

In detection process, $\Delta t_N = P_N^{t+1} - P_N^t$, are computed, where P_N^{t+1} and P_N^t are two patch values of block B_{t+1} and B_t, respectively. Figure 5.10 [Kim et al, 2002] shows Δt_N for watermarked audio signal [Kim et al, 2002].

The peaks shown in this figure refers to the watermark bits locations in audio signal. Then

detection is made according the following criteria, for $\beta > 0$:

- If $\Delta_N^t > \beta N\delta$ and $\Delta_N^{t+1} < 0$ then 1 is detected in block B_t.
- If $\Delta_N^t < -\beta N\delta$ and $\Delta_N^{t+1} > 0$ then 0 is detected in block B_t.
- If previous two conditions are not satisfied, then no watermark bit is detected in this block.

Synchronization bits must be found first to determine the location of watermark bits.

This watermarking system shows a high performance in synchronization and resisting time shifting attack [Kim et al, 2002].

7. CONCLUSIONS

All watermarking systems are designed to achieve one goal that is embedding a hidden robust watermark into digital media. These systems have to satisfy two conflicting requirements. First, watermark must be immune against intentional and unintentional removal. Second, watermarked signal should maintain a good fidelity, i.e. watermark must be perceptually undetectable. To accomplish this task, variety of techniques has been exploited, and different domains are involved to enhance a certain application of watermarking and/or improve fidelity and robustness of watermarked signal.

However, watermarking systems have a number of differences. These differences can be considered in evaluating performance of watermarking systems and suitability of these systems for a specific application. These differences can be explained as follows:

1. Some audio watermarking systems require the original audio signal, or any information derived from it, to be presented in detection process. This will leads to a large number of original works have to be stored and searched during detection.

   Systems that require the original audio signal are not suitable for some type of applications, in case that detection process has no access to the original work or it is not acceptable to disclose it. On the other hand, presenting the original signal yields in efficient watermark extraction consequently efficient detection.

   Audio watermarking systems that are based on patchwork algorithm use a statistical detection process (hypothesis testing) and don't need the original audio for detection purpose. The most techniques that are base on correlation measurement of similarity require that signal except method presented in [Bassia and Pitas, 1998; Bassia et al, 2001].

   In spite of that a number of audio watermarking techniques require only the watermarked signal in detection watermark key is needed in both embedding and detection.

2. In order to maintain the watermark security, watermark would be embedded into selected regions of some domain transform of audio signal. These regions are selected randomly by generating a sequence of indexes. Sequence generation is paramerized by a key called watermarking key. This key is required in both embedding and detection.

   In some watermarking systems, watermarking key is used to generate the watermark itself. In this case, the watermark would be a random sequence of bits or digits generated by some sort of algorithms ensure non-invertiblitiy of watermark in order to maintain the security of watermarking key.

   Watermarking key could be provided by the copyright owner or a combination of information provided by him/her and information derived from original signal. In such case, original signal will be required in detection process for key generation purpose. In all scenarios, the key is used as a seed for random number generator.

   Sometimes, disclosing the watermarking key or having an access to it becomes impossible. Thus, using the same key in detection and embedding will not be acceptable. A solution to such problem could be found in using two keys, one for embedding and another for detection [Hong et al, 2002] (i.e. public-key or asymmetric watermarking system).

3. During embedding process, original audio signal is divided into frames. Then after, each frame is watermarked separately. Some watermarking systems embed the same watermark into a number of frames to enhance watermark robustness. But, in other systems each frame is watermarked with different watermark.

4. Because of sensitivity of HAS, watermark signal must be shaped to rent it inaudible. Masking characteristics of audio signal can be used for this purpose. Psychoacoustics MPEG model is commonly used to calculate masking threshold that is used in weighting the watermark. In some other audio watermarking systems, different techniques are used. These techniques use the original audio signal in modulating the watermark. Therefore; the amplitude of watermark signal is controlled by amplitude of audio signal. Watermark shaping process may effect the existence of the

watermark in cover work, consequently, false negative rate will be increased.

A general work frame for digital audio watermarking systems can be concluded as follows:

1. Watermarking system should be able to embed any set of data in to audio signal, and the detector should be able to retrieve the embedded data (i.e. not just report that watermark is presented or not)
2. Watermark embedded (detection) module should be independent of mode of operating. (e.g. the same watermark is embedded into multiple frames of audio signal or different watermark is embedded into each frame).
3. Watermarking key generation should be independent of watermark embedding and detection (e.g. embedding and detection will not be effected whether original signal is involved in key generation or not).

The above points enables audio watermarking system to be suitable for variety of application and make it possible to put standards (e.g. [SDMI, 2000]) and evaluation benchmarks.

7. REFERENCES

1. Arnold M. 2000, "Audio Watermarking: Features, Applications and Algorithms". *Multimedia and Expo. IEEE international Conf.*, Vol. 2, pp. 1013-1016.
2. Arnold M. 2001, "Audio Watermarking", *Dr. Dobb's Journal*, Vol. 26, Issue 11, pp. 21-26.
3. Bassia P. and Pitas I. 1998, "Robust Audio Watermarking in the Time Domain". *Signal Processing IX, theories and applications: proceeding of Eusipco-98, Ninth European Signal Processing Conf.*, Greece, pp. 8-11.
4. Bassia P., Pitas I., and Nikolaidis 2001, "Robust Audio Watermarking in Time Domain", *IEEE Trans. On Multimedia,* Vol. 3, pp. 232-241.
5. Bender W., Gruhl D., Morimoto N. and Lu A. 1996, "Techniques for Data Hiding", *IBM Systems Journal*, Vol. 35, No. 3&4, pp. 313-335.
6. Boney L. Tewfik A. H. and Hamdy K. N. 1996, "Digital Watermarking for Audio Signal". In *Proc. of EUSIPCO'96,* Sep., Vol. III, pp. 1697-1700.
7. Cox I. J., Kilian J. Leighton F. T. and Shamoon T. 1997, " Secure Spread Spectrum Watermarking for Multimedia". *IEEE Trans. On Image Processing*, Vol. 6, No. 12, pp. 1673-1687.
8. Cox I. J., Miller, M. L. and Bloom J. A. 2002, "Digital Watermarking". *Morgan Kaufmann Publishers*, USA.
9. Daubechies I. 1988, " Orthonormal Bases of Compactly Supported Wavelets". *Comm. Puse and Appk. Math*., Vol. 41, pp. 909-996.
10. Daubechies I. 1992, "Ten Lectures on Wavelets", *SIAM*, Philadelphia.
11. Hong D. G., Park S. H. and Shin J. 2002, " A Public Key Audio Watermarking Using Patchwork Algorithm". *Proceedings of ITC-CSCC* 2002,pp.160-163.
12. Katzenbeisser S. and Petitcolas F. A. P. 2000, " Information Hiding Techniques for Steganography and Digital Watermarking". *Artech House*, UK.
13. Kim H. O., Lee B. K. and Lee N. -Y. 2002, "Wavelet-Based Audio Watermarking Techniques: Robustness and Fast Synchronization". In *http://amath.kaist.ac.kr/research/paper/01-11.pdf.*
14. Neubauer C. and Herre J. 1998, "Digital Watermarking and its Influence on Audio Quality". *105th AES Convention, Audio Engineering Society preprint 4823*, San Francisco.
15. Neubauer C. and Herre J. 2000a, "Audio Watermarking MPEG-2 AAC Bitstream", *108th AES Convention, Audio Engineering Society Preprint 5101*, Paris.
16. Neubauer C. and Herre J. 2000b, "Advanced Audio Watermarking and Applications". *109th AES Convention, Audio Engineering Society Preprint 5176*, Los Angeles.
17. Painter T. and Spanias A. 2000, "Perceptual Coding of Digital Audio". *Proc. of IEEE*, Vol. 88, No. 4, pp. 451-513.
18. Pan D. 1995, "A Tutorial on MPEG / Audio Compression". *IEEE Multimedia*, pp. 60-74.
19. Qiao L. and Nahrstedt K. 1999, "Non-invertible Watermarking Methods for MPEG Encoded Audio", *Conf. on Security and Watermarking of Multimedia Contents*, pp. 194-202.
20. SDMI, 2000, Call for proposal, *http://www.sdmi.org/cfp.html*.
21. Swanson M. D., Zhu B., Tewfik A. H. and L. Boney L. 1998, "Robust Audio Watermarking Using Perceptual Masking", *Elsevier Signal Processing, Sp. Issue on Copyrights Protection and Access Control*, Vol. 66, No. 3, pp. 337-355.
22. Voyatzis G. and Pitas I. 1998, "Chaotic Watermarks for Embedding in Spatial Digital

Image Domain". In *Proc. ICIP98*, Chicago, Vol. II, pp. 432-436.

23. Yeo I.-K. and Kim H. J. 2001, " Modified Patchwork Algorithm: A Novel Audio Watermarking Scheme". *Proc. of the International Conf. On Information Technology: Coding and Computing* , pp. 237 – 242.