AUDIO WATERMARK RESISTANT TO MP3 COMPRESSION

UNDER THE GUIDANCE OF

PROF. INDRANIL SENGUPTA

A SYNOPSIS SUBMITTED

IN PARTIAL FULFILLMENT OF REQUIREMENTS

FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

BY

Muneish Adya

03CS3013

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
KHARAGPUR

1. **Introduction**

The fast growth of the Digital-Internet world and the maturity of audio compression techniques enable the music or song creators to distribute their products digitally over the net. However, an unlimited number of perfect copies of these digital files can be illegally produced. This poses a serious threat to the rights of content owners and leads to musicians and song-writers not getting their due.

Digital watermarking has been proposed as a new, alternative method to enforce the intellectual property rights and protect digital media from tampering. It involves a process of embedding into a host signal a perceptually transparent digital signature, such as an author’s signature, a company logo etc., carrying a message about the host signal in order to "mark" its ownership. The digital signature is called the digital watermark. Although perceptually transparent, the existence of the watermark is indicated when watermarked media is passed through an appropriate watermark detector. It has been generally agreed that an effective watermarking scheme should satisfy three properties:

- Imperceptibility
- Robustness
- Security

However, an ideal method should find the optimum tradeoff exists between the degree of host audio signal degradation and the resistance to common signal processing attacks.

The fundamental process in each watermarking system can be modeled as a form of communication where a message is transmitted from watermark embedder to the watermark receiver. The process of watermarking is viewed as a transmission channel through which the watermark message is being sent, with the host signal being a part of that channel. In Figure 1.1, a general mapping of a watermarking system into a communications model is given. After the watermark is embedded, the watermarked work is usually distorted after watermark attacks. The distortions of the watermarked signal are similar to the data communications model, modeled as additive noise.

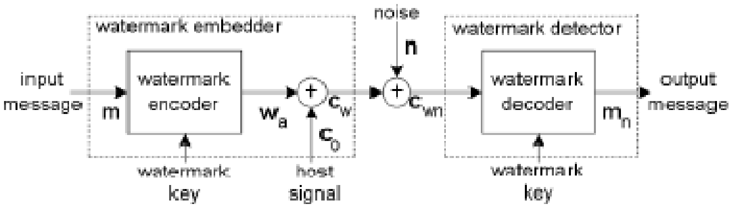


Fig 1.1 A watermarking system and equivalent Communications Model

Most audio watermarking schemes rely on the imperfections of the human auditory system (HAS). In the time domain, it has been demonstrated that the HAS is insensitive to small level changes [5] and insertion of low-amplitude echoes [6]. Data hiding in the frequency domain takes advantage of the insensitivity of the HAS to small spectral magnitude changes [7,8,6]. Quantization index modulation is another type of data hiding algorithms that increases the security of the augmented data at the cost of decreased tolerance to attack noise stronger than the watermark modulation. The Discrete Wavelet Transform has recently provided a new dimension to audio watermarking and a lot of new watermarking algorithms are based on this concept [10].

In our endeavor for developing a robust, imperceptible and blind watermarking scheme, the human auditory system, Spread Spectrum watermarking and the watermarking algorithms based on Discrete Wavelet Transform have been studied extensively and been experimented with. The next few pages discuss HAS model, Spread Spectrum watermarking and few algorithms based on DWT.

2. Human Auditory System

The HAS perceives sounds over a range of power greater than 109:1 and a range of frequencies greater than 103:1. The sensitivity of the HAS to the additive white Gaussian noise (AWGN) is high as well; this noise in a sound file can be detected as low as 70 dB below ambient level. On the other hand, opposite to its large dynamic range, HAS contains a fairly small differential range, i.e. loud sounds generally tend to mask out weaker sounds. Additionally, HAS is insensitive to a constant relative phase shift in a stationary audio signal and some spectral distortions interprets as natural, perceptually non-annoying ones.

Two properties of the HAS dominantly used in watermarking algorithms are frequency (simultaneous) masking and temporal masking which are explained below.

2.1 Frequency Masking

Frequency (simultaneous) masking is a frequency domain phenomenon where a low level signal, e.g. a pure tone (the maskee), can be made inaudible (masked) by a simultaneously appearing stronger signal (the masker), e.g. a narrow band noise, if the masker and maskee are close enough to each other in frequency. A masking threshold can be derived below which any signal will not be audible. The masking threshold depends on the masker and on the characteristics of the masker and maskee (narrowband noise or pure tone).

2.2 Temporal Masking

In addition to 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 postmasking in time. The temporal masking effects appear before and after a masking signal has been switched on and off, respectively (Figure 1.2 b). The duration of the premasking is significantly less than one-tenth that of the post-masking, which is in the interval of 50 to 200 milliseconds. Both pre- and post-masking have been exploited in the MPEG audio compression algorithm and several audio watermarking methods.

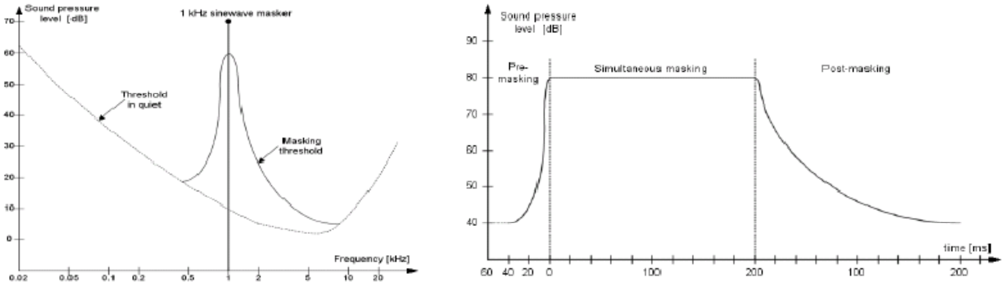


Fig 1.2 a. Frequency masking , b. Temporal masking.

3. Spread Spectrum Watermarking

In a normal communication channel, it is often desirable to concentrate the information in as narrow a region of the frequency spectrum as possible in order to conserve available bandwidth and to reduce power. The basic spread spectrum technique, on the other hand, is designed to encode a stream of information by spreading the encoded data across as much of the frequency spectrum as possible. This allows the signal reception, even if there is interference on some frequencies. While there are many variations on spread spectrum communication, we concentrated on Direct Sequence Spread Spectrum encoding (DSSS).

Let us denote as x the original signal vector to be watermarked. It represents a block of samples of the original audio signal. The corresponding watermarked vector is generated simply by:

$$y = x + w$$

, where the watermark w is has elements w(i) that can assume one of two equi-probable values, i.e. $w(i) \in \{-\Delta,+\Delta\}$, independently of x. Parameter Δ should be set based on the sensitivity of the HAS to amplitude changes. In our case, x is a vector of magnitude frequency components in a decibel scale, so Δ should not be higher than about 1 dB. A correlation detector performs the optimal test for the presence of the watermark

$$C = y \cdot w = (x + w) \cdot w = x \cdot w + N.\Delta^2$$

,where N is the cardinality of the vectors. Since the original audio file is completely uncorrelated to the vector w the product ‘x.w’ theoretically should turns out to be nearly zero. The optimal detection rule is to declare the watermark present if C > T. The choice of the threshold T controls the tradeoff between false alarm and detection probabilities.

In our experiments however, the watermarking method used was a little different from DSSS and is described below.

Firstly, the whole audio file is divided into partitions and one bit is hidden in one partition/block as follows:-

Vector x is considered to be a block of the original host signal. A secret key K is used by a pseudo random number generator (PRN) to produce a chip sequence with zero mean and whose elements are equal to +Δ or - Δ. Let this be denoted as sequence **u**. The sequence **u** is then added to or subtracted from the signal x according to the variable b, the data bit to be hidden in this block, where b assumes the values 1 or 0.

Hence embedding can be performed as:

$$y = x + (2b-1) * u$$

,where b takes value 1 or 0 (i.e. 2b-1 takes value +1 or -1 accordingly).

During watermark extraction, for each block y of the watermarked audio, y.u is calculated.

$$b' = y.u = (x + (2b-1)u).u = x.u + (2b-1).\ N.\Delta^2$$

Since x.u is nearly zero (x and u being uncorrelated) hence if b’ is positive, data bit hidden in that block is taken to be 1 else if the value is negative data bit hidden is taken to be 0.

3.1 Merits of SS

Spread Spectrum holds a lot of advantages against other watermarking methods. Some of them are:

1. Extraction of watermarks does not require the original audio file.
2. Watermark detection is exceptionally resilient to attacks that can be modeled as additive or multiplicative noise.
3. Watermarking algorithm has also been modified to become extremely resistant to dual-watermarking.

3.2 Demerits of SS

However this watermarking also has certain disadvantages which are:

1. The watermarked signal and the watermark have to be perfectly synchronized while computing.
2. The watermark is not robust against simple attacks like mp3 compression – decompression.
3. For a sufficiently small error probability, the vector length N may need to be quite large, increasing detection complexity and delay.

4. Discrete Wavelet Transformation

The wavelet transform is a new tool of signal processing in recent years. Its main character is that it can decompose signals into different frequency components and analyzes signal in the time domain and frequency domain simultaneously. So the wavelet transform is used widely in many fields of signal processing.

The discrete wavelet transform (DWT) of an audio signal f(n) is shown as Fig1.2. Here CA are the approximate components sub-band which mainly represents the low-frequency components of the audio signal, and CD are the detail component sub-band which mainly represents the high frequency component of the audio signal. Basically, CA is the audio signal f(n) after it has been passed through a low-pass filter and CD is obtained after f(n) is passed through a high-pass filter. If the components obtained after one-level of DWT are continued to be decomposed we obtain discrete wavelet transformation at different levels.

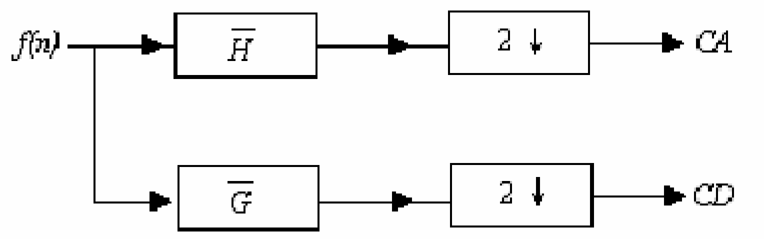


Fig 1.3 The DWT uses a high-pass filter (H) and a low-pass filter (G).

Since the hearing of human ears is not much sensitive to the minute changes in the high-frequency components and the coefficients of the high frequency component are smaller, so we can embed the watermarks into the sub-band of the high frequencies to realize the inaudibility of watermarks effectively. But the high frequency components can be easily destroyed by all kinds of common signal processing and hence robustness of the watermark cannot be ensured. However the low-frequency components are the main components of signal because its coefficients are bigger and they carry more energy. Hence most of watermarking algorithms that utilize DWT do 2 or 3 levels of discrete wavelet transformation and then embed in one of the divisions.

Some algorithms based on DWT are given in the following pages.

4.1 LSB Insertion

This watermarking method is very simple. A block of the original audio signal (say 512 samples) is taken and 2 or 3 levels of DWT is carried out on the block. All 512 wavelet coefficients are then scaled using the largest value inside the given sub-band and converted to binary arrays in two's complement form. A predetermined number of the LSBs are thereupon replaced with bits of information that should be hidden inside the host audio. Coefficients are then converted and scaled back to the original order of magnitude and inverse transformation is performed. This algorithm though robust against many compression-decompression attacks, fails against intelligent audio-processing. Since the watermarking method is simple enough, a attacker/pirate can simply randomize the LSB's in order to remove the watermark and still preserve signal quality.

4.2 Wavelet Quantized Index Modulation

This method is also similar to LSB insertion method. Firstly, a 2 or 3 level DWT of the original signal is carried out. The coefficients then obtained are quantized to different values in order to store a 1 or a 0. Inverse-DWT is carried out to get the watermarked audio. This method also suffers the same issues as the LSB-insertion method.

5. Experiments and Results

In all the experiments conducted, the following binary images were used as the watermarks. All the experiments were carried out using MATLAB which is used for signal processing the world over.

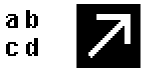
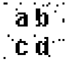
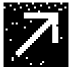
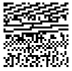
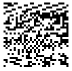


Fig 1.4 The watermark images used in experiments

5.1 SS watermarking

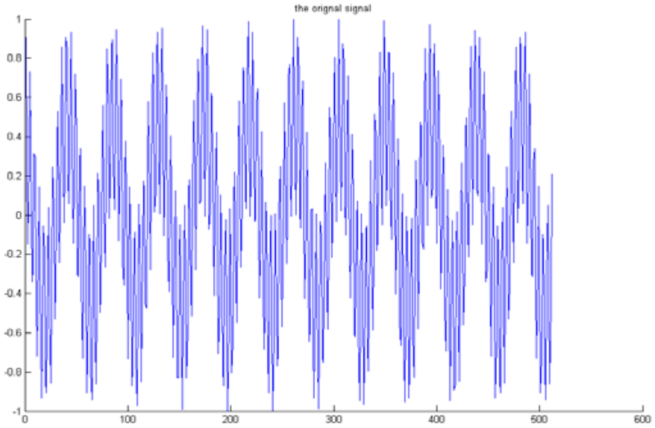
Firstly normal SS watermarking was experimented with. The algorithm is as follows. The audio file is partitioned into blocks of 512 samples. To hide a bit 1 in a block a sequence **u(|u| = 512)** is added to the block and to hide a 0 **u** is subtracted from the block. The performance of this algorithm was tested against 128-bit mp3 compression. The watermarks extracted are shown below. The results presented here are the most occurring result.

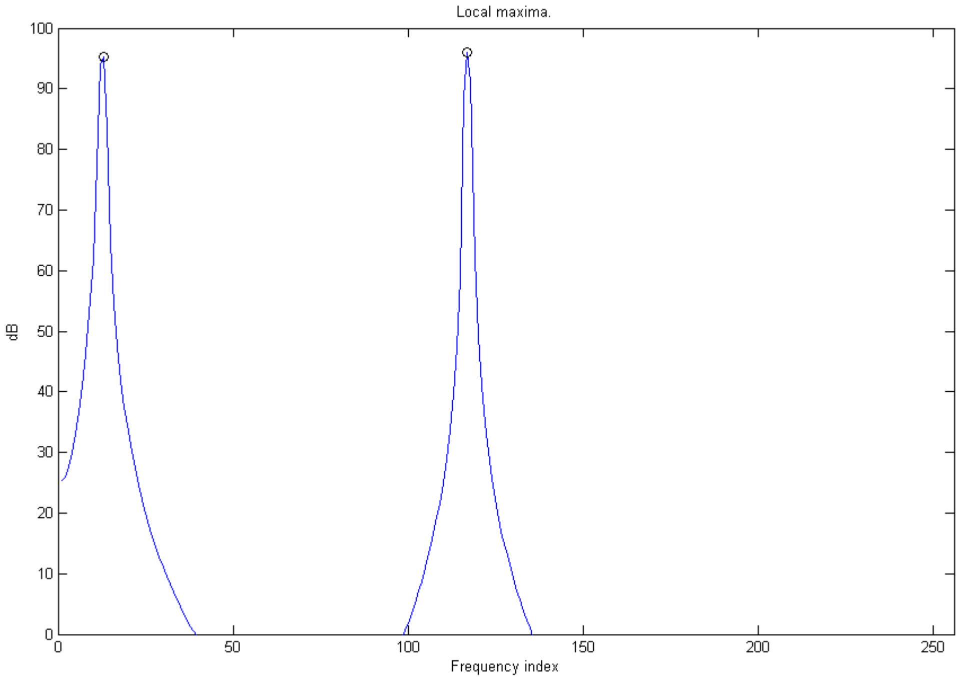
Watermarking Algorithm :	**Spread Spectrum**
SNR = 22.3534	**PSNR = 44.0827**
Extraction without attack:	
Image 1:  NC = 0.9757	Image 2:  NC = 0.9936
Extraction after mp3 attack:	
Image 1:  NC = 0.5464	Image 2:  NC = 0.5192

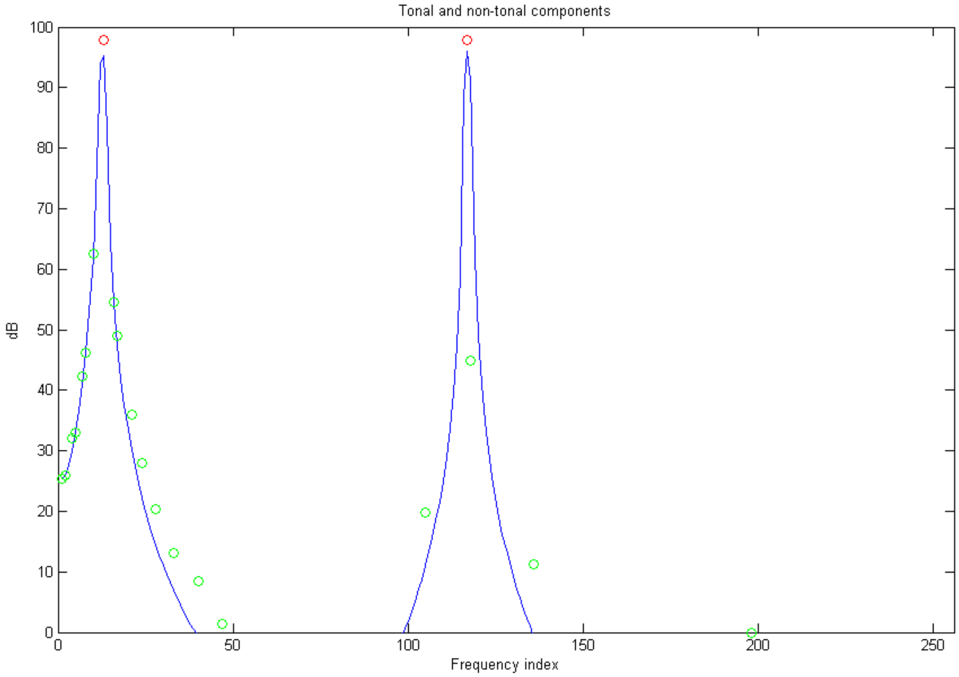
5.2 Spread Spectrum with HAS for watermark shaping

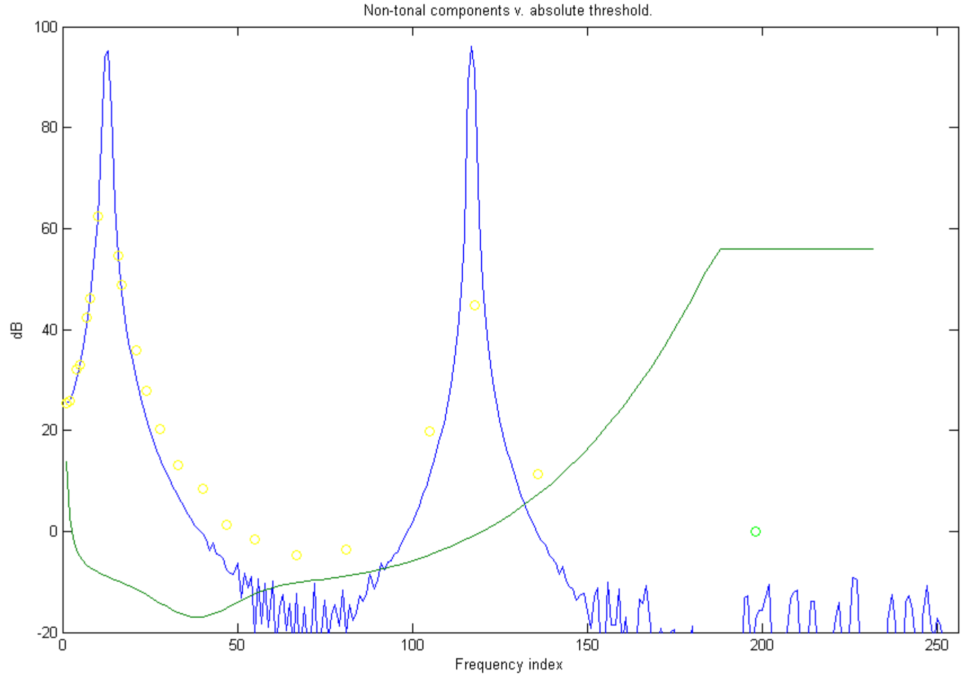
For this experiment, the mp3 compression-type algorithm was used to come up with a shaping of the sequence **u** before it is embedded into the original audio file. It was expected that such a watermark will be resistant to mp3 attack. The steps followed in this watermarking method have been explained below using images (which were generated during the embedding process). The audio file used simply consists of two sinusoidal waves. The results are in end.

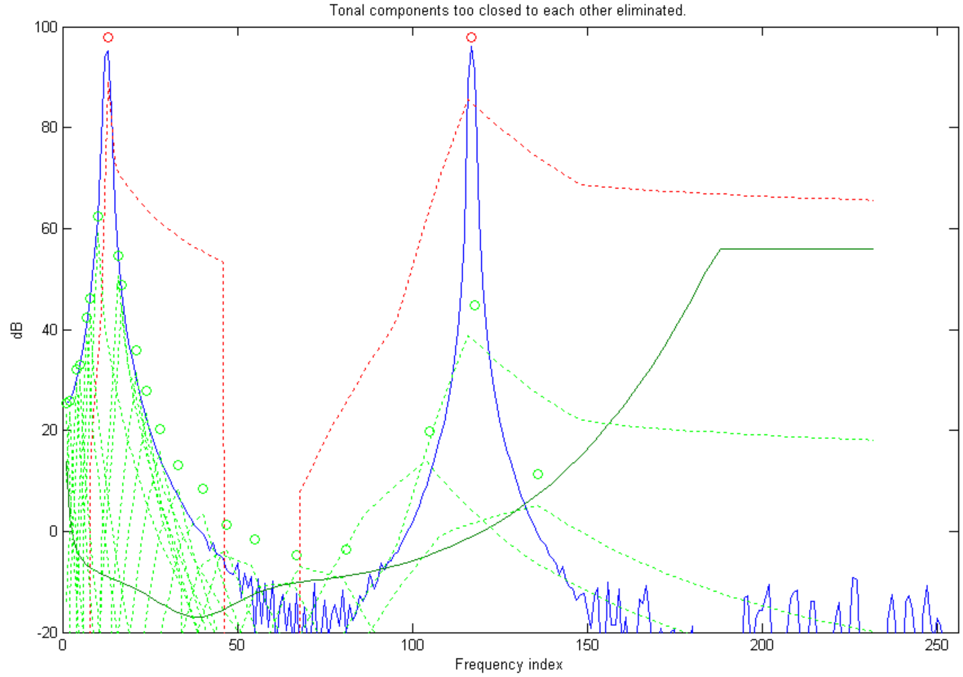
Step 1: Obtaining the masking thresholds for a block of the audio file using an algorithm similar to the MPEG compression model layer I.

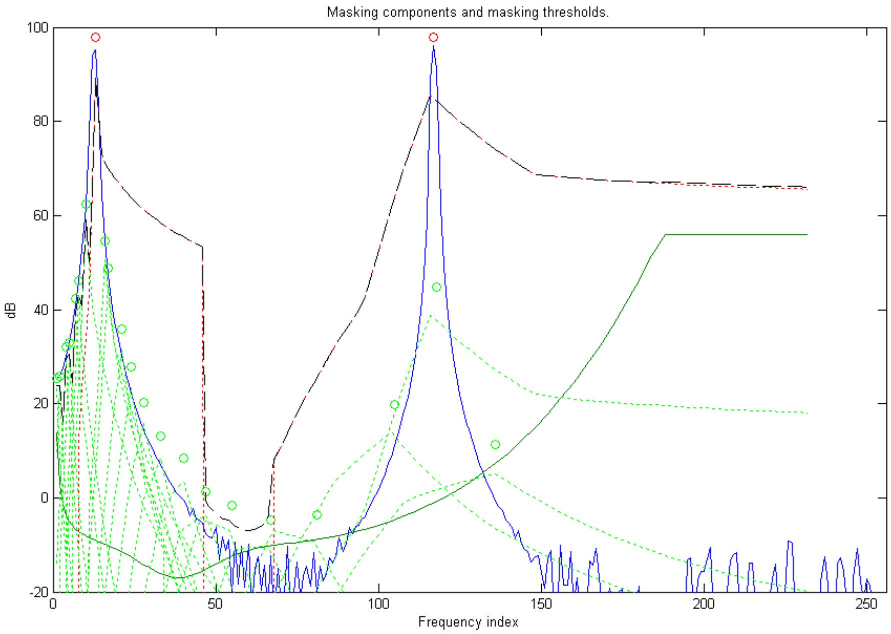



Local maxima.
100
90
80
70
60
50
40
30
20
10
0
dB
0
50
100
150
200
250
Frequency index

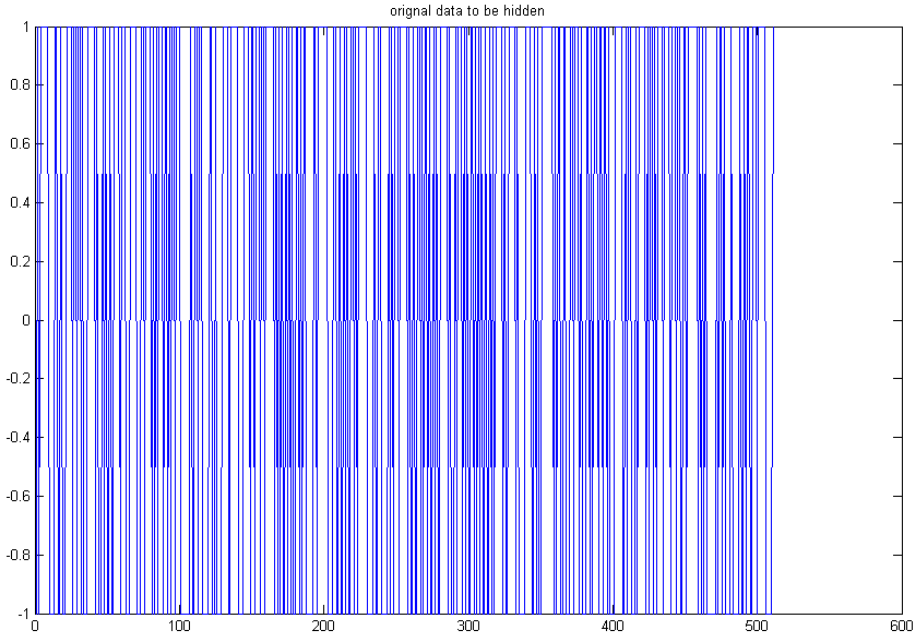

Tonal and non-tonal components
100
90
80
70
60
50
40
30
20
10
0
dB
0
50
100
150
200
250
Frequency index

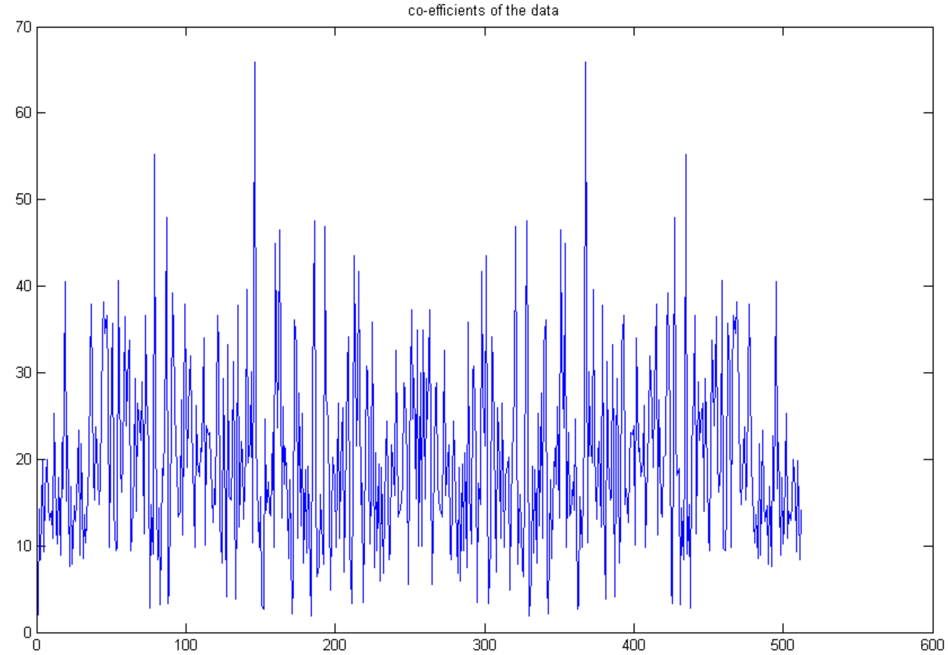

Non-tonal components v. absolute threshold.
dB
Frequency index
100
80
60
40
20
0
-20
0
50
100
150
200
250


Tonal components too closed to each other eliminated.
dB
Frequency index
100
80
60
40
20
0
-20
0
50
100
150
200
250

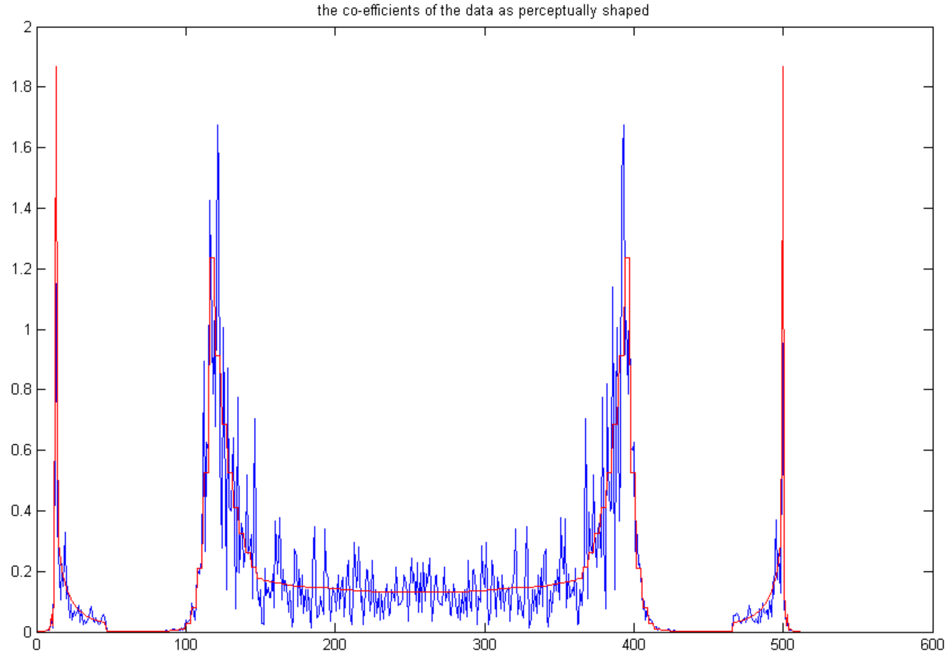


Step 2: Perceptually shaping the key/data that is to be hidden in the audio file.

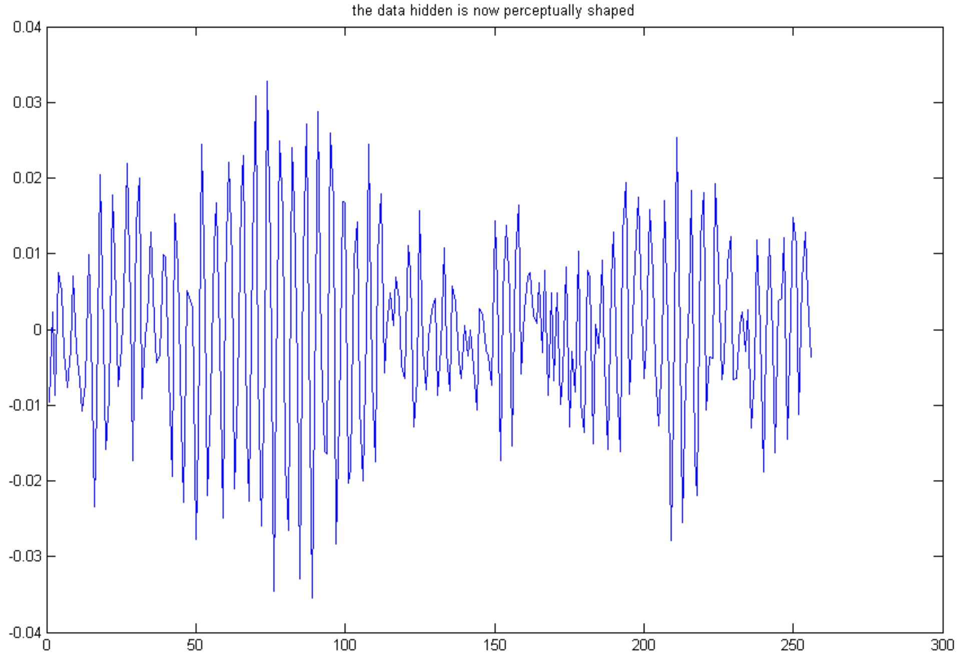


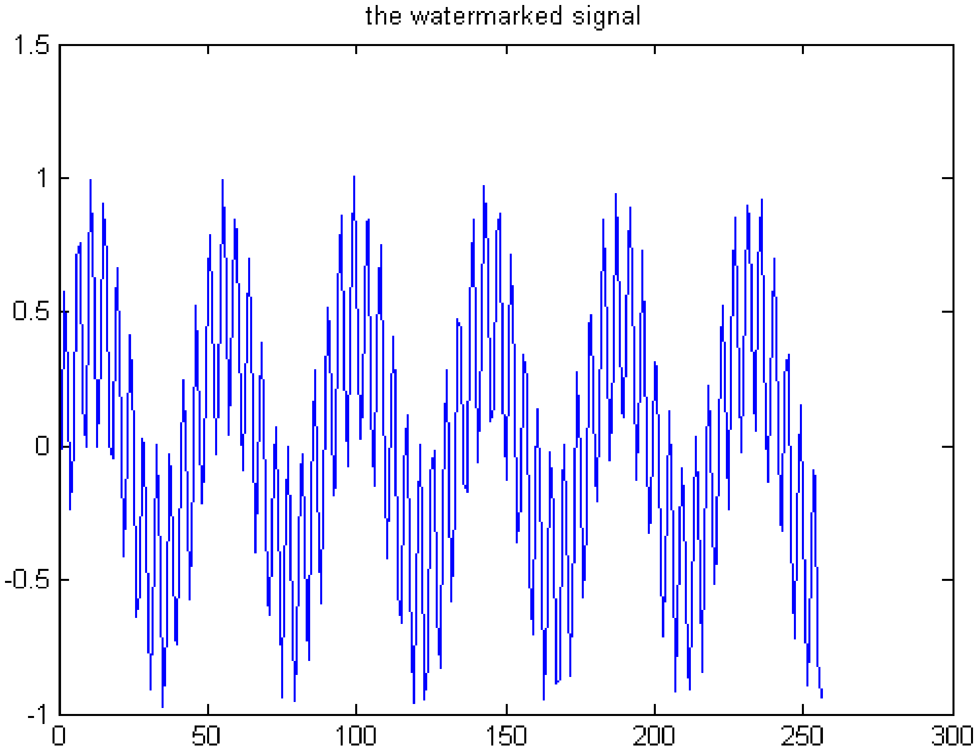


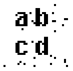
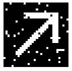
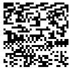
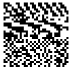
co-efficients of the data



the co-efficients of the data as perceptually shaped

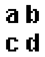
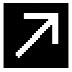

the data hidden is now perceptually shaped
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
0
50
100
150
200
250
300


the watermarked signal
1.5
1
0.5
0
-0.5
-1
0
50
100
150
200
250
300

Watermarking Algorithm :	**Spread Spectrum using HAS model**
SNR = 21.7277	**PSNR = 37.4954**
Extraction without attack:	
Image 1:  NC = 0.9713	Image 2:  NC = 0.9679
Extraction after mp3 attack:	
Image 1:  NC = 0.5077	Image 2:  NC = 0.5192

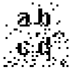
5.3 Spread Spectrum after DWT

In this experiment we tried to combine the two technologies i.e. Spread Spectrum and Discrete Wavelet Transformation so as to integrate the merits of two entirely different approaches into one watermarking algorithm. The original audio file is partitioned into blocks of 2048 samples and 1 bit is hidden per block. To hide a bit in a block, a two-level DWT of the block of the original audio is carried out. One of the portions of the transformation is taken and embedding like normal spread spectrum is carried out in it. After embedding, Inverse-DWT is done to obtain the watermarked audio file. The results of the experiment are as follows:

Watermarking Algorithm :	**Spread Spectrum after DWT**
SNR = 25.0606	**PSNR = 41.3857**
Extraction without attack:	
Image 1:  NC = 1.0000	Image 2:  NC = 1.0000

Extraction after mp3 attack:

Image 1:



NC = 0.8543

Image 2:

NC = 0.7244

5.4 Using two keys/sequences instead of one

In usual Spread Spectrum, only one sequence **u** is used to hide data. To hide bit 1, the sequence is added in a particular block of original audio signal. To hide a 0, the sequence **u** is subtracted from the block. It was thought by us that the use of two sequences **u_1** and **u_0** will improve the robustness of normal Spread Spectrum algorithm. To hide a 1, **u_1** will be added to the block of original audio file and to hide a 0, **u_0** will be added to the block. The reason being that since the two sequences **u_1** and **u_0** (produced independently) are uncorrelated hence this might improve robustness as bit transitions due to attacks will be reduced. The results however of this experiment were not satisfactory since the algorithm is still undergoing reformations.

6. References

[1] Nedeljko Cvejic, "ALGORITHMS FOR AUDIO WATERMARKING AND STEGANOGRAPHY," Oulu 2004.

[2]. Darko Kirovski and Henrique S. Malvar, "Spread-Spectrum Watermarking of Audio Signals," in IEEE transactions on Signal Processing , VOL.51, NO.4, APRIL2003.

[3] NedeGko Cvejic, Tapio Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography," IEEE 2002.

[4] Mitchell D Sanson, Bin Zhu, Ahmed H Tewk and Laurence Boney, "Robust audio watermarking using perceptual masking," 1996.

[5] P. Bassia and I. Pitas, "Robust audio watermarking in the time domain," Proc. EUSIPCO 98, ol. 1, pp. 25–28, Rodos, Grece, Sept. 1998.

[6] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in Information Hiding, Springer Lecture Notes in Computer Science, v1174, pp. 295–315, 1996.

[7] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "A secure robust watermark for multimedia," information Hiding Worshop, Univ. of Cambridge, pp.185–206, 1996.

[8] C. Neubauer and J. Herre, “Digital watermarking and its influence on audio quality,” Proc. 105th Convention, Audio Engineering Society, San Francisco, CA, Sept. 1998.

[9] M.D. Swanson, B. Zhu, A.H. Tewfik, and L. Boney, “Robust audio watermarking using perceptual masking,” Signal Processing, vol.66, pp. 337–355, 1998.

[10] B. Chen and G. W. Wornell, “Digital watermarking and Information embedding using dither modulation,” Proc. IEEE Workshop on Multimedia Signal Processing, Redondo Beach, Cpp. 273–278, Dec. 1998.