

# DIGITAL AUDIO WATERMARKING USING PSYCHOACOUSTIC MODEL AND SPREAD SPECTRUM THEORY

Manish Neoliya

School of Electrical and Electronics Engineering

Nanyang Technological University

Singapore-639789

Email: [neoliya@pmail.ntu.edu.sg](mailto:neoliya@pmail.ntu.edu.sg)

## Abstract

There is a need of an effective watermarking technique for copyright protection and authentication of intellectual property. The project proposes a watermarking technique which makes use of simultaneous frequency masking to hide the watermark information. The algorithm is based on Psychoacoustic Auditory Model and Spread Spectrum theory. It generates a watermark signal using spread spectrum theory and embeds it into the signal by measuring the masking threshold using Psychoacoustic Auditory model. Since the watermark is shaped to lie below the masking threshold, the difference between the original and the watermarked copy is imperceptible. Recovery of the watermark is performed without the knowledge of the original signal. A software system is implemented using MATLAB and the characteristics studied.

## Introduction

In today's world every form of information, be it text, images, audio or video, has been digitized. Widespread networks and internet has made it easier and far more convenient to store and access this data over large distances. Although advantageous, this same property threatens the copyright protection.

Media and information in digital form is easier to copy and modify, and distribute with the aid of widespread internet. Every year thousands of sound tracks are released and within a few days are readily available on the internet for download. Without any information on the track itself, it's easy for some one to make profit out of them by modifying the original and selling under a different name. As a measure against such practices and other intellectual property rights, digital watermarking techniques can be used as a proof of the authenticity of the data.

*Digital watermarking* is the process of embedding or inserting a digital signal or pattern in the original data, which can be later used to identify the author's work, to authenticate the content and to trace illegal copies of the work.

Some of the requirements of the digital watermarking are:

- The original media should not be severely degraded and the embedded data should be minimally perceptible. The words hidden, inaudible, imperceptible, and invisible mean that an observer does not notice the presence of the hidden data.
- The hidden data should be directly embedded into the media, rather than into the header of it.
- The watermark should be robust. It should immune to all types of modifications including channel noise, filtering, re-sampling, cropping, encoding, lossy compressing, printing and scanning, digital-to-analog (D/A) conversion, and analog-to-digital (A/D) conversion, etc.
- It should be easy for the owner or a proper authority to embed and detect the watermark.
- It should not be necessary to refer to the original signal when extracting a watermark.

Compared with the image and video watermarking, digital audio watermarking is especially challenging, because the human auditory system (HAS) is extremely more sensitive than Human Visual System (HVS). There are many methods we can use to embed audio watermarks. Currently, audio watermarking techniques mainly focus on four aspects: *low bit coding*, *phase coding*, *spread spectrum-based coding* and *echo hiding*.

In this project I have used principles of Spread Spectrum Theory and Psychoacoustic Auditory Model to embed the watermark. The watermarking system uses an algorithm that relies on the above principles, to generate a digital watermark. (bit-stream, character string, etc.), and embed it into the original audio file (which is in .wav format) by spectrally shaping the watermark signal.

## Psychoacoustic Auditory Model

The psychoacoustic auditory model is an algorithm to imitate the Human Auditory System (HAS). HAS is sensitive to a very wide dynamic range of amplitude of one billion to one and of frequency of one thousand to one. It is also acutely sensitive to the additive random noise. These properties of HAS makes it all the more challenging to tamper with any kind of audio signal.

However, there are a few “holes” in the auditory system. While the HAS has a very large dynamic range, it has a very small differential range. This is called the masking effect of HAS. There are two types of masking observed in the HAS – *frequency masking* and *temporal masking*.

In this project, simultaneous frequency masking is closely studied to be used for watermark shaping purposes. The auditory model processes the audio information to produce information about the final masking threshold. The final masking threshold information is used to shape the generated audio watermark. This shaped watermark is ideally imperceptible for the average listener. To overcome the potential problem of the audio signal being too long to be processed all at the same time, and also extract quasi-periodic sections of the waveform, the signal is segmented in short overlapping segments, processed and added back together. Each one of these segments is called a “frame.” The steps involved in creating a *Psychoacoustic Auditory Model* include:

- Fast Fourier Transform
- Power Spectra
- Energy per critical band
- Spread masking
- Masking threshold Estimation

Figure 1 shows an example of threshold estimation (Blue line above the threshold).

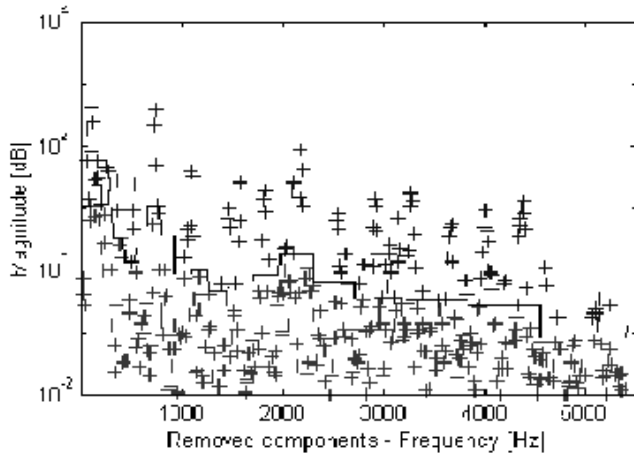


Figure 1 Threshold  $T(z)$ .

## Noise Shaping using Masking Threshold

The components that lie below the masking threshold are imperceptible to the human ear. Those components can be conveniently discarded without losing the perceptual quality of the audio signal. Not only that, these components can even be replaced by other information. This is the idea implemented while embedding the watermark into the audio signal. The signal component whose power lies below the threshold are discarded and replaced by the respective noise (watermark in

this case) component. After replacing the signal components with the noise components, shaping needs to be done in order to contain the noise below the threshold. Figure 2, 3 and 4 show the process of removal of components and shaping of the watermark. It clearly emphasizes upon the importance of shaping.

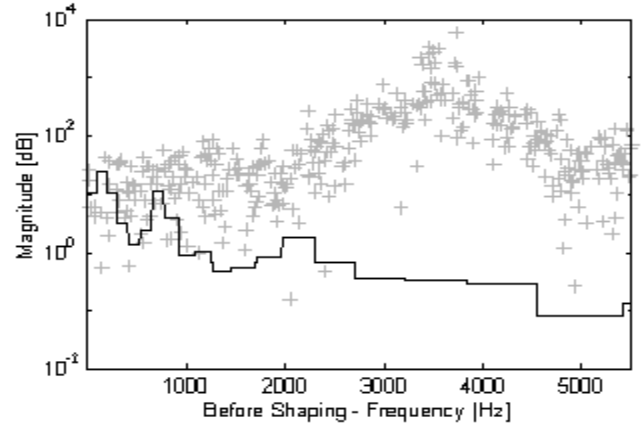


Figure 2: Watermark before shaping.

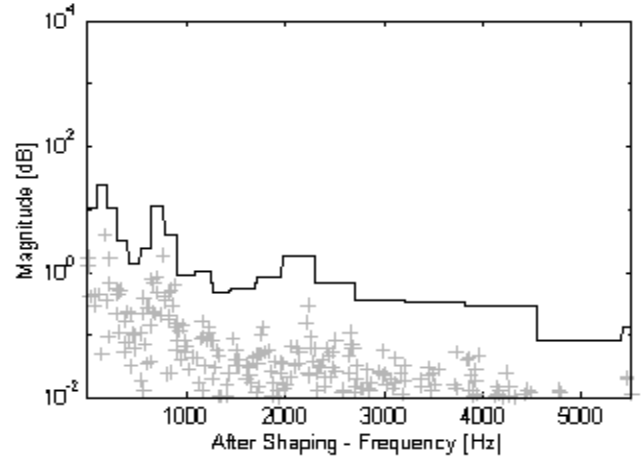


Figure 3: Watermark power spectra after shaping

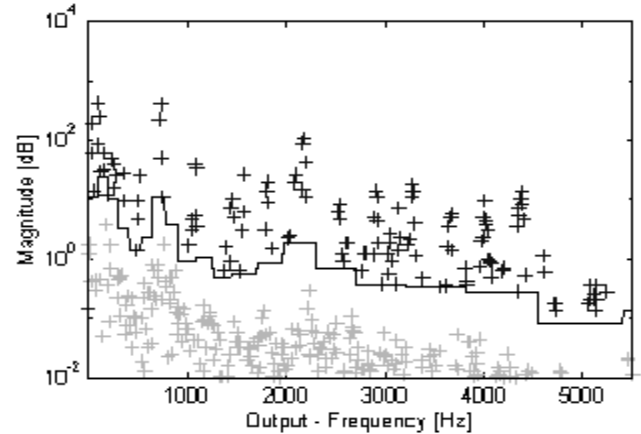
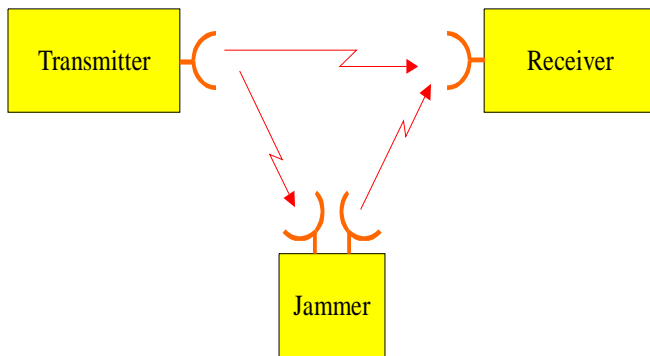


Figure 4: The output after adding the audio and watermark components.

## Spread Spectrum Theory

Spread spectrum is a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by means of a code which is independent of the data, and a synchronized reception with the code at the receiver is used for de-spreading and subsequent data recovery. The process of watermark embedding can be viewed as intentional jamming of the watermark signal with the music or the audio signal. In this case the signal (watermark) has much less power than the jammer (music). It is one of the problems to be overcome at the receiver end. The following analysis expresses the process of watermark generation in spread spectrum terminology. The approach selected in this algorithm is the Direct Sequence Spreading. Figure 3 shows the basic spread spectrum communication system.



### Direct Sequence Spreading

Coherent direct-sequence systems use a pseudorandom sequence and a modulator signal to modulate and transmit the data bit stream. The main difference between the uncoded and coded versions is that the coded version uses redundancy and “scrambles” the data bit stream before the modulation is done and reverses the process at the reception. The watermarking algorithm uses the coded scheme.

### Uncoded Direct Sequence Spread Binary Phase-Shift-Keying (DS/BPSK)

#### Coded Direct Sequence Binary Phase-Shift-Keying

Coded DS/BPSK differs from the uncoded DS/BPSK in the sense that it uses redundancy and scrambles. There are two more steps each at the transmitter and the receiver. At the transmitter, the watermark code is first repeated the specified number of times to generate a repeat code. This repeat code is then scrambled or interleaved using an interleaver matrix. This scrambled code is then modulated using the uncoded DS/BPSK.

At the receiver end, the data received is first de-scrambled by passing it through de-interleaver and then decode using hard

The figure below sums up the process of Uncoded DS/BPSK. The system modulates input the data bit-stream with the help of Pseudorandom Number sequence and modulator signal which is usually a cosine function of time with the centre frequency,  $f_0$ .

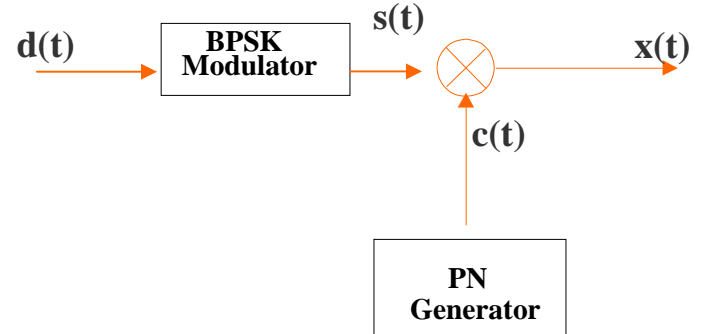


Figure 5: Uncoded DS/BPSK modulation.

In the figure,  $d(t)$  is the input bit-stream,  $s(t)$  is the amplitude modulated signal, and  $x(t)$  is the signal output which has been spread over the entire bandwidth. The figure that follows clearly illustrates the process of spreading the data.

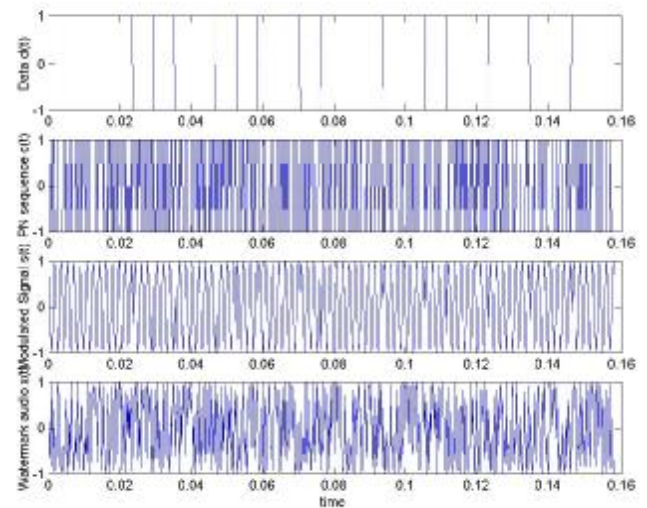


Figure 6: Uncoded DS/BPSK

decision rule. The output of this decoder is the recovered watermark.

### Despreading and data recovery

At the receiver end, the signal is multiplied with the PN sequence. The property of the PN sequence,  $c(t)^2 = 1$ , is utilized here. Product  $r(t)$  is then demodulated and integrated to get the output signal. This output signal is then put through de-interleaver and decoder to recover the watermark code. The figure below summarizes the process.

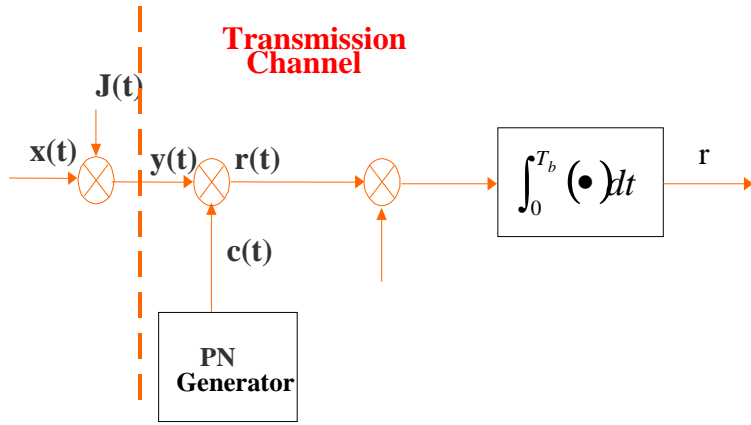


Figure 7: Despreading

## Proposed Watermarking System

The watermarking algorithm used in this project mixes the psychoacoustic auditory model and the spread spectrum communication technique to achieve its objective. It is comprised of two main steps: first, the watermark generation and embedding and second, the watermark recovery. The watermark generation and embedding process is shown in Figure 8. A bit stream that represents the watermark information is used to generate a noise-like audio signal using a set of known parameters to control the spreading. At the same time, the audio (i.e. music) is analyzed using a psychoacoustic auditory model. The final masking threshold information is used to shape the watermark and embed it into the audio. The output is a watermarked version of the original audio that can be stored or transmitted.

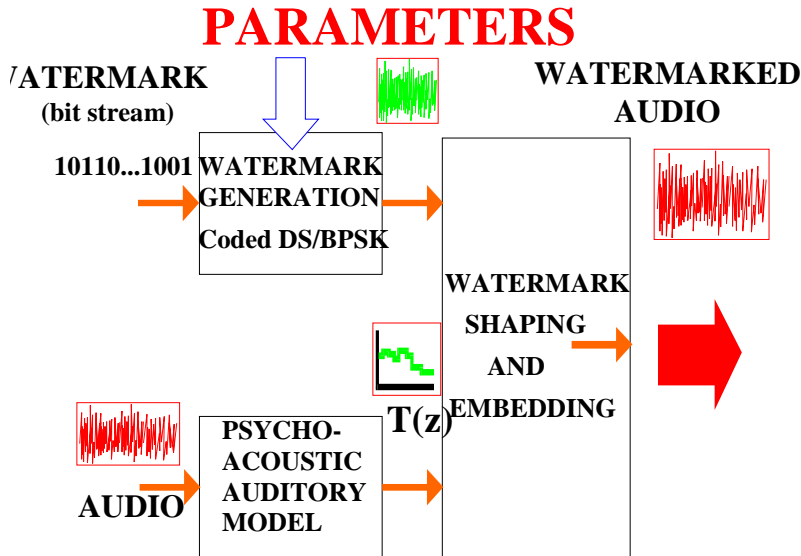


Figure 8: Watermark embedding and transmission

## Watermark Generation and Embedding

The watermark is generated from a specified watermark code through following steps;

- Watermark repeat code generation
- Interleaving
- Header concatenation
- Spreading using coded DS/BPSK

The list of spread spectrum parameters required is:

- $R_d$  = is the data bits per second
- $m$  = is the repetition code factor
- $N$  = is the spreading factor.
- $R_b = R_d * m$  is the coded bits per second
- $T_b = 1/R_b$  is the time of each coded bit
- $R_c = N * R_b$  is the PN sequence bits per second
- $T_c = T_b / N$  is the time of each PN bit or “chip”

## Threshold estimate

The threshold of the audio signal is estimated using the Psychoacoustic Auditory Model and based on that the spectral shaping of the watermark is done. The resulting watermark signal is then added to the modified audio signal to get the watermarked audio signal as the output. Figure below shows the experimental result of the watermarking algorithm.

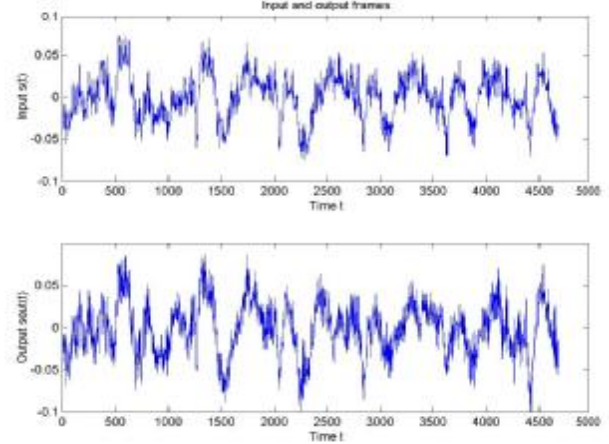


Figure 9: Input (original) and output (watermarked) audio frame

## Watermark recovery

The watermark recovery is shown in Figure 10. The input is the watermarked audio after transmission (i.e. music + noise, low quality, etc). An auditory psychoacoustic model is used to generate a residual. At the same time as the known parameters are used to generate the header of the watermark. Using an adaptive high-resolution filter, all the residual is scanned to find all the occurrences of the known header and therefore the initial position of each possible watermark. After this, the

same known parameters used to generate the header are used to de-spread and recover the watermark.

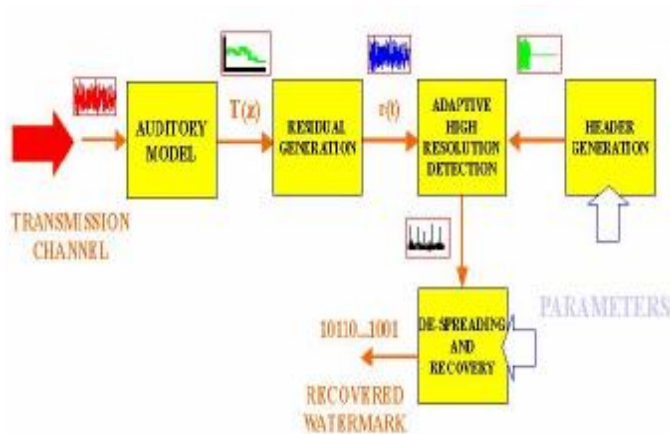


Figure 10: Watermark recovery

## Discussion

After the whole process it is realized that not all the watermarks could be recovered with 100% accuracy. This occurs because of the multiple factors that affect the quality of the embedded watermark, such as: the number of audio components replaced the gain of the watermark, and the masking threshold. It is important to note that in some frames the watermark information can be very weak, even null. The spread spectrum technique employed can partially solve these problems, but if many consecutive frames have no watermark information, that specific watermark can not be recovered.

An important thing to notice is that the watermark shaping differs for different kinds of music. The figure below shows the threshold of two types of music. The line represents the threshold of a Heavy metal song, “Feuer Frei” by Rammstein. The Bold line is for MLTR song “More than just friends”.

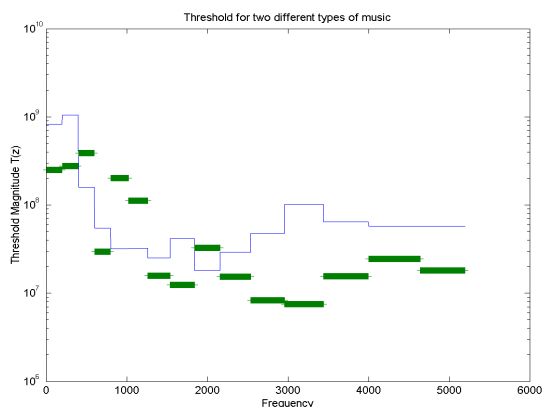


Figure 11: Threshold for different kinds of music. Line represents a heavy metal song while bold segments represent soft music.

The percentage of correct bits recovered measures the quality of the recovery for each watermark. It is noticed that not all the watermark bits are recovered, and not all the watermarks are recovered in their totality. In out case, the maximum percentage of the correct bits was 71% of the bits. A good bit error detection/correction algorithm or averaging technique could substantially improve the recovery of the watermark.

Another thing that is noticed is that after converting it into mp3 format and then reconverting it, the watermark retrieval isn't accomplished. This might be because some encoders tend to shift the audio signal in time domain and hence synchronization fails. Also in the proposed system, a rectangular window is used. In real time system, this might lead to Gibbs phenomenon after the frames are added together.

## Conclusion

The proposed digital watermarking system for audio signal is based on psychoacoustic auditory model which is used to shape the watermark generated using spread spectrum techniques. To recover the watermark, the original signal is not required. The only information necessary is the parameters that are selected originally for the watermark generation. The method retains the perceptual quality of the audio signal.

Further research can be done on the performance of the watermarking technique on various kinds of music. The resistance of the watermark to various kinds of attack like frequency shifting and re-sampling should be looked into. The effect of the spread spectrum parameter on the watermark generation and recovery needs to be studied in detail. Also, proper bit error detection and recovery could enhance the watermark reliability.

## Acknowledgement

I would like to express my sincere gratitude to A/P Anthony Ho for his support and guidance throughout the project. I would also like to thank the graduate and post-graduate students in the Computer Engineering Lab for their timely assistance during the project.

## References

1. Ricardo A. Garcia, Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory, University of Miami, School Of Music, 1999.
2. Yuval Cassuto, Michael Lustig and Shay Mizrachy, Real Time Digital Watermarking System for Audio Signals Using Perceptual Masking, Internal Report, Israel Institute of Technology, 2001.
3. E. Zwicker and U. T. Zwicker, “Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System,” *J. Audio Eng. Soc.*, vol. 39, pp. 115 -126 (1991 March)