Chapter 3

Audio Watermarking using Spectral Modifications

3.1 Introduction

In this chapter, watermark is multiplicatively embedded in discrete fourier transform (\textit{DFT}) magnitude of audio signal using spread spectrum (\textit{SS}) based technique. A new perceptual model for magnitude of \textit{DFT} coefficients is developed which finds the regions of highest watermark embedding capacity with least perceptual distortion. Theoretical evaluation of detector performance using correlation detector and likelihood ratio detector is undertaken under the assumption that host feature (\textit{DFT} magnitude) follows Weibull distribution. Also, experimental results are presented in order to show the performance of the proposed scheme under various attacks such as presence of multiple watermarks, additive white gaussian noise and \textit{MP3} compression.

In \textit{SS} based watermarking schemes, watermark embedding is mostly preferred in transform domain (e.g. \textit{DFT}, \textit{DCT}, \textit{DWT} etc) of audio in order to obtain high robustness. Numerous watermarking techniques [69] - [81] in literature, as discussed in section 2.4, exploit psychoacoustic characteristics of \textit{HAS} while embedding the watermark additively or multiplicatively in spectral domain. These techniques explored the fact that \textit{HAS} is insensitive to small amplitude changes in spectral domain. Whereas, phase discontinuity of an audio signal causes perceptible distortion when the phase relation between each frequency component of the signal is changed. Hence \textit{DFT} magnitude would be a better option for inserting watermark. However, in literature
no perceptual model is defined for DFT magnitude which can decide the location and strength of watermark to be embedded in audio spectrum. Also, these techniques have two major drawbacks. First, the psychoacoustic modeling used by existing techniques require rigorous complex computations. Second, the watermark embedding capacity of these schemes is low i.e. there is not much space to accommodate watermark in the host feature within the defined perceptual limits.

To overcome these two problems, a new method of evaluating masking threshold for DFT magnitude is proposed which requires lesser computations as compared to traditional psychoacoustic model based thresholds. The technique finds best possible locations in spectra for watermark embedding and finds scaling factor of watermark to achieve high watermark embedding capacity.

Existing techniques generally employ statistical characterization of the host signal to develop an optimal or suboptimal watermark detector. Kalantari [88] and Akhaee [89] obtained optimal performance of watermark detector from Maximum Likelihood decoder. Similarly, correlation detector which is the Maximum Likelihood (ML) optimal detector, is applied in majority of watermarking techniques [87]. These two detectors are optimal and minimize the error probability only in cases when the host feature follows a Gaussian distribution. However, generalized Gaussian distribution is not suited to describe positive valued random variables such as DFT magnitude. The best suited distribution for magnitude of discrete fourier transform (DFT) coefficients is Weibull distribution which is defined for positive real axis only.

### 3.2 Discrete Fourier Transform (DFT)

DFT is used to calculate the spectrum of a waveform in terms of a set of harmonically related sinusoids, each with a particular amplitude and phase. This transform is most commonly used in audio signal processing, as it has the Fast Fourier Transform (FFT) algorithm to increase the processing speed. Also, it is an important representation of audio data because human hearing is based on a kind of real-time spectrogram encoded by the cochlea of the inner ear. Spectrogram is a sequence of FFTs of windowed audio segments.

The angular frequencies of these sinusoids are represented by $\omega_k = k\omega$, where $k$ is an integer varying from 0 to $N - 1$ and $\omega = 2\pi f_s / N = 2\pi / NT$. Here $f_s = 1/T$ denotes
sampling frequency of discrete time signal \( s \) given as

\[
s = [s(t_0), s(t_1), \cdots, s(t_{N-1})] = [s(0), s(T), \cdots, s((N-1)T)]
\]  \hspace{1cm} (3.2.1)

For convenience \( s(nT) \) is often written as \( s(n) \) in literature. The \( k^{th} \) component of DFT, \( S(k) \), of signal \( s(n) \) is given as

\[
S(k) = \sum_{n=0}^{N-1} s(n) e^{-j\omega_k t_n} \\
= \sum_{n=0}^{N-1} s(n) e^{-j(2\pi k/T)(nT)} \\
= \sum_{n=0}^{N-1} s(n) e^{-j2\pi kn/N}
\]  \hspace{1cm} (3.2.2)

The samples of discrete time signal \( s(n) \) is recovered using the inverse discrete Fourier transform of \( S(k) \) as,

\[
s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k)e^{j2\pi kn/N}
\]  \hspace{1cm} (3.2.3)

3.3 Description of Watermarking Model

A watermarking system encompasses three major functionalities, namely, watermark generation, watermark embedding, and watermark detection, as shown in figure 3.1. The aim of watermark generation is to construct a sequence \( W \) using an appropriate function \( f \). Hence the watermark vector \( W = [W(0), W(1), \cdots, W(N-1)] \), such that \( W(i) \in R \), where \( R \) is real number, is given as

\[
W = f(K, N)
\]  \hspace{1cm} (3.3.4)

here \( K \) is the watermark key, \( N \) is the length of watermark. Watermarked feature \( F' \) is obtained by multiplicatively embedding watermark \( W \) in host feature \( F \) given as

\[
F' = F(1 + aW)
\]  \hspace{1cm} (3.3.5)

here \( F' = [F'(0), F'(1), \cdots, F'(N-1)] \) and \( a \) is the scaling factor lying between 0 and 1. The scaling factor is introduced to maintain imperceptibility of the distortions caused to the host signal due to watermarking.
Watermark detector is used to examine whether the signal under test $F_t$ contains a watermark $W$ or not under a binary-decision hypothesis test framework. Each module is now discussed in detail, in the following subsections.

### 3.3.1 Watermark Generation

The steps required for generation of watermark are as follows:

- To construct watermark $W$, a white pseudo-random (PN) sequence or chip $W_0 = [W_0(0), W_0(1), \ldots, W_0(N_w - 1)]$, where $W_0(i) \in (-1, 1)$, is generated. The sequence is generated using secret key $K$ such that they are mutually independent with respect to the host signal.

- The magnitude nature of host feature needs to be preserved implying that $F'$
given in (3.3.5), should always be greater then zero. Such condition is obtained when \(aW(i)'s\ \forall \ 0 \leq i \leq N - 1\) take the value in the finite interval \([-1,1]\) keeping scaling factor \(a \leq 1\).

- The \(N\) point \(DFT\) region hosting the watermark is usually split in number of subregions, which in our case are the critical bands. The start location \((m)\) and end location \((n)\) of watermark embedded in these critical bands is decided by a pre-defined masking threshold. Hence the length of watermark \(N_w\) is evaluated as

\[
N_w = \lceil (n - m)N \rceil \tag{3.3.6}
\]

- To maintain the symmetry of \(DFT\) magnitude a reflected version of \(W_0\) is required to be generated as

\[
W'_0(i) = W_0(N_w - i - 1), \quad 0 \leq i \leq N_w - 1 \tag{3.3.7}
\]

The reflected chip \(W'_0\) is embedded in the frequency components around coefficient \(N - 1\). This is essential to obtain real valued audio in time domain.

### 3.3.2 Masking threshold for DFT magnitude

In this work, the magnitude of \(DFT\) coefficients of host audio signal are modified by adding watermark, such that the modified spectra is always below the predefined masking threshold, termed as maximum amplitude spread (MAS). The MAS is defined as the maximum of all amplitude spreads (AS) of \(DFT\) components at a particular frequency location within a frame. Following steps are involved to find MAS.

**STEP-I: Finding Amplitude Spread (AS)**

The AS of \(DFT\) components is evaluated from the energy spreading function given by Schroeder [103] and its effect is seen at all the \(N\) frequency locations of a frame. Schroeder presented a real nonnegative energy spreading function which approximated the basilar spreading as a triangular spreading function and is given as

\[
SF_{dB}(i, j) = 15.81 + 7.5(\Delta_z + 0.474) - 17.5\sqrt{1 + (\Delta_z + 0.474)^2}, \tag{3.3.8}
\]

here \(SF_{dB}(i, j)\) is the energy spread in decibels (dB) from \(i^{th}\) to \(j^{th}\) frequency location. The bark separation between these two points is given as \(\Delta_z = z_j - z_i\), where \(z_i\) and
$z_j$ denote the bark frequencies of $i^{th}$ and $j^{th}$ frequency locations respectively. Let the audio signal $s$, given by (3.2.1), is sampled at frequency, $f_s$ Hertz (Hz). Since audio is real valued signal, its DFT will satisfy the symmetry property i.e. $S(k) = S(N - k)^*$, where $k = 1, ..., N/2 - 1$. The DFT coefficients $S(k)$ corresponds to frequencies $f_k$ given as

$$f_k = f_s \times k/N,$$

(3.3.9)

here $0 \leq k \leq N - 1$, $N$ being a power of 2. Considering the duplication in the spectra for $k \geq N/2$, we evaluate the masking spread $A_1(i, j)$ for amplitude of $N/2$ components only, given as

$$A_1(i, j) = \sqrt{SF(i, j)} , \quad 0 \leq i \leq N/2 - 1$$

(3.3.10)

where $SF(i, j)$ is the inverse decibel of $SF_{dB}(i, j)$. The square root is to convert the masking spread from energy scale to amplitude scale. Now respecting the symmetry property of DFT components, we define $A(i, j)$ as,

$$A(i, j) = \begin{cases} A_1(i, j), & 0 \leq j \leq N/2 \\ A_1(i, N - j), & N/2 + 1 \leq j \leq N - 1 \end{cases}$$

(3.3.11)

The amplitude spread of $i^{th}$ DFT component is then defined as,

$$A'(i, j) = |A(i, j)S(i)|, \quad for \quad 0 \leq i \leq N/2 - 1, 0 \leq j \leq N - 1,$$

(3.3.12)

where $S(i)$ is given by (3.2.2). This gives $N/2 \times N$ matrix showing amplitude spread of each of the $N/2$ DFT components at $N$ frequency locations. Figure 3.2 shows a plot of Amplitude spread $A'(i, j)$ of $i = 17^{th}$ and $20^{th}$ frequency components at all the frequency location $f_j$ for $0 \leq j \leq N - 1$ given in (3.3.9) where $N = 512$ and $f_s = 44.1kHz$.

**STEP II: Evaluation of Maximum Amplitude Spread (MAS)**

The amplitude spreads of neighboring DFT components overlap each other. Maximum amplitude spread (MAS) is the maximum of all the overlapping amplitude spreads at $f_i$ frequency due to DFT coefficients $S(j)$, $\forall 0 \leq j \leq N/2 - 1$ and $j \neq i$. $MAS, Y(i)$, at location $i$ can therefore be evaluated as

$$Y(i) = \max(A'(i, j)) \quad for \quad 0 \leq j \leq N/2 - 1$$

(3.3.13)
Figure 3.2: Overlapping of Amplitude Spread of 17th and 20th DFT components and magnitude of 19th DFT component.

Now maximum amplitude spread $MAS$ for a critical bands $z$ will be the minimum of all $Y(i)$ in that critical band. From (3.3.13), we evaluate the Maximum Amplitude Spread $Y(z)$ for critical bands $z = 1, 2, \cdots, z_t$ as

$$Y(z) = \min(|Y(i)|) \text{ for } LB_z \leq i \leq HB_z,$$

(3.3.14)

where $LB_z$ and $HB_z$ are lower and upper frequency components of $z^{th}$ critical band. Figure 3.3 shows the plot between maximum amplitude spread $Y(i)$ and the magnitude of DFT coefficients $F(i)$ at all the frequency locations $f_i$ for $i = 0, 1, \cdots, N-1$.

### 3.3.3 Watermark Embedding

In watermark embedding the watermark $W$ is added to host signal $F$ in a way that the symmetry of $F$ is not disturbed. Also, the dc component and nyquist component of $DFT$ spectrum should remain unchanged. This is essential in order to retrieve real valued audio signal after watermarking process. The magnitude of $DFT$ coefficients of host audio signal are modified by multiplicative watermarking, such that the modified
spectra is always below the maximum amplitude spread of original signal. Hence, the DFT magnitudes are modified only in certain critical bands to maintain the transparency of audio signal. The embedding steps are described as follows

- The original audio is first segmented into frames of size $N = 512$ samples. Let $s(i)$, $i = 1, 2, \cdots, N$, be the samples of any one frame of host audio signal $s$.

- Let $S(k)$, $k = 1, 2, \cdots, N$, be the spectral coefficients of $s(i)$, $i=1, 2, \cdots, N$. These coefficients are obtained by directly taking N point non-overlapping DFT of given audio segment.

- The magnitude $F(k) = |S(k)|$ and phase $\phi(k) = \angle S(k)$ of the spectral coefficients are evaluated for $k = 1, 2, \cdots, N$, where $S(k)$ is given by (3.2.2).

- The distribution of magnitude of DFT coefficients per critical band $F_z(k)$, for $LB_z \leq k \leq HB_z$ is found by translating frequency into bark scale. Here $z = 1, 2, \cdots, z_t$ are the critical bands, $z_t$ is total number of critical bands and
$LB_z$ and $HB_z$ are the respective lower and higher frequencies in the critical band $z$.

- The watermark is embedded in critical bands in which magnitude of DFT coefficients is less than the defined masking threshold, $Y(z)$.

- The final watermark is now generated as

\[
W(k) = W_0(i), \quad \text{if} \quad mN \leq k \leq nN
\]
\[
= W'_0(i), \quad \text{if} \quad (1-n)N \leq k \leq (1-m)N
\]
\[
= 0, \quad \text{otherwise}
\]  

(3.3.15)

Here $0 \leq i \leq N_w$ and $0 \leq k \leq N - 1$ and $(0 < m < n < 0.5)$ to maintain symmetry of final watermark.

- Once location of embedding is decided, the watermark scaling factor $a$ has to be calculated for each critical band to ensure inaudibility of the embedded watermark. The scale factor $a_z$ of $z^{th}$ critical band is obtained by dividing masking threshold $Y(z)$ by the maximum magnitude component of the DFT coefficient in each critical band as

\[
a_z = A \frac{Y(z)}{\max(|F(k)|)}, \quad \text{for} \quad z = 1, 2, \ldots, z_l
\]  

(3.3.16)

Here $A$ is the gain factor that controls the overall magnitude of the watermarked signal $F'(k)$ given in (3.3.5). The value of $A$ varies from 0 to 1. The scaling factor $a_z$ decides how much the amplitude of watermark is to be suppressed in the selected critical band before adding it to the spectrum of host signal.

- The scaled watermark is now added according to rule

\[
F'(k) = F(k) \quad \text{if} \quad F(k) \geq Y(z)
\]
\[
= F(k)(1 + a_z W(k)) \quad \text{if} \quad F(k) < Y(z)
\]  

(3.3.17)

here $0 \leq k \leq N - 1$. Watermark is embedded in all frames in which the DFT spectrum is found lower than the maximum amplitude spread of original signal. Consequently, the spectral coefficients are modified only in certain critical bands of a frame so that transparency of audio signal is maintained. Hence embedding
of watermark is not in consecutive blocks of signal, but only in those blocks which satisfy the above said criterion.

- The length of pn sequence is taken such that one sequence is completely inserted in a given frame. Hence marking is done framewise where the embedding capacity of watermark will vary frame to frame, according to the embedding criterion given by equation (3.3.17).

- The modified amplitude of DFT coefficient $F'(k)$ is now combined with their corresponding phases $\phi(k)$, to get watermarked DFT coefficients $S'(k)$.

- The corresponding time domain watermarked signal $s'(n)$ is obtained by calculating inverse discrete fourier transform (IDFT) of $S'(k)$ given by (3.2.3).

3.3.4 Optimal Watermark Detection

The aim of watermark detection is to verify, whether or not the given watermark $W_d$ at receiver end resides in the test signal $F_t$. The detection is blind i.e. secret key is the only information that detector has at the receiver end. Also, the detection of watermark sequence is performed frame-wise.

The detector uses salient points for synchronizing the embedded information, so that audio can be analyzed for salient point extraction. Watermark detection can be considered as a binary hypothesis test, given as

- $H_0$: The test signal $F_t$ does not contain the watermark $W_d$

  \[ F_t = F, \quad a = 0 \quad (3.3.18) \]

- $H_1$: The test signal under investigation $F_t$ contains watermark $W_d$.

  \[ F_t = F(1 + aW_d), \quad a \neq 0, \quad (3.3.19) \]

The hypothesis test is solved by means of a correlation detector [104] and log-likelihood ratio detector [105]. However, few assumptions are done before performing the detector tests.

Assumptions about the Host signal and Watermark: In our proposed technique the watermark is embedded multiplicatively in the magnitude of the DFT
domain. These DFT coefficients of watermarked signal are deterministic however, due to watermark insertion at random locations the DFT magnitude of marked audio becomes random in nature. Hence for statistical detection techniques (such as Likelihood ratio test) a basic step for the construction of optimal watermark detector is the computation of the marked DFT magnitude coefficient distribution. Hence for optimal detection, following assumptions are made.

- The sample of the watermarked signal can be modelled as independent identically distributed (i.i.d) random variables obeying certain distribution. This approach has been taken in many research papers as well [106], [107], [108]. Hence, the samples of DFT magnitude can be modelled as i.i.d random variables which follow the Weibull distribution.

- DFT magnitude is wide sense stationary process.

- For large number of samples likelihood ratio and correlation coefficient attain Gaussian distribution due to central limit theorem.

Likelihood Ratio Detector

The watermarked signal $F'$ given in (3.3.5), may undergo various signal processing or noisy channel attacks before reaching the receiver end. The received signal $F_t$ is now used for watermark detection, by using log-likelihood ratio test. In Bayes theory of hypothesis testing, the criterion of optimum watermark detection is minimization of Bayes risk which leads to a decision criterion based on the likelihood ratio $l(F_t)$, given as

$$l(F_t) = \frac{p_{F_t}(f_t|H_1)}{p_{F_t}(f_t|H_0)}$$

(3.3.20)

where $p_{F_t}(f_t|H_0)$ and $p_{F_t}(f_t|H_1)$ are the conditional probability density functions of $F_t$ under the hypotheses $H_0$ and $H_1$. Now considering the second assumption, the likelihood ratio rewritten as

$$l(F_t) = \prod_{i=1}^{N} \frac{p_{F_t}(f_i|H_1)}{p_{F_t}(f_i|H_0)}$$

(3.3.21)

The detector compares likelihood detection ratio $l(F_t)$ with a detection threshold $\lambda$. The value of $\lambda$ should be such that it minimizes the overall error probability. Hence,
decision function $\Phi$ is given as

$$\Phi(F_t) = 1, \quad l(F_t) \geq \lambda$$
$$= 0, \quad otherwise \quad (3.3.22)$$

The value of $\lambda$ should be such that it minimizes the overall error probability. It is often convenient to replace the likelihood ratio with log-likelihood ratio, defined as

$$L(F_t) = \ln[l(F_t)] \quad (3.3.23)$$

leading to equation

$$\Lambda = \ln \lambda \quad (3.3.24)$$

where $\Lambda$ is the threshold of log likelihood detector.

**Performance evaluation of Detector**

We now evaluate the detection threshold and Receiver Operating Characteristics (ROC) for watermarked DFT magnitude.

**STEP-I: Selection of Distribution Function**

The modeling PDF for DFT magnitudes is evaluated for short frame size ($< 100\text{ms}$) [109], [110]. We have assumed that each DFT frame $F = [f(1), \cdots, f(N)]$ follows Weibull distribution [111], however the parameters of distribution (mean and variance of PDF) may defer frame to frame. Hence we have estimated the parameters of distribution for every frame separately in which watermark is embedded. In this case, we need to do the estimation once for each frame.

The probability density function (pdf) of Weibull distribution is defined as

$$p_F(f) = \frac{\beta}{\alpha} \left(\frac{f(i)}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{f(i)}{\alpha}\right)^{\beta}\right], \quad (3.3.25)$$

here $f(i) > 0$ for $i = 1, 2, \cdots, N$. Scale parameter ($\alpha$) and shape parameter ($\beta$) are positive real valued parameters, which control the mean, variance and shape of distribution. The mean $\mu_f$ and variance $\sigma_f^2$ can be written in terms of the scale and shape parameters as

$$\mu_f = \alpha \Gamma(1 + \frac{1}{\beta}),$$

$$\sigma_f^2 = \alpha^2 \Gamma(1 + \frac{2}{\beta}) - \mu_f^2 \quad (3.3.26)$$
where gamma function $\Gamma(x)$ is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt$$

(3.3.27)

The two specific cases $\beta = 1$ and $\beta = 2$ correspond to exponential and Rayleigh distribution.

**STEP-II: Estimation of Parameter**

Once the underlying distribution has been chosen, the next step is the estimation of parameters that govern the characteristics of the selected probability function. The parameter estimation problem consists of finding the underlying distribution parameters by observing samples of random variable described in [112]. Given $N$ sample values $[f(1), \cdots, f(N)]$, from the random variable $F$, which can be modeled by a two parameter Weibull distribution with a pdf as given by (3.3.25) the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ of $\alpha$ and $\beta$ respectively [113] are known to satisfy the equations

$$\hat{\alpha} = \left(\frac{1}{N} \sum_{i=1}^{N} (f(i))^{\hat{\beta}}\right)^{1/\hat{\beta}}$$

(3.3.28)

and

$$\hat{\beta} = \left[\left(\frac{1}{N} \sum_{i=1}^{N} f(i)^{\hat{\beta}} \log f(i)\right) \left(\frac{1}{N} \sum_{i=1}^{N} (f(i))^{\hat{\beta}}\right)^{-1} - \frac{1}{N} \sum_{i=1}^{N} \log f(i)\right]^{-1}$$

(3.3.29)

The value of $\hat{\beta}$ has to be obtained from (3.3.29) by the use of standard iterative procedures (i.e. Newton-Raphson method) and then used in (3.3.28) to obtain $\hat{\alpha}$.

Although the optimum decoder structure requires knowledge of the distribution underlying the magnitude of the non-watermarked coefficients $f_i$, this information is not present at the decoder side. Since decoding is done without resorting to the original audio, the decoder has no access to the original coefficients. Hence the distribution of the non-watermarked coefficients $f_i$ needs to be approximated by the distribution of the watermarked coefficients $f'_i$. Figure 3.4 shows the plot of PDF (on y axis) vs DFT magnitude (on x axis) of watermarked (right curve) and original audio signal (left curve). As long as the embedding strength and thus the watermark power is kept small, the difference between the two distributions will be negligible, as shown in figure 3.4. Also, the density function tends to $\infty$ as DFT magnitude approaches zero from above and is strictly decreasing. The plot obtained ($\gamma = 0.6833$) exactly matches with the results suggested in [111]. The values of $\alpha$ and $\beta$ obtained from
maximum likelihood estimator are 2.9369 and 0.6833 respectively.

**STEP-III: Evaluation of Threshold**

Having identified a suitable model for host feature, we now find the likelihood ratio, as given in [114].

**Under hypothesis** \(H_0\), \(f_t(i)=f(i)\), and each \(f_t(i)\) follows Weibull pdf with parameters equal to those of the original coefficients. Hence putting \(f_t(i)\) in place of \(f(i)\) in (3.3.25), we get

\[
p_{F_t}(f_t|H_0) = \frac{\beta}{\alpha} \left( \frac{f_t(i)}{\alpha} \right)^{\beta-1} \exp \left[ - \left( \frac{f_t(i)}{\alpha} \right)^{\beta} \right], \quad \text{for } f_t > 0 \quad (3.3.30)
\]

**Under hypothesis** \(H_1\), \(f_t(i) = f(i)(1 + aW(i))\), hence putting \(f(i) = \frac{f_t(i)}{(1+aW(i))}\) in (3.3.25) the pdf under hypothesis \(H_1\) is obtained as

\[
p_{F_t}(f_t|H_1) = \frac{\beta}{\alpha(1+aW(i))} \left( \frac{f_t(i)}{\alpha(1+aW(i))} \right)^{\beta-1} \exp \left[ - \left( \frac{f_t(i)}{\alpha(1+aW(i))} \right)^{\beta} \right]. \quad (3.3.31)
\]

where \(f_t(i) > 0\) for \(i = 1, 2, \ldots, N\).

Substituting the values of \(p_{F_t}(f_t|H_0)\) and \(p_{F_t}(f_t|H_1)\) in (3.3.21), we get the likelihood
as
\[ l(F_t) = \prod_{i=1}^{N} \left( \frac{1}{1 + aW(i)} \right)^\beta \exp \left[ - \left( \frac{f_t(i)}{\alpha(1 + aW(i))} \right)^\beta + \left( \frac{f_t(i)}{\alpha} \right)^\beta \right] \] (3.3.32)

Taking the log of likelihood ratio, we get
\[ L(F_t) = \sum_{i=1}^{N} \left[ - \beta \ln(1 + aW(i)) \right] + \sum_{i=1}^{N} \left[ - \left( \frac{f_t(i)}{\alpha(1 + aW(i))} \right)^\beta + \left( \frac{f_t(i)}{\alpha} \right)^\beta \right] \] (3.3.33)

Solving the equation we finally get
\[ L(F_t) = \sum_{i=1}^{N} \left[ - \beta \ln(1 + aW(i)) \right] + \sum_{i=1}^{N} (f_t(i))^\beta v(i) \]
\[ = \sum_{i=1}^{N} \left[ - \beta \ln(1 + aW(i)) \right] + \sum_{i=1}^{N} z(i) \] (3.3.34)

By applying some algebra and by retaining only the terms depending on \( f_t \), we get
\[ z(i) = v(i) f_t(i)^\beta \] (3.3.35)

where
\[ v(i) = \left[ \frac{(1 + aW(i))^\beta - 1}{\alpha^\beta(1 + aW(i))^\beta} \right] \] (3.3.36)

The values of \( \alpha \) and \( \beta \) in (3.3.26) are estimated using maximum likelihood method and the value of threshold \( \Lambda \) is found for a given probability of false alarm. We now get the sufficient statistic \( z \) for watermark detection given as
\[ z = \sum_{i=1}^{N} z(i) \]
\[ = \sum_{i=1}^{N} v(i) f_t(i)^\beta \] (3.3.37)

• **Under Hypothesis** \( H_0 \), the signal under test \( f_t(i) = f(i) \). Hence \( f_t(i) \) will also have Weibull distribution with parameter \( \alpha \) and \( \beta \), \( f_t^\beta(i) \) is exponential pdf with mean \( \alpha^\beta \) and variance \( \alpha^{2\beta} \). By also taking into account the presence of \( v_i \), we conclude that second term of the sum in (3.3.34) follows an exponential pdf, given as
\[ p(z(i)|H_0) = p \left( v(i)(f_t(i))^\beta | H_0 \right) \]
\[ = \eta_i \exp \left( -\eta_i z(i) \right) u[z(i)] , \] (3.3.38)
Here,
\[ \eta_i = \frac{(1 + aW_i)^\beta}{(1 + aW_i)^\beta - 1} \]  \hspace{1cm} (3.3.39)

The value of \( L(F_t) \) is now given as
\[ L(F_t) = \sum_{i=1}^{N} \left[ -\beta \ln(1 + aW(i)) \right] + z \]  \hspace{1cm} (3.3.40)

By invoking central limit theorem, \( z \) is assumed to follow Gaussian distribution with mean and variance
\[ \mu_{z|H_0} = \sum_{i=1}^{N} \frac{1}{\eta_i} = \sum_{i=1}^{N} \frac{(1 + aW(i))^\beta - 1}{(1 + aW(i))^\beta} \]  \hspace{1cm} (3.3.41)
\[ \sigma_{z|H_0}^2 = \sum_{i=1}^{N} \left( \frac{1}{\eta_i^2} \right) = \sum_{i=1}^{N} \left( \frac{(1 + aW(i))^\beta - 1}{(1 + aW(i))^\beta} \right)^2 \]  \hspace{1cm} (3.3.42)

**Under Hypothesis** \( H_1 \), the signal under test becomes \( f_t(i) = f(i)(1 + aW(i)) \).
Then \( f_t(i) \) will have Weibull distribution with parameter \( \alpha' = \alpha(1 + aW(i)) \) and \( \beta' = \beta \). The pdf of \( z(i) \) under hypothesis \( H_1 \) will be given as,
\[ p(z(i)|H_1) = p\left(v(i)(f_t(i))^\beta|H_1\right) \]
\[ = \eta'_i \exp\left(-\eta'_i z(i)\right) u[z(i)], \text{ if } W(i) > 0 \]
\[ = |\eta'_i| \exp\left(-\eta'_i z(i)\right) u[-z(i)], \text{ if } w_i < 0 \]  \hspace{1cm} (3.3.43)

Here,
\[ \eta'_i = \frac{1}{(1 + aW(i))^\beta - 1} \]  \hspace{1cm} (3.3.44)

The value of mean and variance of \( z \) will now be given as
\[ \mu_{z|H_1} = \sum_{i=1}^{N} \frac{1}{\eta'_i} = \sum_{i=1}^{N} \left( 1 + aW(i))^\beta - 1 \right] \]  \hspace{1cm} (3.3.45)
\[ \sigma_{z|H_1}^2 = \sum_{i=1}^{N} \left( \frac{1}{(\eta'_i)^2} \right) = \sum_{i=1}^{N} \left( (1 + aW(i))^\beta - 1 \right)^2 \]  \hspace{1cm} (3.3.46)

**STEP-IV: Performance Analysis of Likelihood ratio detector**
The decision on the valid hypothesis is taken by comparing log-likelihood ratio \( L(F_t) \) against a properly selected threshold \( \Lambda \). A decision to adopt hypothesis \( H_0 \) is taken when \( L(F_t) < \Lambda \) and \( H_1 \) when \( L(F_t) \geq \Lambda \). The detection threshold \( \Lambda \) controls the tradeoff between the probabilities of false positive and false negative decisions.
The performance of a log-likelihood based technique can be measured in terms of probability of false alarm $P_f$ and probability of misdetection $P_m$ defined as

$$ P_f = P(L(F_t) > \Lambda | H_0) = \int_{\Lambda}^{\infty} p_{L(F_t)|H_0}(x)dx \quad (3.3.47) $$

$$ P_m = P(L(F_t) \leq \Lambda | H_1) = \int_{-\infty}^{\Lambda} p_{L(F_t)|H_1}(x)dx \quad (3.3.48) $$

Considering fourth assumption (likelihood ratio follows Gaussian distribution), the probability of false alarm becomes

$$ P_f = \int_{\Lambda}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{L(F_t)|H_0}}} exp\left(-\frac{(x - \mu_{L(F_t)|H_0})^2}{2\sigma_{L(F_t)|H_0}^2}\right)dx \quad (3.3.49) $$

and probability of false rejection is given as

$$ P_m = \int_{-\infty}^{\Lambda} \frac{1}{\sqrt{2\pi\sigma_{L(F_t)|H_1}}} exp\left(-\frac{(x - \mu_{L(F_t)|H_1})^2}{2\sigma_{L(F_t)|H_1}^2}\right)dx \quad (3.3.50) $$

Evaluating above equation we obtain

$$ P_f = \frac{1}{2} - \frac{1}{2}erf\left(\frac{\Lambda - \mu_{L(F_t)|H_0}}{\sqrt{2\sigma_{L(F_t)|H_0}}}\right) = \frac{1}{2}erfc\left(\frac{\Lambda - \mu_{L(F_t)|H_0}}{\sqrt{2\sigma_{L(F_t)|H_0}}}\right) \quad (3.3.51) $$

$$ P_m = \frac{1}{2} + \frac{1}{2}erf\left(\frac{\Lambda - \mu_{L(F_t)|H_1}}{\sqrt{2\sigma_{L(F_t)|H_1}}}\right) = 1 - \frac{1}{2}erfc\left(\frac{\Lambda - \mu_{L(F_t)|H_1}}{\sqrt{2\sigma_{L(F_t)|H_1}}}\right) \quad (3.3.52) $$

Threshold is calculated so that the desired $P_f$ is achieved. The value of threshold $\Lambda$ in terms of $P_f$ and $P_m$ respectively is given as,

$$ \Lambda = \sqrt{2\sigma_{L(F_t)|H_0}}erf^{-1}(1 - 2P_f) + \mu_{L(F_t)|H_0} \quad (3.3.53) $$

$$ \Lambda = \sqrt{2\sigma_{L(F_t)|H_1}}erf^{-1}(2P_m - 1) + \mu_{L(F_t)|H_1} \quad (3.3.54) $$

In case of multiplicative Weibull channel $L(F_t)$ is replaced by $z$ in (3.3.51). $P_f$ is then given as

$$ P_f = \frac{1}{2}erfc\left(\frac{\Lambda - \mu_z|H_0}{\sqrt{2\sigma_z|H_0}}\right) \quad (3.3.55) $$

giving $\Lambda$ equal to

$$ \Lambda = \sqrt{2\sigma_z|H_0}erf^{-1}(1 - 2P_f) + \mu_z|H_0 \quad (3.3.56) $$

For $P_f=10^{-6}$

$$ \Lambda = 3.3\sqrt{2\sigma_z|H_0} + \mu_z|H_0 \quad (3.3.57) $$
Similarly from (3.3.52), \( P_m \) is given as

\[
P_m = \frac{1}{2} \text{erfc} \left( \frac{\mu_{z|H_1} - \Lambda}{\sqrt{2} \sigma_{z|H_1}} \right)
\]

(3.3.58)

The Bayes decision criterion then becomes

\[
\Phi(F) = 1, \quad \text{if} \quad z > \Lambda
\]

\[
= 0, \quad \text{otherwise}
\]

(3.3.59)

According to above equation \( P_f \) and \( P_m \) depend on threshold \( \Lambda \). A possible change of \( \Lambda \) increases one probability and decrease the other. Hence an appropriate threshold should be selected. By solving both (3.3.55) and (3.3.58), one can express \( P_f \) as a function of \( P_m \). The plot of \( P_f \) versus \( P_m \) is called the Receiver Operating Characteristic (ROC) curve of the corresponding watermarking system. This curve conveys all the information required in order to judge the detection performance of a such a system. Substituting the value of \( \Lambda \) in \( P_f \) we obtain the relation

\[
P_f = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\sqrt{2} \sigma_{z|H_1} \text{erf}^{-1}(2P_m - 1) + \mu_{z|H_1} - \mu_{z|H_0}}{\sqrt{2} \sigma_{z|H_0}} \right) \right]
\]

(3.3.60)

### 3.3.5 Correlation detector

The correlation detector, which is the Maximum Likelihood (ML) optimal detector, is applied to additive or multiplicative watermarking system. These detectors give optimal results while considering Gaussian distribution for the host signals. The correlation detection can be performed by computing the correlation \( c \) between pseudorandom sequence \( W \) and watermarked signal \( F_t \) in time or frequency domain given as

\[
c = F_t W = [F(1 + \alpha W)] W = FW + \alpha FWW
\]

(3.3.61)

The watermark chips generated, are such that they are mutually independent with respect to the original audio signal, hence the first product term (\( FW \)) vanishes. The second product term \( WW \) gives high correlation, if the watermark under test is correct. The correlation is therefore compared to a predefined threshold to determine whether watermark is present in the signal or not. This is due to a special property of pseudo-random sequences. The autocorrelation calculated between the pseudo-random sequences generated by the same initiator has the maximal value. On the
contrary, the cross-correlation calculated between pseudorandom sequences generated by different initiators has a very small value. Most popular pseudo-random sequence is the maximum length sequence (also known as M-sequence) [66]. Even though there is no malicious attack, the correlation-based detection can be erroneous due to this characteristic of the host signal. These errors are called false-positive and false-negative errors. The probability of these errors should be as low as $10^{-8}$.

The received signal $F_t$ is used for watermark detection, by using correlation test. In case of correlation detector we substitute $L(F_t)$ by correlation coefficient $c$ and threshold $\Lambda$ is replaced by $T_c$ in (3.3.22). Under Bayes decision criterion the detector compares correlation coefficient $c$ with a detection threshold $T_c$. The value of $T_c$ should be such that it minimizes the overall error probability. Hence Bayes decision criterion becomes

$$\Phi(F_t) = \begin{cases} 1, & if \quad c > T_c \\ 0, & if \quad otherwise \end{cases} \quad (3.3.62)$$

Performance analysis of correlation detector

The correlation detector examines whether the signal under test $F_t(k)$, $0 \leq k \leq N-1$, contains the watermark $W(k)$ or not under a binary hypothesis test framework. The correlation between the signal under investigation $F_t$ and the watermark sequence available at detector end $W_d$ is given by

$$c = \frac{1}{N} \sum_{n=0}^{N-1} F_t(n)W_d(n)$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} [F(n)W_d(n) + aF(n)W(n)W_d(n)] \quad (3.3.63)$$

here $a$ is the scaling factor ranging between 0 and 1. In order to decide on the valid hypothesis, $c$ is compared against a suitably selected threshold $T_c$. To proceed with the extraction of the correlation detector statistics, we first need to make certain assumptions about the host signal and the watermark. These assumptions are as follows

- The host signal, that is, the magnitude of the DFT coefficients, is assumed to be wide-sense stationary, thus

$$E[F(n)] = \mu_f \quad for \quad n = 0, 1, \ldots, N - 1$$
\[ E[F(n)F(n + k)] = R_f(k) \quad \text{for} \quad n = 0, 1, \ldots, N - 1 \quad (3.3.64) \]

Furthermore, the host signal is assumed to follow an exponential autocorrelation function [115] given as,

\[ R_f(k) = \mu_f^2 + \sigma_f^2 \beta_f^k, \quad k \geq 0, \quad |\beta_f| \leq 1 \quad (3.3.65) \]

where \( \beta_f \) is the parameter of autocorrelation function and \( \sigma_f^2 \) is the host signal variance, given as

\[ \sigma_f^2 = E[F^2(n)] - E[F(n)]^2 \quad (3.3.66) \]

- The watermarks generated by pseudo-random number generator can be accurately modeled as i.i.d. random variables obeying a uniform distribution. These sequences are wide sense stationary and ergodic. For such watermarks the terms of sum in correlator equation can be assumed independent. From the above assumptions the pseudo-random watermark signal moments are given as

\[
E[W^m(i)] = \begin{cases} 
0, & \text{if } m \text{ is odd} \\
\frac{1}{(m+1)2^m}, & \text{if } m \text{ is even} 
\end{cases} \quad (3.3.67)
\]

\[
E[W^l(i)W^m(j)] = E[W^l(i)]E[W^m(j)] \quad \text{for} \quad i \neq j \quad (3.3.68)
\]

The mean and variance of correlation coefficient is calculated as follows

\[
\mu_c = E[c] = \frac{1}{N} \sum_{n=0}^{N-1} E[F(n)](1 + aE[W(n)])E[W_d(n)] \\
= \frac{1}{N} \sum_{n=0}^{N-1} [E[F(n)]E[W_d(n)] + \frac{1}{N} \sum_{n=0}^{N-1} aE[F(n)]E[W(n)W_d(n)]] \\
= \frac{1}{N} \sum_{n=0}^{N-1} aE[F(n)]E[W(n)W_d(n)] \quad (3.3.69)
\]

\[
\sigma_c^2 = E[c^2] - E[c]^2 \\
= E[(\frac{1}{N} \sum_{n=0}^{N-1} [F(n)W_d(n) + aF(n)W(n)W_d(n)])^2] - \mu_c^2 \\
= \frac{1}{N^2} \left[ \sum_{n=0}^{N-1} (E[F^2(n)]E[W_d^2(n)] + a^2 E[F^2(n)]E[W^2(n)W_d^2(n)] \\
+ 2aE[F^2(n)]E[W(n)W_d^2(n)]) + \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} aE[F(n)F(m)]E[W_d(n)W_d(m)] \right]
\]
\[ + aE[F(n)F(m)]E[W(n)W_d(n)W_d(m)] \\
+ aE[F(n)F(m)]E[W(m)W_d(n)W_d(m)] \\
+ a^2E[F(n)F(m)]E[W(n)W_d(n)W(m)W_d(m)] \] \right) - \mu_c^2 \quad (3.3.70)\

The parameters \( p_{c|H_0}, p_{c|H_1} \) are the conditional probability density functions of the correlation \( c \) under the hypotheses \( H_0 \) and \( H_1 \), respectively with mean and variance \((\mu_{c|H_0}, \sigma_{c|H_0})\) and \((\mu_{c|H_1}, \sigma_{c|H_1})\) respectively. Hence the performance of an multiplicative embedding correlation-based watermarking scheme for this type of watermark signals depends only on these four parameters, given as

- Under Hypothesis \( H_0 \), the values of mean and variance are
  \[
  \mu_{c|H_0} = 0, \\
  \sigma_{c|H_0}^2 = \frac{1}{12N^2} \sum_{n=0}^{N-1} E[F^2(n)].
  \]

- Under Hypothesis \( H_1 \), value of mean and variance are
  \[
  \mu_{c|H_1} = \frac{a}{12} \\
  \sigma_{c|H_1}^2 = \frac{1}{N^2} \left[ \sum_{n=0}^{N-1} \left( E[F^2(n)]\left(\frac{1}{12} + \frac{a^2}{80}\right) \right) \\
  + \sum_{n=0}^{N-1} \sum_{m, m \neq n}^{N-1} \frac{a^2}{144} E[F(n)F(m)] \right] - \left(\frac{a}{12}\right)^2 \]

For a given threshold, the system performance can be measured in terms of the probability of false alarm \( P_f(T_c) \), (i.e., the probability to detect a watermark in a signal that is not watermarked and the probability of misdetection \( P_m(T_c) \) (i.e., the probability to erroneously neglect the watermark existence in the signal).

\[
P_f(T_c) = P\{c \geq T_c | H_0\} \\
P_m(T_c) = P\{c < T_c | H_1\}
\]

\( P_f(T_c) \) and \( P_m(T_c) \) can be calculated as follows:

\[
P_f(T_c) = \int_{T_c}^{\infty} p_{c|H_0}(t)dt \\
P_m(T_c) = \int_{-\infty}^{T_c} p_{c|H_1}(t)dt
\]
For watermarks generated by pseudo-random sequence generators, the correlation function \( c \) follows Normal distributions. The central limit theorem for random variables with small dependency may be used, assuming \( N \) is large. Under this assumption \( P_f(T_c) \) is given as

\[
P_f(T_c) = \frac{1}{2} \text{erf} \left( \frac{T_c - \mu_{c|H_0}}{\sqrt{2} \sigma_{c|H_0}} \right)
\]

(3.3.77)

Similarly \( P_m(T_c) \) will be

\[
P_m(T_c) = \frac{1}{2} \text{erf} \left( \frac{\mu_{c|H_1} - T_c}{\sqrt{2} \sigma_{c|H_1}} \right)
\]

(3.3.78)

solving (3.3.77) for \( T_c \) we get

\[
T_c = \sqrt{2} \sigma_{c|H_0} \text{erf}^{-1}(1 - 2P_f) + \mu_{c|H_0}
\]

(3.3.79)

Solving (3.3.77) and (3.3.78) we obtain the ROC as

\[
P_f = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\sqrt{2} \sigma_{c|H_1} \text{erf}^{-1}(2P_m - 1) + \mu_{c|H_1} - \mu_{c|H_0}}{\sqrt{2} \sigma_{c|H_0}} \right) \right]
\]

(3.3.80)

Putting the values of \( \mu_{c|H_0} \) and \( \mu_{c|H_1} \) from (3.3.71) and (3.3.73) respectively we get

\[
P_f = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\sqrt{2} \sigma_{c|H_1} \text{erf}^{-1}(2P_m - 1) + a/12}{\sqrt{2} \sigma_{c|H_0}} \right) \right]
\]

(3.3.81)

Further the values of \( \sigma_{c|H_0} \) and \( \sigma_{c|H_1} \) can be obtained from (3.3.72) and (3.3.74) respectively and thereby ROC for a given correlation coefficient can be evaluated.

### 3.4 Experimental Results

To generate experimental results a total of 10 standard audio test sequences are taken which are listed in table 3.1. These test sequences are adopted to analyze the performance of the proposed watermarking algorithm. These items represent a corpus of audio material which contains excerpts from single instruments, multiple instruments, complex sound sources and speech. This delineation was chosen because all items have different spectral properties. Each signal was sampled at 44.1 kHz, represented by 16 bits per sample, and eight seconds in length. The DFT magnitude of audio signal was assumed to follow Weibull distribution with pdf given by (3.3.25) and the value of the parameters was evaluated using maximum likelihood method as shape parameter, \( \beta = 0.6833 \) and scale parameter, \( \alpha = 2.9369 \).
Table 3.1: Audio test sequences (44.1 kHz, 16 bit)

<table>
<thead>
<tr>
<th>TS.No.</th>
<th>audio</th>
<th>TS.No.</th>
<th>audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>drums</td>
<td>6</td>
<td>clarinet</td>
</tr>
<tr>
<td>2</td>
<td>flute</td>
<td>7</td>
<td>waltz</td>
</tr>
<tr>
<td>3</td>
<td>speech(mono)</td>
<td>8</td>
<td>jazz</td>
</tr>
<tr>
<td>4</td>
<td>speech(stereo)</td>
<td>9</td>
<td>Synth</td>
</tr>
<tr>
<td>5</td>
<td>violin</td>
<td>10</td>
<td>haffner</td>
</tr>
</tbody>
</table>

3.4.1 Experimental Performance Evaluation

The value of scaling factor $a$ is changed and its effect is seen on performance of likelihood ratio detector and correlation detector respectively. For this the values of $a$ for various critical bands are obtained using (3.3.16).

- **Effect on detection threshold**
  In case of LLR detector, the effect of scale factor $a$ is observed on detection threshold $\Lambda$. First we have shown the curves between $\Lambda$ and $P_f$ keeping the value of $a$ fixed. The upper and lower portion of figure 3.5 shows the variations of $\Lambda$ with respect to $P_f$ for two values of $a$, 0.0024 and 0.8 respectively. The first value, $a = 0.0024$, is obtained from MAS threshold and second value, 0.8 is selected close to the maximum limit of $a$ to show the effects clearly visible. As can be seen from figure, for the same range of $P_f$ the variations in $\Lambda$ is only $0 < |\Lambda| \leq 0.08$ when $a = 0.0024$. Whereas the variations are quite high ($0 \leq |\Lambda| \leq 20$) for $a = 0.8$. Next we analyse the variations of $\Lambda$ with respect to $a$ for all the values in the range of $0 < a \leq 1$. Figure 3.6 shows the variation of $\Lambda$ with respect to $a$ for a fixed value of $P_f$ ($\simeq 10^{-6}$). From the plot we observe that the value of $\Lambda$ remains constant for $a \leq 0.04$. However, as the value of $a$ is increased beyond 0.04 a steep rise in $\Lambda$ is obtained. Another observation from figure is that with decreasing $a$, the value of detection threshold $\Lambda$ also decreases which in turn degrades the detector response.

In case of correlation detector the effect of $a$ is observed on detection threshold $T_c$. A plot between $T_c$ and $P_f$ for two different values of $a$ (i.e. 0.0024 and 0.8) is
Figure 3.5: Threshold versus probability of false detection for LLR detector for two values of scaling factor $a$. 

Detection Threshold $\Lambda$ Vs Probability of false alarm $P_f$ for LLR detector
Figure 3.6: Threshold versus scaling factor for Log-LLR detector for $P_f = 10^{-6}$

shown in upper and lower portion of figure 3.7. From the figure we observe that $T_c = 0.27$ when $P_f = 10^{-3}$ for both the values of $a$. Also the value ofthreshold lies within the range of $0 \leq T_c \leq 0.5$ for a wide variation of $a$ (i.e. $0 \leq a \leq 1$). Hence it can be inferred from the curves that the output of correlation detector is not much effected by scalingfactor $a$. Instead the output depends mainly on $pn$ sequence, taken as watermark during the embedding process. The detector gives high correlation, if the watermarkunder test is correct. It is due to the special property of $pn$ sequence. The autocorrelation calculated between the $pn$ sequences generated by the same initiator has the maximal value. On the contrary, the cross-correlation calculated between $pn$ sequences generated by different initiators has a very small value.

- **Effect on ROC**

The Receiver Operating Characteristic (ROC) curve is obtained from likelihood ratio and correlation watermark detectors, as shown in figure 3.8 and figure 3.9 respectively. The results are compared with actual experimental curve for both detectors with two different values of $a$, i.e. 0.0024 and 0.8. It is observed that for $a=0.0024$ the three curves nearly coincide with each other, whereas the
Figure 3.7: Threshold versus probability of false detection for correlation detector for two values of scaling factor $a$. 

Detection Threshold $T_c$ Vs Probability of false alarm $P_f$ for correlation detector
same is not true for the case $a=0.8$. For the proposed value of $a$, given in section 3.3.2, the statistical detectors give optimum results which are close to actual experimental value. Further we observe that $LLR$ detector gives better approximation to experimental results as compared to correlation detector, for all values of scaling factor.

### 3.4.2 Objective and Subjective Quality Evaluation

The proposed watermarking technique provides high embedding capacity with a very low perceptual distortion. Figure 3.10 shows the superimposed picture of original and watermarked audio for $a=0.0024$. It can be observed from figure that there are very low deviations between original and watermarked audio in time domain. Subjective and objective quality tests are performed to evaluate the quality of watermarked audio signal [96].

The **subjective audio quality** of watermarked audio is evaluated by double-blind A-B-C triple-stimulus hidden reference comparison test, as explained in section 2.4.2. Stimuli A contains the reference signal, whereas B and C are pseudo-randomly selected from the watermarked and the reference signal. After listening to all three,
Figure 3.9: Receiver operating characteristic curve for scaling factor $\alpha=0.8$

Figure 3.10: Watermarked signal imposed on original signal for a frame of $N = 512$ samples and scaling factor $\alpha=0.0024$. 
the subject was asked to identify either B or C as the hidden reference, and then grade the watermarked signal relative to the reference stimulus using the SDG. The standard species 20 subjects as an adequate size for the listening panel. Since expert listeners participated in the test, the number of listeners has been reduced to 10 for an informal test. A training session preceded the grading session where a trial was conducted for each signal. The tests were performed with headphones in a special cabin dedicated to listening tests. Ten test signals selected from the sound with a length of 10-20 s have been presented to the listeners.

The results of the listening test are shown in figure 3.4.2. For the different audio files, the mean SDG value and the 95% confidence interval are plotted as a function of the different audio tracks to clearly reveal the distance to transparency (SDG=0).
<table>
<thead>
<tr>
<th>S.No.</th>
<th>ODG</th>
<th>$a$</th>
<th>$\Lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.624</td>
<td>0.8</td>
<td>11.8741</td>
</tr>
<tr>
<td>2</td>
<td>-1.436</td>
<td>0.4</td>
<td>6.2374</td>
</tr>
<tr>
<td>3</td>
<td>-0.999</td>
<td>0.1</td>
<td>1.9353</td>
</tr>
<tr>
<td>4</td>
<td>-0.710</td>
<td>0.05</td>
<td>1.0094</td>
</tr>
<tr>
<td>5</td>
<td>-0.641</td>
<td>0.005</td>
<td>0.1050</td>
</tr>
<tr>
<td>6</td>
<td>-0.065</td>
<td>0.0024</td>
<td>0.0505</td>
</tr>
<tr>
<td>7</td>
<td>+0.045</td>
<td>0.001</td>
<td>0.0211</td>
</tr>
<tr>
<td>8</td>
<td>+0.145</td>
<td>0.0005</td>
<td>0.0105</td>
</tr>
</tbody>
</table>

Table 3.2: Results of ODG and scaling factor for multiplicative embedding in DFT magnitude of audio signals

It is observed from the above results that the quality degradation of the proposed watermarking scheme is very small for the vast majority of the test items, given in table 3.1. For all test items the SDG is within -0.7 to -0.065 which indicates that there is no significant distortion introduced by this scheme.

For objective quality measure, software ‘PQevalAudio’ for perceptual evaluation of audio quality (PEAQ) is utilized to evaluate an objective difference grade (ODG), which is an objective measurement of SDG. Table 3.2 lists the average value of PEAQ/ODG with the give test items for varying value of $a$. It shows that as value of scaling factor $a$ decreases, perceptual quality of watermarked audio becomes better. However, if the value of scaling factor is lowered below the value obtained from MAS (0.0024), ODG obtained is positive. The ITU recommendation does not allow positive ODGs, because this could also happen in listening tests, where the file under test is rated better than the reference file. The value of ODG obtained from watermarked audio is $-0.065$ for the optimum value of $a = 0.0024$. From ROCs plotted in figure 3.8, figure 3.9 and the objective quality given in table 3.2 we observe that for small values of $a$ the detector response is poor, but perceptual quality is within acceptable limits. On the contrary, for larger values of scale factor ($a \simeq 0.04$) the detector response improves, but then the perceptual transparency is deteriorated. It can be inferred from these results that proposed technique gives a good tradeoff between perceptual transparency and detector performance.
Table 3.3: Comparison of ODG and watermark embedding capacity between available literature schemes

<table>
<thead>
<tr>
<th>Technique</th>
<th>ODG</th>
<th>Embedding Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megias [80]</td>
<td>-0.5 to -2</td>
<td>61bps</td>
</tr>
<tr>
<td>Fujimoto [82]</td>
<td>–</td>
<td>1 Kbps</td>
</tr>
<tr>
<td>Fallahpour [83]</td>
<td>-0.5</td>
<td>3kbps</td>
</tr>
<tr>
<td>Proposed</td>
<td>-0.065 to -0.7</td>
<td>1.42 to 4kbps</td>
</tr>
</tbody>
</table>

3.4.3 Watermark Embedding Capacity

The proposed scheme provides high watermark embedding capacity with least perceptual distortions. Table 3.3 compares the embedding capacity and perceptual quality of proposed scheme with other schemes present in literature.

The scheme of Megias [80] achieves value of ODG between -0.5 to -2, which is not an acceptable range. The technique proposed by Fujimoto [82] provides high watermark embedding rate of 1 kbps and is simple to implement. However, the scheme only considers mp3 compression attack and ODG value is not mentioned. The algorithm proposed by Fallahpour [83] achieved a high capacity of about 3 kbps and it is robust against most attacks. The average ODG score achieved is -0.5, which is not too satisfactory. This could be due to the fact that the manipulation based on the estimated FFT coefficients introduces distortions.

The embedding capacity of proposed scheme was found to be 1.4 kbps, with \( ODG = -0.065 \), when embedding was done in only one critical band. The average watermark capacity increased to 4 kbps, with \( ODG = -0.7 \), when embedding was performed in more than one critical bands (i.e. 3). As compared to HAS, MAS enables relatively higher watermark embedding rate in DFT magnitude within acceptable limits of perceptual quality. The proposed method is thus able to provide large capacity whilst keeping imperceptibility in the admitted range (1 to 0).

3.4.4 Robustness to Attacks

The other major issue in watermarking is robustness to various attacks. We will now present the robustness of watermark against additive white gaussian noise (AWGN)
Addition of AWGN noise

Additive White Gaussian Noise is added to watermarked audio signal and Signal to Noise Ratio (SNR) is varied from 1dB to 10dB. The % watermark recovery indicates the number of watermark bits correctly detected out of total mark bits inserted by watermark embeder. As can be seen from the figure 3.12, 99 percent of watermark recovery is achieved for $SNR$ value of 6$dB$ and above. This implies high robustness of watermark against AWGN noise.

Presence of Multiple Watermark

To see the effect of presence of multiple watermark, 1000 normally distributed pseudo-random sequences with mean 0 and variance 1 are generated. The real valued normally distributed pn sequences with given mean and variance can be generated using $randn$ function in MATLAB. These sequences are used as test sequences $W_t$ for watermark detection process. The correct test watermark $W_d$ was used at 501$^{th}$ iteration. Further both types of detectors i.e. likelihood ratio and correlation detectors, are
Figure 3.13: Output of Log-LLR detector for scaling factor $a=0.8$

Figure 3.14: Correlation detector response for scaling factor $a=0.0024$
### Table 3.4: Robustness to common audio processing operations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time stretch (preserve pitch)</td>
<td>0.9129</td>
<td>0.8660</td>
<td>0.8660</td>
</tr>
<tr>
<td>Pitch Shift (preserve tempo)</td>
<td>0.8774</td>
<td>0.7773</td>
<td>0.7519</td>
</tr>
<tr>
<td>Resample (preserve neither)</td>
<td>0.7994</td>
<td>0.7790</td>
<td>0.7999</td>
</tr>
<tr>
<td>Low Pass Filter</td>
<td>0.8272</td>
<td>0.8022</td>
<td>0.8024</td>
</tr>
<tr>
<td>High Pass Filter</td>
<td>1</td>
<td>0.9553</td>
<td>0.9522</td>
</tr>
<tr>
<td>MP3 (128kbps)</td>
<td>0.8972</td>
<td>0.955</td>
<td>0.9079</td>
</tr>
<tr>
<td>Resample(16k/16bps)</td>
<td>0.7994</td>
<td>0.8320</td>
<td>0.8054</td>
</tr>
<tr>
<td>Cropping (with Half left)</td>
<td>0.9102</td>
<td>0.92</td>
<td>0.9501</td>
</tr>
<tr>
<td>Delay (10.2 msec)</td>
<td>0.7962</td>
<td>0.8938</td>
<td>0.7961</td>
</tr>
<tr>
<td>Invert</td>
<td>0.9351</td>
<td>0.9523</td>
<td>0.9975</td>
</tr>
</tbody>
</table>

In case of LLR detector the value of $\Lambda$ obtained is 9.8 for $P_f = 10^{-6}$ with scaling factor $a = 0.8$. The LLR detector output is shown for high value of $a$, as the response of this detector is poor for small values of $a$, as can be seen from figure 3.6. Figure 3.13 shows the response of LLR detector in the presence of multiple watermarks. It can be observed from figure that log-likelihood ratio $\Lambda$ of correct watermark is above 9.8 whereas the LLR ratio of other watermarks is well below the threshold.

Similarly in case of correlation detector the value of threshold $T_c$ obtained statistically was 0.271. As can be seen from figure 3.14 the correct watermark at 501th can be very easily distinguished from other watermarks. The correlation coefficient $c$ of correct watermark is above 0.271 whereas other watermarks are quite below the threshold. Hence the threshold values evaluated statistically matches with the experimental results.

#### 3.4.5 Robustness to common audio manipulations

Further, we test the robustness of our work against several kinds of common audio manipulations (or attacks). The audio editing tools adopted in our experiment are Cool Edit Pro v2.1 and Goldwave v5.10 to generate all the following attacks. The correlation of template matching is given in table 3.4 to show the applicability of the proposed scheme in searching for watermark protected audio clips.
• **MP3 compression**: To test the robustness against lossy compression, the watermarked audio is compressed and decompressed by MPEG-I Layer 3 (MP3) at 128 kbps. Results indicate high values of the correlation.

• **Re-sampling**: The watermarked audio with original 44100 Hz sampling rate and 16 bits/sample is re-sampled down to 16000 Hz and 16 bits per sample. Then the low-resolution audio is up-sampled to 44100 Hz and re-quantized to 16 bits/sample. Although the above procedure caused audible noise, there is almost no effect on the correlation of template matching and the extracted owner’s information is hardly affected.

• **Low-pass filtering**: To test the robustness against filtering, a low-pass filter was applied to the watermarked audio sampled at 44100 Hz. A lowpass filter with less than 3 dB of ripple in the passband defined from 0 to 4kHz and at least 40 dB of ripple in the stopband defined from 6 kHz to the Nyquist frequency (22050 Hz) was designed. The loss of high frequency components is clearly audible; however, the symmetrically embedded watermark can be detected successfully from the low frequency components.

• **Random cropping**: The watermarked audio is randomly cropped and left a half segment in length. Due to the fact that each slice is an independent processing unit, we can extract watermarks from the remaining frames after block synchronization. We can successfully recognize the hidden information and the correlations of template matching are reasonably high.

• **High Pass Filter**: A 6th-order highpass Butterworth filter with cutoff frequency of 7000 Hz was applied on watermarked data sampled at 44100 Hz. The symmetrically embedded watermark can be detected successfully from the high frequency components.

• **Time scaling**: The watermarked audio is scaled by 1.2% for testing, including the following three different kinds: time stretching (preserves pitch), pitch stretching (preserves tempo) and resampling (preserves neither). The time scaling of resampling attacks, time stretching and pitch stretching attacks has very
low effect on our extracting scheme. The shifting and scaling of each slice can be detected by template matching.

### 3.4.6 Robustness to Stirmark Audio Benchmark

Stirmark for Audio [116] is a standard robustness evaluation benchmark tool for audio watermarking techniques. The test results for all test functions in Stirmark Benchmark for Audio V0.2 are listed in Table 5 and are performed with the default parameters included in the version of the tool available online [116]. For that experiment, we have selected 10 standard audio clips (Table 3.1), watermarked it, and then detected watermarks in the original, the marked copy, and all 49 clips created by the Stirmark Audio suite of attacks. The detection results are presented in Table 5. The detection threshold is set to $T_c = 0.27$, which results in an estimated probability of a false positive smaller than $10^{-3}$ for a variety of audio clips. From Table 5, we observe that most of the attacks had minimal effect on the correlation value. The attacks

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>addbrumm_100</td>
<td>0.7681</td>
<td>addbrumm_1100</td>
<td>0.7741</td>
<td>addbrumm_2100</td>
<td>0.8553</td>
</tr>
<tr>
<td>addbrumm_3100</td>
<td>0.7958</td>
<td>addbrumm_4100</td>
<td>0.7958</td>
<td>addbrumm_5100</td>
<td>0.8361</td>
</tr>
<tr>
<td>addbrumm_6100</td>
<td>0.8065</td>
<td>addbrumm_7100</td>
<td>0.7516</td>
<td>addbrumm_8100</td>
<td>0.7092</td>
</tr>
<tr>
<td>addbrumm_9100</td>
<td>0.7092</td>
<td>addbrumm_10100</td>
<td>0.7851</td>
<td>addfftnoise</td>
<td>0.8573</td>
</tr>
<tr>
<td>addnoise_100</td>
<td>0.7730</td>
<td>addnoise_300</td>
<td>0.8756</td>
<td>addnoise_500</td>
<td>0.7448</td>
</tr>
<tr>
<td>addnoise_700</td>
<td>0.7946</td>
<td>addnoise_900</td>
<td>0.8066</td>
<td>addsinus</td>
<td>0.9029</td>
</tr>
<tr>
<td>amplify</td>
<td>0.5774</td>
<td>compressor</td>
<td>0.9037</td>
<td>copysample</td>
<td>0.7698</td>
</tr>
<tr>
<td>cutsamples</td>
<td>1</td>
<td>dynnoise</td>
<td>0.7997</td>
<td>echo</td>
<td>0.7715</td>
</tr>
<tr>
<td>exchange</td>
<td>0.6251</td>
<td>extrastereo_30</td>
<td>1</td>
<td>extrastereo_50</td>
<td>1</td>
</tr>
<tr>
<td>extrastereo_70</td>
<td>1</td>
<td>fft_hilpass</td>
<td>0.8117</td>
<td>fft_invert</td>
<td>0.7707</td>
</tr>
<tr>
<td>fft_real_reverse</td>
<td>0.7730</td>
<td>fft_stat1</td>
<td>1</td>
<td>fft_test</td>
<td>1</td>
</tr>
<tr>
<td>flipsample</td>
<td>0.7645</td>
<td>invert</td>
<td>0.8525</td>
<td>lsbzero</td>
<td>1</td>
</tr>
<tr>
<td>normalize</td>
<td>1</td>
<td>nothing</td>
<td>1</td>
<td>original</td>
<td>0.4008</td>
</tr>
<tr>
<td>rc_highpass</td>
<td>0.7180</td>
<td>rc_lowpass</td>
<td>1</td>
<td>resampling</td>
<td>0.8660</td>
</tr>
<tr>
<td>smooth</td>
<td>0.8660</td>
<td>smooth2</td>
<td>0.7303</td>
<td>stat1</td>
<td>fail</td>
</tr>
<tr>
<td>stat2</td>
<td>1</td>
<td>voiceremove</td>
<td>fail</td>
<td>zerocharge</td>
<td>1</td>
</tr>
<tr>
<td>zerolength</td>
<td>0.5774</td>
<td>zeroremove</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5: Watermark detection results on audio clips attacked with the Stirmark Audio Benchmark
that reduced significantly the correlation value (such as stat1 and voiceremove), had a strong impact on the fidelity of the recording, so that the attacked clip almost did not resemble the original.

### 3.4.7 Computational complexity

The execution time (etime) for evaluating masking threshold from existing psychoacoustic model and from proposed model on MATLAB 7.7 using Intel Core i3 CPU, 32 bit operating system was evaluated. Table 3.6 shows that computation time required for evaluating masking threshold from DFT magnitude (proposed technique) is much less then the execution time for frequency masking threshold of MPEG/audio psychoacoustic model.

### 3.5 Conclusion

The proposed method introduces the watermark in the spectral domain by exploiting the amplitude spreading effect of DFT components of the audio signal. Our watermark is imperceptibly embedded into the audio signal and is easy to detect by the author.

In case of likelihood detection, for the same range of probability of false alarm the variations in detection threshold is very small when scaling factor is obtained from proposed technique. Whereas the variations are quite high for other values of scale factor. Hence with decreasing value of scale factor, the detection threshold also decreases which in turn degrades the LLR detector response. However as value of scaling factor $a$ decreases, perceptual quality of watermarked audio becomes better. The proposed scale factor gives good tradeoff between LLR detector response and perceptual quality. From statistical analysis, we also find that in case of correlation detector the detection threshold is not effected by scaling factor. It remains constant for almost

<table>
<thead>
<tr>
<th>execution time</th>
<th>Mono(Speech)</th>
<th>Drum</th>
<th>Flute</th>
<th>Stereo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global masking threshold</td>
<td>7.89 msec</td>
<td>8.2057 msec</td>
<td>6.4116 msec</td>
<td>7.0668 msec</td>
</tr>
<tr>
<td>Maximum Amplitude Spread</td>
<td>2.49 msec</td>
<td>2.46 msec</td>
<td>2.4336 msec</td>
<td>2.776 msec</td>
</tr>
</tbody>
</table>

Table 3.6: Comparison of execution time between proposed and existing model.
all values of scale factor at a given probability of false alarm. Therefore embedder can adjust the watermark strength as per desire, without informing the correlation detector.

For the proposed value of scale factor both the statistical detectors give receiver operating characteristics which are close to actual experimental value. However \( LLR \) detector gives better approximation to experimental ROC as compared to correlation detector, for other values of scaling factor.

The major contributions of proposed work to the field are as follows:

- **Masking threshold for DFT magnitude**: The novelty of the presented work is that perceptual masking is defined in terms for DFT magnitudes. As a result, for multiplicative watermarks embedded in the magnitude of DFT coefficients, correlation detector gives optimal results (minimal error probability) when magnitudes follow Weibull distribution. The computations done for evaluating the frequency and temporal masking thresholds is not required in this method.

- **High embedding capacity**: The watermark embedding factor, for various critical bands of the signal spectrum is evaluated for finding the regions of highest watermark embedding capacity with least perceptual distortion.

- **Blind detection**: The scheme uses blind detection for watermark detection i.e. detector does not require original copy of the audio signal to detect watermark from the received audio signal. The detector has access to the secret key that is the only information that detector has about the embedding.

- **Robust to manipulation and signal processing operations**: Performance of proposed correlation detector and log-likelihood ratio detector for watermarked audio is investigated in terms of perceptual quality and robustness. The scheme is found robust to various signal processing attacks like presence of multiple watermarks, Additive white gaussian noise (AWGN) and MP3 compression.