UNIVERSITY OF MIAMI

SCHOOL OF MUSIC

DIGITAL WATERMARKING OF AUDIO SIGNALS USING A PSYCHOACOUSTIC AUDITORY MODEL AND SPREAD SPECTRUM THEORY

BY

RICARDO A. GARCÍA

A Research Project

Submitted to the faculty of the University of Miami in partial fulfillment of the requirements for the degree of Master of Science in Music Engineering Technology.

Approved by:

____
Prof. Ken Pohlmann, Project Advisor
Program Director
Music Engineering Technology

____
Prof. Will Pirkle
Music Engineering Technology

____
Dr. Michael Scordilis
Electrical and Computer Engineering

Coral Gables, Florida
April 21, 1999

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DIGITAL WATERMARKING OF AUDIO SIGNALS USING A PSYCHOACOUSTIC AUDITORY MODEL AND SPREAD SPECTRUM THEORY

Abstract:

A new algorithm for embedding a digital watermark into an audio signal is proposed. It uses spread spectrum theory to generate a watermark resistant to different removal attempts and a psychoacoustic auditory model to shape and embed the watermark into the audio signal while retaining the signal's perceptual quality. Recovery is performed without knowledge of the original audio signal. A software system is implemented and tested for perceptual transparency and data-recovery performance.

1 INTRODUCTION

Every day the amount of recorded audio data and the possibilities to distribute it (i.e. by the Internet, CD recorders, etc) are growing. These factors can lead to an increase in the illicit recording, copying and distributing of audio material without respect to the copyright or intellectual property of the legal owners. Another concern is the tracking of audio material over broadcast media without the use of human listeners or complicated audio recognition devices. Audio watermarking techniques promise a solution to some of these problems.

The concept of watermarking has been used for years in the fields of still and moving images. The basic idea of a watermark is to include a special "code" or information within the transmitted signal. This code should be transparent to the user (non-perceptible) and resistant against removal attacks of various types.

In audio signals, the desired characteristics can be translated into:

- Not perceptible (the audio information should appear "the same" to the average listener before and after the code is embedded).
- Resistant to degradation because of analog channel transmission. (i.e. TV, radio and tape recording).
- Resistant to degradation because of uncompressed-digital media. (i.e. CD, DAT and wav files).
- Resistant to removal through the use of sub-band coders or psychoacoustic models. (i.e. MPEG, Atrac, etc).

The proposed algorithm generates a digital watermark (i.e. a bit stream) that is spectrally shaped and embedded into an audio signal. Spread spectrum theory is used in the generation of the watermark. The strength of coded direct-sequence/binary-phase-shift keying (DS/BPSK) is used to create a robust watermark. The concepts are adapted to better deal with audio signals in a restricted audio bandwidth. A psychoacoustic auditory model is applied to shape and embed the watermark into the audio signal while retaining its perceptual quality for the average listener.

A complete psychoacoustic auditory model is explained in detail. This information is useful for other applications involving auditory models. The spread spectrum encoding and decoding processes are then presented. The algorithm performs an analysis of the incoming signal and searches the frequency domain for "holes" in the spectrum where the spread spectrum data can be placed without being perceived by the listener. The psychoacoustic auditory model is used to find these frequency "holes."

After transmission, the receiver recovers the embedded spread spectrum information and decodes it in order to reconstruct the original bit stream (watermark). There is no need for the receiver to have access to the original audio signal.

The algorithm is implemented in a software system to create an encoder and decoder, and its performance is evaluated for diverse channels and audio signals. The survival of the watermark (number of correct bytes/second) is analyzed for different configurations of the encoding system. Each one of these configurations is tested for transparency using an ABX listening test and for different channels (i.e. AM Radio, FM stereo radio, Mini Disc, MPEG layer 3, D/A – A/D conversion, etc).

2 PSYCHOACOUSTIC AUDITORY MODEL

An auditory model is an algorithm that tries to imitate the human hearing mechanism. It uses knowledge from several areas such biophysics and psychoacoustics.

From the many phenomena that occur in the hearing process, the one that is the most important for this model is “simultaneous frequency masking.” 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 needed to form a psychoacoustic auditory model are condensed in Figure 1. The first step is to translate the actual audio frame signal into the frequency domain using the Fast Fourier Transform. In the frequency domain the power spectrum, energy per critical band and the spread energy per critical band are calculated to estimate the masking threshold. This masking threshold is used to shape the “noise or watermark” signal to be imperceptible (below the threshold). Finally frequency domain output is translated into the time domain and the next frame is processed.

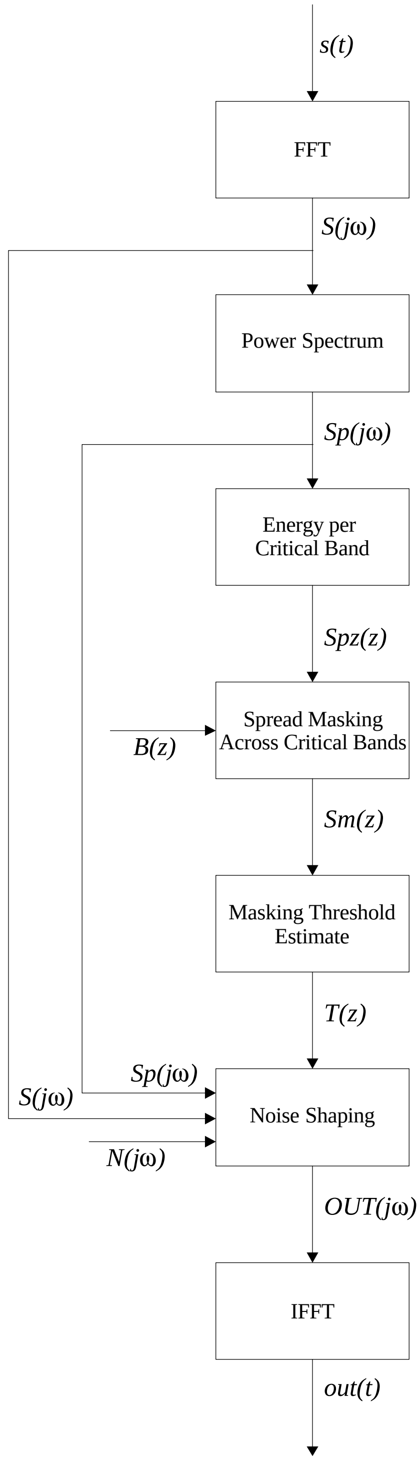

Figure 1. Psychoacoustic auditory model

2.1 SHORT TIME FOURIER TRANSFORM (STFT)

The cochlea can be considered as a mechanical to electrical transducer, and its function is to make a time to frequency transformation of the audio signal. To be more specific, the audio information, in time, is translated in first instance into a frequency-spatial representation inside the basilar membrane. This spatial representation is perceived by the nervous system and translated into a frequency-electrical representation.

This phenomenon is modeled using the short time Fourier Transform (STFT). The STFT uses successive, overlapped windows from the time domain input signal. If the input and the window signals are represented in discrete form:

s(m)=input signal in time domain

w(n-m)=window in time domain

The short term FT of $Sw(j\boldsymbol{w})$ is defined as:

$$\begin{aligned} Sw(j\boldsymbol{w}) &= \mathrm{FT}\{w(n)s(n)\} \\ Sw(j\boldsymbol{w}) &= \sum_{m=-\infty}^{\infty} w(n-m)s(m)e^{-j\boldsymbol{w}} \end{aligned} \tag{1}$$

where *n* is an integer representing the time index, and the window, *w(n-m)* is called an analysis window that determines what section of the input data is being analyzed using the offset index, *m*. The inverse short term FT is denoted IFT and is defined as:

$$\begin{aligned} w(n-m)s(n) &= \mathrm{IFT}\{Sw(j\boldsymbol{w})\} \\ w(n-m)s(n) &= \frac{1}{2\boldsymbol{p}}\int_{-\boldsymbol{p}}^{\boldsymbol{p}} Sw(j\boldsymbol{w})e^{j\boldsymbol{w}m}d\boldsymbol{w} \end{aligned} \tag{2}$$

2.2 SIMULTANEOUS FREQUENCY MASKING AND BARK SCALE

Simultaneous masking of sound occurs when two sounds are played at the same time and one of them is masked or “hidden” because of the other. The formal definition says that masking occurs when a test tone or “maskee” (usually a sinusoidal tone) is barely audible in the presence of a second tone or “masker.” The difference in sound pressure level between the masker and maskee is called the “masking level.”[1]

It is easier to measure the masking level for narrow band noise maskers (with a defined center frequency) and sinusoidal tone maskees. Figure 2 (a) and (b) display some curves that show the masking threshold for different narrow band noise maskers centered at 70, 250, 1000 and 4000 Hz. The level of all the maskers is 60 dB. The broken line represents the “threshold in quiet.” Average listeners will not hear any sound below this threshold. Figure 2 (a) uses a linear and (b) uses a logarithmic frequency scales.

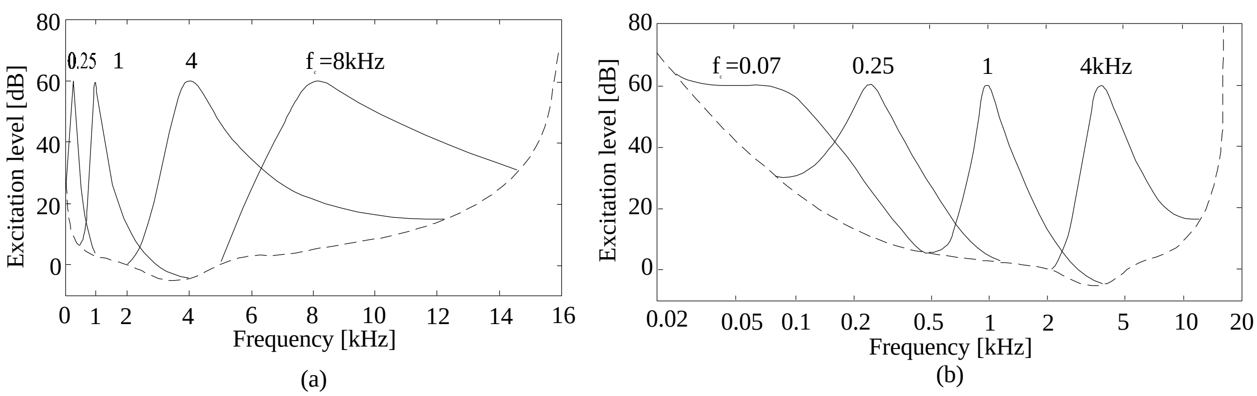

Figure 2. Masking curves in (a) linear and (b) logarithmic frequency scale [1]

The shape of all the masking curves is very different across the frequency range in both graphs. There are some similarities in the shape of the curves below 500 Hz in the linear frequency scale (a), and some similarities above 500 Hz in the logarithmic frequency scale (b). A more useful scale has been introduced that is known as “critical band rate” or “Bark scale.” The concept of the Bark scale is based on the well-researched assumption [1] that the basilar membrane in the hearing mechanism analyzes the incoming sound through a spatial-spectral analysis. This is done in small sectors or regions of the basilar membrane that are called “critical bands.” If all the critical bands are added together in a way that the upper limit of one is the lower limit of the next one, the critical band rate scale is obtained. Also a new unit has been introduced, the “Bark” that is by definition one critical band wide.

Figure 3 (a) and (b) show the relationship between the Bark scale and the linear and logarithmic frequency scales.

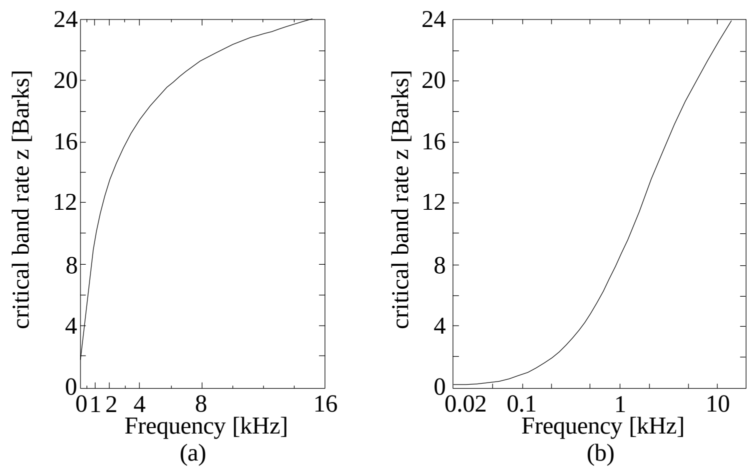

Figure 3. Relationship between Bark scale and (a) linear and (b) logarithmic frequencies

Figure 4 shows the same masking curves from Figure 2 in a Bark scale. Notice that the shape of the masking curves is almost identical across the frequency range. Various approximations may be used to translate frequency into a Bark scale [2]:

$$z = 13\tan^{-1}\left(\frac{0.76 * f}{1000}\right) + 3.5\tan^{-1}\left(\left(\frac{f}{7500}\right)^2\right) \qquad \textbf{(3)}$$

and [3]:

$$z = \frac{26.81 * f}{1960 + f} - 0.53 \tag{4}$$

where *f* is the frequency in Hertz and *z* is the mapped frequency in Barks.

Eq. (3) is more accurate, but the Eq. (4) is easier to compute. Figure 4 shows the excitation level of several narrow band noises with diverse center frequencies in a Bark scale.

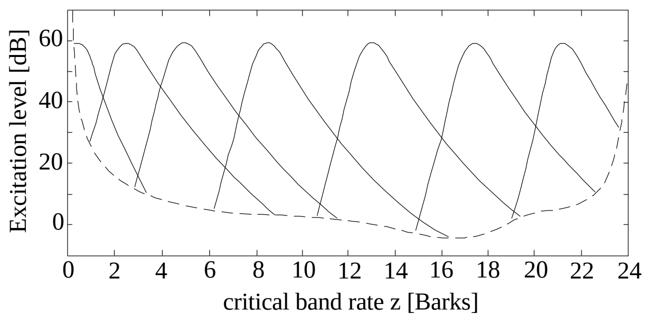

Figure 4. Excitation level versus critical band rate for narrow band noises with various center frequencies [1]

2.3 POWER SPECTRA

The first step in the frequency domain (linear, logarithmic or bark scales) is to calculate the power spectra of the incoming signal. This is calculated with:

$$\begin{aligned} Sp(j\boldsymbol{w}) &= \mathrm{Re}\{Sw(j\boldsymbol{w})\}^2 + \mathrm{Im}\{Sw(j\boldsymbol{w})\}^2 \\ &= |Sw(j\boldsymbol{w})|^2 \end{aligned} \tag{5}$$

The energy per critical band, $Spz(z)$, is defined as:

$$Spz(z) = \sum_{\boldsymbol{w}=LBZ}^{HBZ} Sp(j\boldsymbol{w}) \tag{6}$$

with:

$$\begin{aligned} z &= 1,2,\ldots,Zt \\ LBZ &= \text{Lower frequency in critical band z} \\ HBZ &= \text{Higher frequency in critical band z} \end{aligned}$$

The power spectrum *Sp(jw)* and the energy per critical band *Spz(z)* are the base of the analysis in the frequency domain. They will be used to compute the spread masking threshold. Figure 5 shows the power spectrum *Sp(jw)* and the energy per critical band *Spz(z)* for a particular frame.

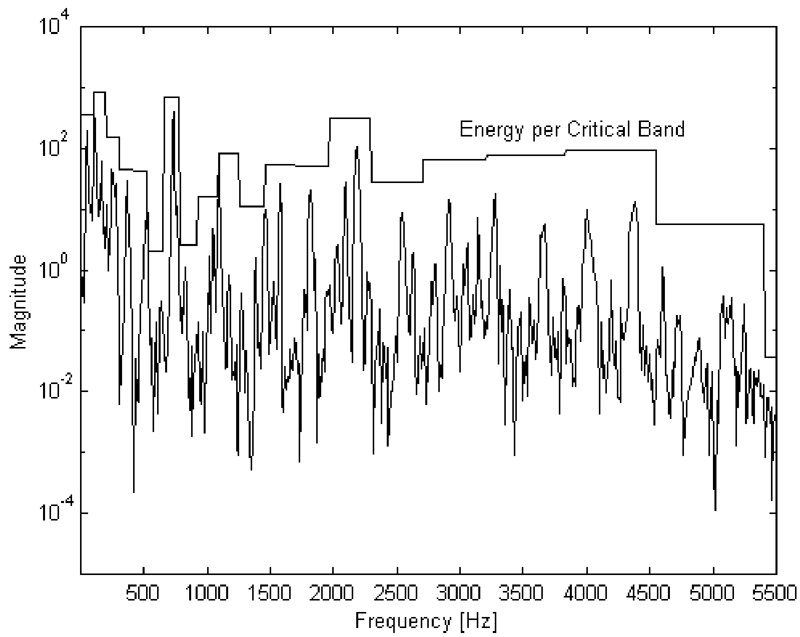

Figure 5. Power Spectrum *Sp(jw)* **and Energy per Critical Band** *Spz(z)*

2.4 BASILAR MEMBRANE SPREADING FUNCTION

Figure 6 shows some experimental data of basilar-membrane spreading-curves of a narrow-band noise masking a tone of a given frequency. The level of the masker is varied from 20 dB to 100 dB SPL in steps of 20 dB. Note the difference in the right slope with the different levels. The left side is almost the same for all masker levels.

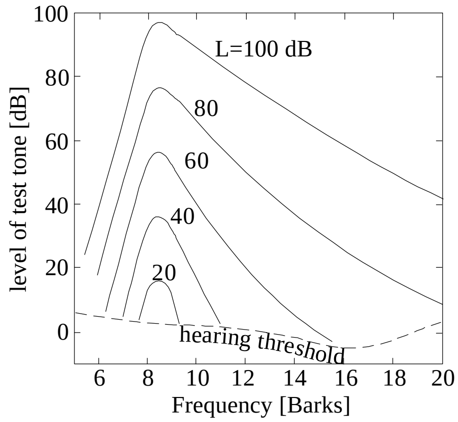

Figure 6. Spreading curves of narrow band noise masking a tone of given frequency [1]

A model that approximates the basilar membrane spreading function, without taking in account the change in the upper slope is defined [3]:

$$B(z) = 15.91 + 7.5(z + 0.474) - 17.5\sqrt{1 + (z + 0.474)^2} \qquad \textbf{(7)}$$

where *z* is the normalized Bark scale. Figure 7 shows $B(z)$.

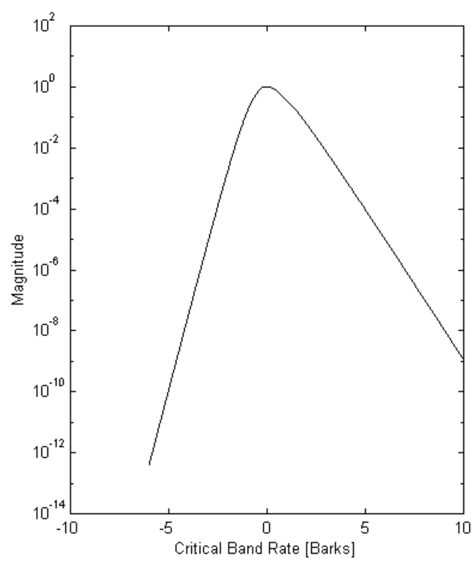

Figure 7. Model of the spreading function, $B(z)$, using Eq. (7)

The auditory model uses the information about the energy in each critical band given by Eq. (6) and uses Eq. (7) to calculate the spread masking across critical bands $Sm(z)$. This is done using:

$$Sm(z) = Spz(z) * B(z) \tag{8}$$

This operation is a convolution between the basilar membrane spreading function and the total energy per critical band. A true spreading calculation should include all the components in each critical band, but for the purposes of this algorithm, the use of the energy per critical band $Spz(z)$ is a close approximation. $Sm(z)$ can be interpreted as the energy per critical band after taking in account the masking occasioned by neighboring bands. Figure 8 shows $Sm(z)$.

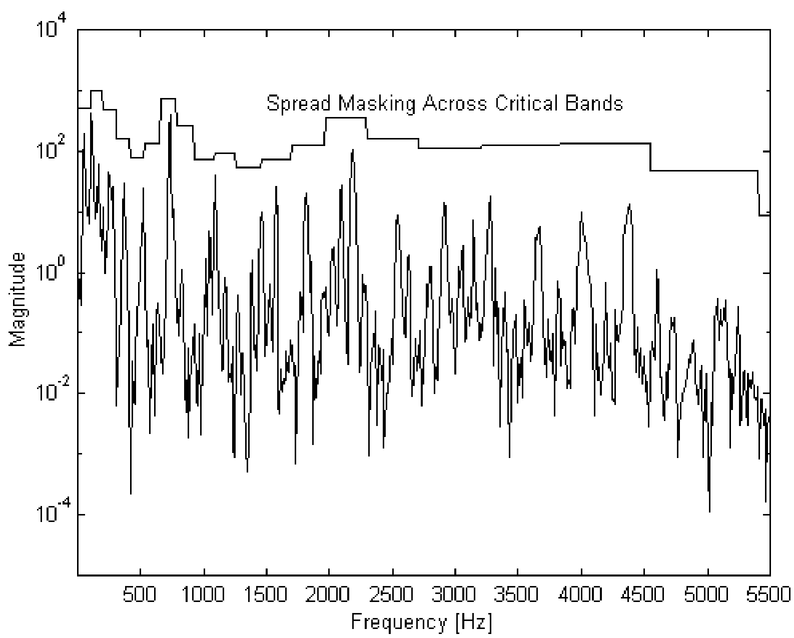

Figure 8. Spread masking across critical bands $Sm(z)$

2.5 MASKING THRESHOLD ESTIMATE

2.5.1 MASKING INDEX

There are two different indexes used to model masking. The first one is used when a tone is masking noise (masker = tone, maskee = noise), and it is defined to be *14.5 + z dB* below the spread masking across critical bands $Sm(z)$. In this case *z* is the center frequency of the masker tone using a bark scale. The second index is used when noise is masking a tone (masker = noise, maskee = tone), and is defined to be 5.5 dB below $Sm(z)$, regardless of the center frequency [4].

2.5.2 SPECTRAL FLATNESS MEASURE (SFM) AND TONALITY FACTOR (a)

The spectral flatness measure (*SFM*) is used to determine if the actual frame is noise-like or tone-like and then to select the appropriate masking index. The *SFM* is defined as the ratio of the geometric to the arithmetic mean of $Spz(z)$, expressed in dB as:

$$SFM_{dB} = 10\log 10 \left\{ \frac{\left[\prod_{z=1}^{Zt} Spz(z) \right]^{\frac{1}{Zt}}}{\frac{1}{Zt}\sum_{z=1}^{Zt} Spz(z)} \right\} \qquad (9)$$

with *Zt* = total number of critical bands on the signal

The value of the SFM is used to generate the “tonality factor” that will help to select the right masking index for the actual frame. The tonality factor is defined in [3], [4] as the minimum of the ratio of the calculated SFM over a SMF maxima and 1:

$$\alpha = \min\left(\frac{SFM_{dB}}{SFM_{dB\max}},\ 1\right) \tag{10}$$

with $SFM_{dB\max} = -60dB$.

Therefore, if the analyzed frame is tone-like, the tonality factor α will be close to 1, and if the frame is noise-like, α will be close to 0. This is shown for different types of frames and they respective spectrums in Figure 9 and Figure 10. The *SFM* and $\boldsymbol{\alpha}$ are shown in Table 1 and in Table 2, respectively.

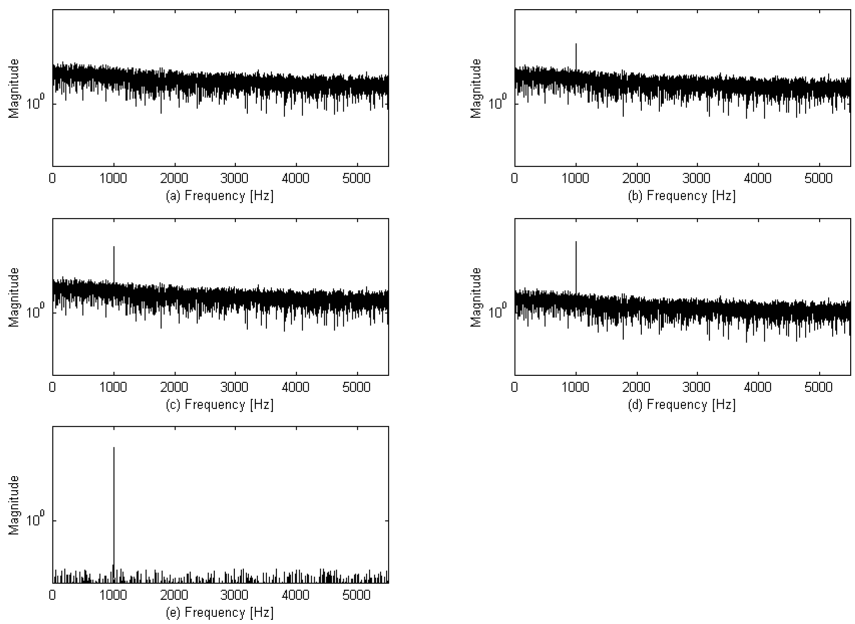

Figure 9. Spectrum of different noise+tone frames (a) 100% noise, (b) 75% noise - 25% tone, (c) 50% noise - 50% tone, (d) 25% noise - 75% tone, and (d) 100% tone

	SFM dB	$\boldsymbol{\alpha}$
(a)	*-0.358655*	*0.005978*
(b)	*-1.855607*	*0.030927*
(c)	*-7.426313*	*0.123772*
(d)	*-21.032517*	*0.350542*
(e)	*-93.419982*	*1*

Table 1. SFM and α for noise+tone frames

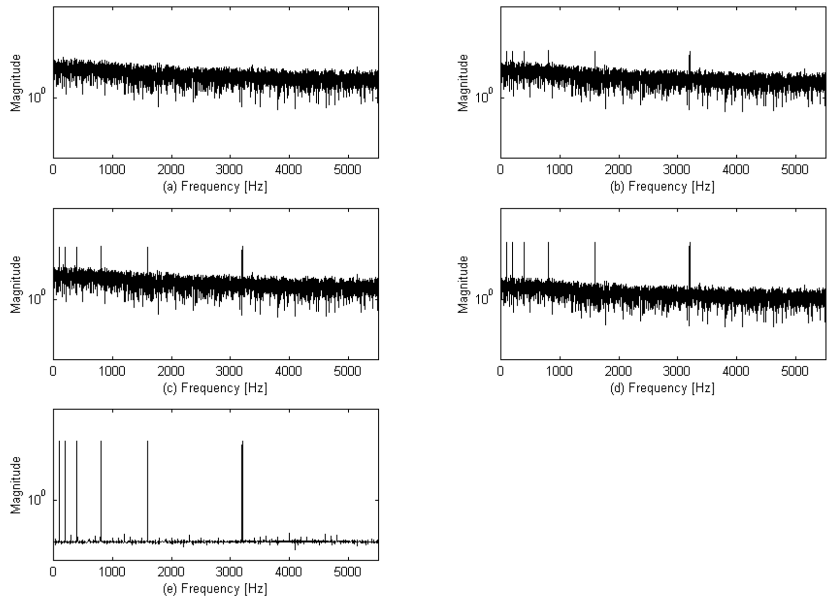

Figure 10. Spectrum of different noise+ harmonic frames (a) 100% noise, (b) 75% noise – 25% tone, (c) 50% noise - 50% tone, (d) 25% noise - 75% tone, and (d) 100% tone

	SFM dB	***a***
(a)	*-0.358655*	*0.005978*
(b)	*-0.591993*	*0.009867*
(c)	*-2.8772893*	*0.047955*
(d)	*-12.154526*	*0.202575*
(e)	*-48.920751*	*0.815346*

Table 2. SFM and α for noise+harmonic frames

The tonality factor α is used to calculate the masking energy offset *O(z)*, defined as [3], [4]:

$$O\ (z) = \boldsymbol{a}(14.5 + z) + (1 - \boldsymbol{a})5.5 \qquad \textbf{(11)}$$

The offset *O(z)*is subtracted from the spread masking threshold to estimate the raw masking threshold $Traw(z)$.

$$Traw(z) = 10^{\left(\log_{10}(Sm(z)) - \frac{O(z)}{10}\right)} \tag{12}$$

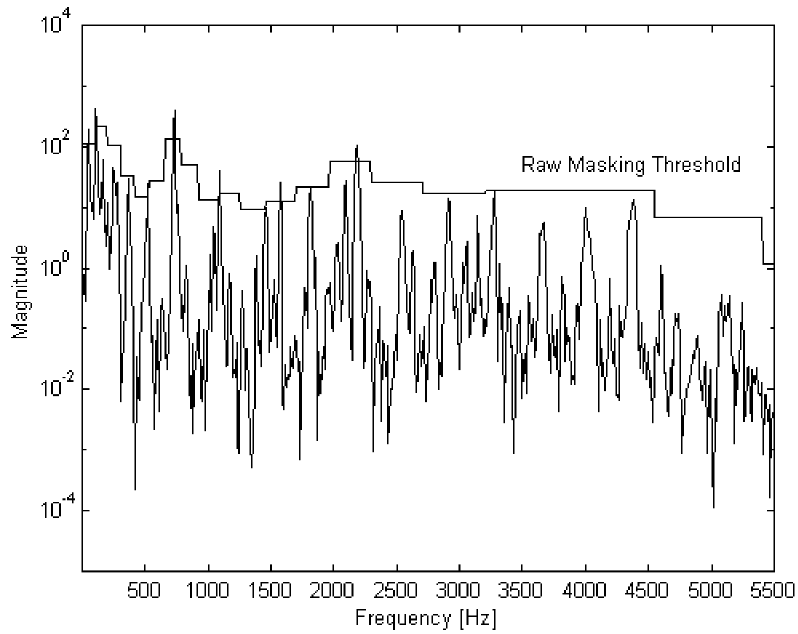

Figure 11. Raw masking threshold *Traw(z)*

2.5.3 THRESHOLD NORMALIZATION

The use of the spreading function $B(z)$ increases the energy level in each one of the critical bands of the spectrum $Sm(z)$. This effect has to be undone using a normalization technique, to return $Traw(z)$ to the desired level. The energy per critical band calculated with Eq. (6) is also affected by the number of components in each critical band. Higher bands have more components than lower bands, affecting the energy levels by a different amount. The normalization used [4] simply divides each one of the components of $Traw(z)$ by the number of points in the respective band P_z.

$$Tnorm(z) = \frac{Traw(z)}{P_z} \tag{13}$$

Where:

$$P_z = \text{number of points in each band } z$$
$$z = 1,2,\ldots.Zt$$

Figure 12 shows the normalized masking threshold *Tnorm(z)* and the hearing threshold *TH*.

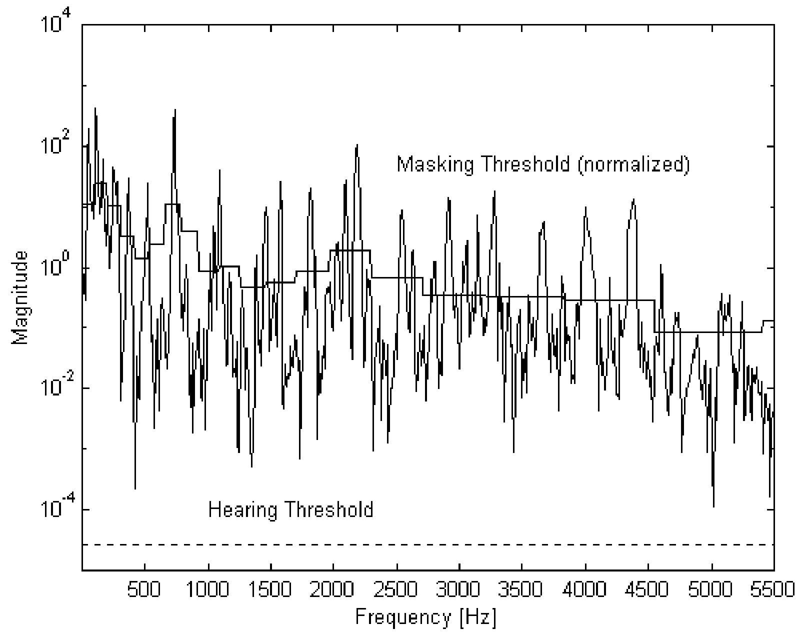

Figure 12. Masking threshold normalized *Tnorm(z)* and hearing threshold *TH* (dotted line)

2.5.4 FINAL MASKING THRESHOLD

After normalization, the last step is to take in to account the absolute auditory threshold or "hearing threshold." The hearing threshold varies across the frequency range as stated in Zwicker and Zwicker [1]. In the proposed auditory model the hearing threshold will be simplified to use the worst case threshold (the lowest). That is defined as a sinusoidal tone of 4000 Hz with one bit of dynamic range [4]. These values are chosen based on the data from experimental research that shows that the most sensitive range of the human ear is in the range of 2500 to 4500 Hz [1]. For a frequency of 4000 Hz, the measured sound intensity is 10^{-12} Watt/m^2, that equals a loudnes of 0 phons at that frequency [12].The chosen amplitude (one bit) is the smallest possible amplitude value in a digital sound format. The hearing threshold is then calculated with [4]:

$$TH = \max\left(\left|Pp(j\omega)\right|\right) \qquad (14)$$

where:

$$Pp(j\omega) = \text{power spectrum of the probe signal } p(t)$$
$$p(t) = \sin(2\pi 4000t)$$

The final threshold $T(z)$ is:

$$T(z) = \max(Tnorm(z), TH) \qquad (15)$$

2.6 NOISE SHAPING USING THE MASKING THRESHOLD

The objective of the auditory model is to find a usable masking threshold. The final masking threshold is always compared with the values of the power spectrum of the signal $Sp(j\omega)$. This can be interpreted as “below this threshold, the information is not relevant for human hearing.” This means that if the frequency components that fall below the masking threshold are removed; the average listener will notice no difference between the original sound signal and the altered version.

Another very important consequence of this is that if these components are not just discarded but replaced with new components they will be, as before, inaudible for the listener. This assumes that the new components do not change the average energy considerably in their critical band. Let the frame with the new components be called $N(j\boldsymbol{w})$. The objective is to use the final masking threshold to select which components from $Sp(j\boldsymbol{w})$ can be replaced with components from $N(j\boldsymbol{w})$. The components of $N(j\boldsymbol{w})$ are shaped to stay below the final masking threshold. The final signal, that includes components from $Sw(j\boldsymbol{w})$ and $N(j\boldsymbol{w})$, ideally retains the perceptual quality of the original signal for the average listener.

The following steps are used to remove the components from $Sw(j\boldsymbol{w})$, shape the vector $N(j\boldsymbol{w})$ and mix them:

Calculate the "new" version of the sound signal (after removing some components):

$$Swnew_i(j\boldsymbol{w}) = \begin{cases} Sw_i(j\boldsymbol{w}) & Sp_i(j\boldsymbol{w}) \geq T(z) \\ 0 & Sp_i(j\boldsymbol{w}) < T(z) \end{cases}$$
$$i = 1,2\ldots \text{number of components}$$
$$\text{z}, \boldsymbol{w} \text{ according to component } i \tag{16}$$

Remove the unneeded components in the $N(j\boldsymbol{w})$ vector:

$$Nnew_i(j\boldsymbol{w}) = \begin{cases} 0 & Sp_i(j\boldsymbol{w}) \geq T(z) \\ N_i(j\boldsymbol{w}) & Sp_i(j\boldsymbol{w}) < T(z) \end{cases}$$
$$i = 1,2\ldots \text{number of components}$$
$$\text{z}, \boldsymbol{w} \text{ according to component } i \tag{17}$$

Calculate the power spectrum of $Nnew(j\boldsymbol{w})$:

$$Nnewp(j\boldsymbol{w}) = \left| Nnew(j\boldsymbol{w}) \right|^2 \tag{18}$$

and then, the energy per critical band:

$$Nnewpz(z) = \sum_{\boldsymbol{w}=LBZ}^{HBZ} Nnewp(j\boldsymbol{w}) \tag{19}$$

with:

$$\begin{aligned} \mathrm{z} &= 1,2,\ldots,\mathrm{Zt} \\ \mathrm{LBZ} &= \text{Lower frequency in critical band z} \\ \mathrm{HBZ} &= \text{Higher frequency in critical band z} \end{aligned}$$

The shaping is done applying a factor F_z to each critical band. These factors are given by:

$$\begin{aligned} F_z &= A\frac{\sqrt{T(j\boldsymbol{w})}}{\max(|Nnew(j\boldsymbol{w})|)} \\ z &= 1,2\ldots Zt \\ \boldsymbol{w} &= LBZ \text{ to } HBZ \text{ for each band } z \end{aligned} \tag{20}$$

The coefficient A is used as the "gain of the noise signal". Varies from 0 to 1 and weights the embedded noise below the threshold of masking. The factors F_z are applied using:

$$\begin{aligned} Nfinal(j\boldsymbol{w}) &= Nnew(j\boldsymbol{w})F_z \\ z &= 1,2\ldots Zt \\ \boldsymbol{w} &= LBZ \text{ to } HBZ \text{ for each band } z \end{aligned} \tag{21}$$

The final step is to mix both spectrums, the altered $Swnew(j\boldsymbol{w})$ and the shaped $Nfinal(j\boldsymbol{w})$ to form the composite signal $OUT(j\boldsymbol{w})$:

$$OUT(j\boldsymbol{w}) = Swnew(j\boldsymbol{w}) + Nfinal(j\boldsymbol{w}) \tag{22}$$

2.7 DISCUSSION

In this chapter, a complete psychoacoustic auditory model was introduced, and the intended application of shaping the noise using the masking threshold was explained step by step. This will be used in the watermarking algorithm to embed the watermark into the audio signal while retaining the perceptual quality of the original audio signal. The next chapter will cover the watermark generation.

3 SPREAD SPECTRUM

One of the requirements of a watermarking algorithm is that the watermark should resist multiple types of removal attacks. A removal attack is considered as anything that can degrade or destroy the embedded watermark. Another factor to be considered is that the masking threshold of the actual audio signal determines the embedding of the watermark, because the watermark is embedded in the “spare components” found using the psychoacoustic auditory model. From this point of view, the watermark has to be the least intrusive to the audio signal, and therefore, the actual audio data can be seen as the main obstacle for a good watermarking algorithm. This is because the audio will use all the needed bandwidth and the watermark will use what is left after the auditory model analysis.

The desired watermarking technique should be resistant to degradation because of:

- The used transmission channel: analog or digital.
- High-level wide-band noise (in this case, the “noise” is the actual audio signal). This is often related as “low signal to noise ratio”.
- The use of psychoacoustic algorithms on the final watermarked audio.

A communication theory technique that meets the requirements is the “spread spectrum technique”, as described thoroughly in Simon et al. [5] and Pickholtz et al. [6].

"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 despreading and subsequent data recovery." [6]

In the following analysis, the process of generating a watermark that will be embedded in an audio signal is expressed in spread spectrum terminology. The original audio signal will be called "noise" and the bit stream that conforms the watermark sequence will be the data signal. The watermark sequence is transformed in a watermark audio signal and then the audio signal (noise) is added to it. This process of adding noise to a channel or signal is called "jamming." The objective of a jammer in a communication system is to degrade the performance of the transmission, exploiting knowledge of the communication system. In the watermark algorithm the audio signal (i.e. music) is considered the jammer, and it has much more power than the transmitted bit stream (watermark).

3.1 BASIC CONCEPTS

The primary challenge that a receiver must overcome is intentional jamming, especially if the jammer has much more power than the transmitted signal. Classical communications theoretical investigations about additive white Gaussian noise help to analyze the problem. White Gaussian noise is a signal which has infinite power spread uniformly over all frequencies; but even under these circumstances communication can be achieved due to the fact that on each of the "signal coordinates" the power of the noise component is limited (not infinite). Therefore, if the noise component in the signal coordinates is not too large, communication can be made. This is usually applied in a typical narrow-band signal, where just the noise components in the signal bandwidth are taken into account as possible factors that can do harm to the communication. With this knowledge, the best strategy to combat intentional jamming is to select signal coordinates where the jammer to signal ratio is the smallest possible.

Assume a communication link with many signal coordinates available to choose from, and only a small subset of these is used at any time. If the jammer can not determine which subset is being used, it is forced to jam all the coordinates and therefore, all its power will be distributed among all the coordinates, with little power in each of them. If the jammer chooses to jam only some of the coordinates, the power over each of them is larger, but the jammer lacks the knowledge of which coordinates to jam. The protection against the jammer is enhanced, as more signal coordinates are available to choose from.

Having a signal of bandwidth *W* and duration *T*, the number of coordinates available is given by:

$$N \cong \begin{cases} 2WT & \text{coherent signals} \\ WT & \text{non - coherent signals} \end{cases} \tag{23}$$

T is the time used to send a standard symbol. To make *N* larger when *T* is fixed, two techniques can be applied:

- Direct sequence spreading (DS)
- Frequency hopping (FH)

The signals created with these techniques are called “spread spectrum signals.”

3.1.1 MODELS AND FUNDAMENTAL PARAMETERS

The basic system is shown in Figure 13, with the following parameters:

W_{ss} = Total spread spectrum signal bandwidth available
R_b = Data rate (bits / second)
S = Signal power (at the input of the receiver)
J = Jammer power (at the input of the receiver)

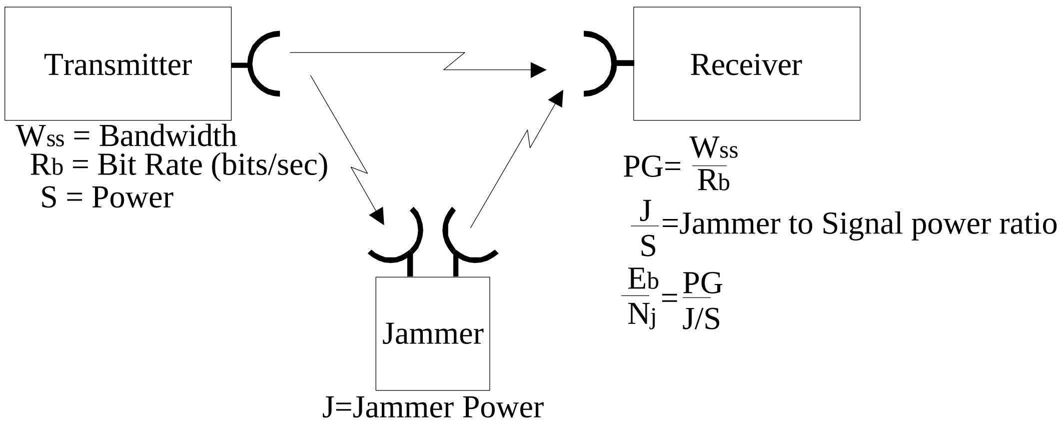

Figure 13. Basic spread spectrum communications system

W_{ss} is defined as the total available spread spectrum bandwidth that could be used by the transmitter, but it is not guaranteed that it will be used during the actual transmission. Neither is it guaranteed that the spectrum will be continuous. R_b is the uncoded bit data rate used during transmission. The signal and the jammer powers *S* and *J* are the averaged power at the receiver. This does not change even if the jammer and/or the signal are pulsating.

Regardless of the signal and jammer waveforms, equivalent bit energy to jammer noise ratio is defined as:

$$\frac{E_b}{N_J} = \frac{W_{ss}S}{R_b J} \qquad \textbf{(24)}$$

And the processing gain is defined as:

$$PG = \frac{W_{ss} S}{R_b} \quad \textbf{(25)}$$

where $\frac{J}{S}$ = jammer to signal power ratio

Translating to decibels, the *bit energy-to-jammer noise ratio:*

$$\frac{E_b}{N_J}{}_{dB} = [PG]_{dB} - [J/S]_{dB} \quad \textbf{(26)}$$

where:

$$\begin{aligned} [PG]_{dB} &= 10\log_{10}\left(\frac{W_{ss}}{R_b}\right) \\ [J/S]_{dB} &= 10\log_{10}\left(\frac{J}{S}\right) \end{aligned} \quad \textbf{(27)}$$

3.1.2 JAMMER WAVEFORMS

The number of possible jammer waveforms that a jammer can apply to a communication system is infinite, but the two principal types will be analyzed and the result will be used to determine some desired characteristics of the system.

3.1.2.1 Broadband and Partial-Band Noise Jammers

A broadband noise jammer spreads Gaussian noise of a total power *J* evenly over the total frequency range of the spread bandwidth W_{ss}. This is shown in Figure 14.

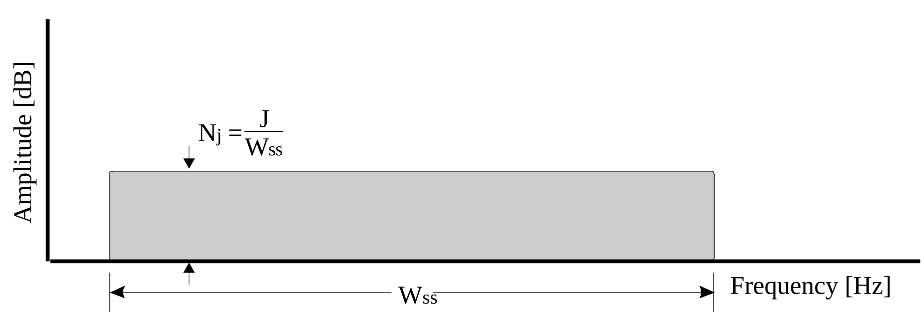

Figure 14. Broadband noise jammer

The equivalent single sided noise power spectral density is:

$$N_J = \frac{J}{W_{SS}} \tag{28}$$

The signal energy per bit is ST_b where

$$\begin{aligned} T_b &= \frac{1}{R_b} \\ E_b &= \frac{S}{R_b} \end{aligned} \tag{29}$$

In this case, Eq. (29) is exactly the bit energy to jammer noise ratio.

The only knowledge of the communication system that the broad band noise jammer is exploiting is its spread bandwidth W_{SS}. The bit error probability of this system with a broadband noise jammer is the same that with additive white Gaussian noise of one-sided spectral density equal to N_J. A partial band noise jammer is shown in Figure 15:

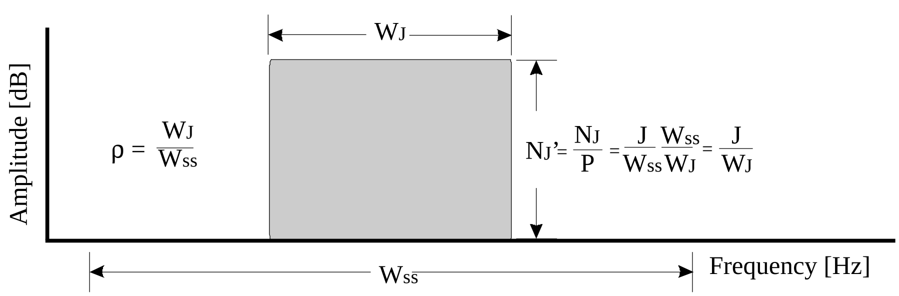

Figure 15. Partial band noise jammer

This noise jammer spreads noise of total power J evenly over a frequency range of bandwidth W_J, which is contained in the total spread bandwidth W_{ss}. Then, $\boldsymbol{r}$ is defined as the ratio:

$$\boldsymbol{r} = \frac{W_J}{W_{ss}} \leq 1 \tag{30}$$

This is the fraction of the total spread spectrum bandwidth that is being jammed. That band has a noise of total power J, and spectral power density:

$$\begin{aligned} \frac{J}{W_J} &= \frac{J}{W_{SS}} \cdot \frac{W_{SS}}{W_J} \\ &= \frac{N_J}{\boldsymbol{r}} \end{aligned} \tag{31}$$

3.1.2.2 Pulse Jammer

When a jammer transmits the jammer waveform during a fraction $\boldsymbol{r}$ of the time, the average power is J, but the peak power during transmission is given by:

$$J_{\text{peak}} = \frac{J}{\boldsymbol{r}} \tag{32}$$

3.2 COHERENT DIRECT-SEQUENCE SYSTEMS

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, but the uncoded is studied because is easier to understand and is the foundation of the coded scheme.

3.2.1 UNCODED DIRECT-SEQUENCE SPREAD BINARY PHASE-SHIFT-KEYING

Uncoded Direct-Sequence Spread Binary Phase-Shift-Keying is known as uncoded DS/BPSK. It may be explained with a simple example. BPSK signals are often expressed as:

$$
\begin{aligned}
&s(t) = \sqrt{2S}\,\sin\left[\boldsymbol{w}_0 t + \frac{d_n \boldsymbol{p}}{2}\right] \\
&nT_b \le t < (n+1)T_b, \qquad n = \text{integer}
\end{aligned}
\qquad \textbf{(33)}
$$

where

T_b is the data bit time $\left(\dfrac{1}{R_b}\right)$
$\{d_n\}$ is the sequence of data bits, with the possible values of 1 or –1; and equal probability of occurrence.

Eq. (33) can be expressed as:

$$
\begin{aligned}
&s(t) = d_n \sqrt{2S}\,\cos(\boldsymbol{w}_0 t) \\
&nT_b \le t < (n+1)T_b, \qquad n = \text{integer}
\end{aligned}
\qquad \textbf{(34)}
$$

BPSK can be seen as phase modulation in Eq.(33) or amplitude modulation in Eq. (34)).

The spectrum of a BPSK signal is usually of the form shown in Figure 16:

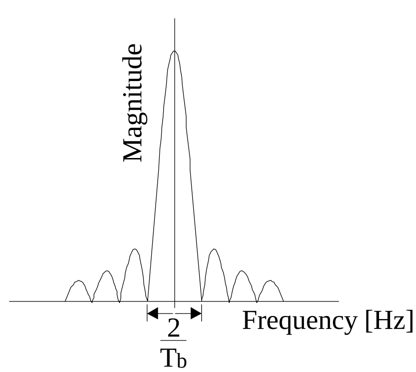

Figure 16. Spectrum of signal BPSK

This is a $\frac{(\sin^2 x)}{x^2}$ function, and the first null bandwidth is $\frac{1}{T_b}$. This shows the minimum bandwidth needed to transmit the signal *s(t)* and to recover it at the receiver.

Spread spectrum theory requires the signal to be spread over a larger spectrum than the minimum needed for transmission. The spreading of the direct sequence is done using a pseudorandom (PN) binary sequence $\{c\}$. The values of this sequence are 1 or –1 and its speed is *N* times faster than the $\{d\}$ data rate. The time, T_c, of each bit on a PN sequence is known as a "chip" and is given by:

$$T_c = \frac{T_b}{N} \tag{35}$$

The direct sequence spread spectrum signal has the form:

$$\begin{aligned} x(t) &= \sqrt{2S}\sin\left[\omega_0 t - d_n c_{nN+k}\pi/2\right] \\ &= d_n c_{nN+k}\sqrt{2S}\cos(\omega_0 t) \\ nT_b + kT_c &\le t < nT_b + (k+1)T_c \\ k &= 0,1,2\ldots N-1 \\ n &= \text{integer} \end{aligned} \tag{36}$$

The signal is very similar to the common BPSK, except that the bit rate is *N* times faster and the power spectrum is *N* times wider:

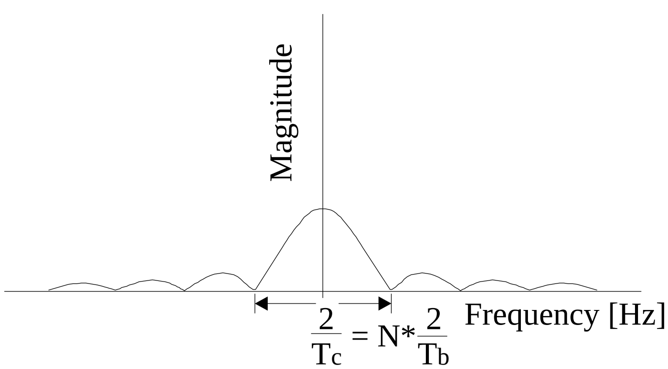

Figure 17. Spectrum of signal BPSK after spreading

The processing gain is given by:

$$PG = \frac{W_{SS}}{R_b} = N \tag{37}$$

W_{SS} is the direct sequence spread spectrum bandwidth $\frac{1}{T_c} = N\frac{1}{T_b}$.

If the data function is defined as:

$$\begin{aligned} d(t) &= d_n, \quad nT_b \le t < (n+1)T_b \\ n &= \text{integer} \end{aligned} \tag{38}$$

and the PN sequence is:

$$\begin{aligned} c(t) &= c_k, \quad kT_c \le t < (k+1)T_c \\ k &= \text{integer} \end{aligned} \tag{39}$$

Eq. (36) can be expanded as:

$$\begin{aligned} x(t) &= \sqrt{2S}\sin\left[\omega_0 t + c(t)d(t)\pi/2\right] \\ &= c(t)d(t)\sqrt{2S}\cos(\omega_0 t) \end{aligned} \tag{40}$$

Figure 18 shows the block diagram for the normal DS/BPSK modulation; and Figure 19 shows an equivalent model used in the next step of the analysis. Figure 20 shows the signals *d(t), c(t)* and *c(t)d(t)* with *N=7.*

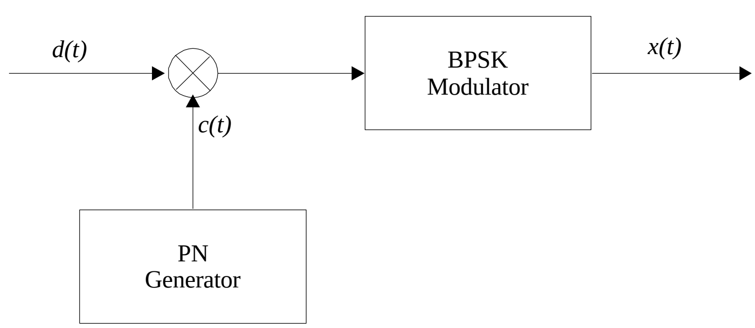

Figure 18. DS/BPSK modulation

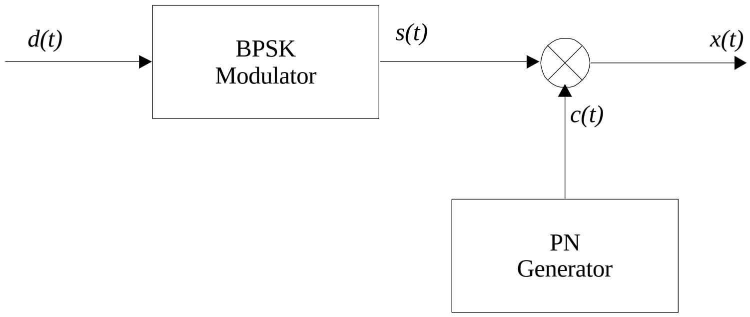

Figure 19. DS/BPSK modified

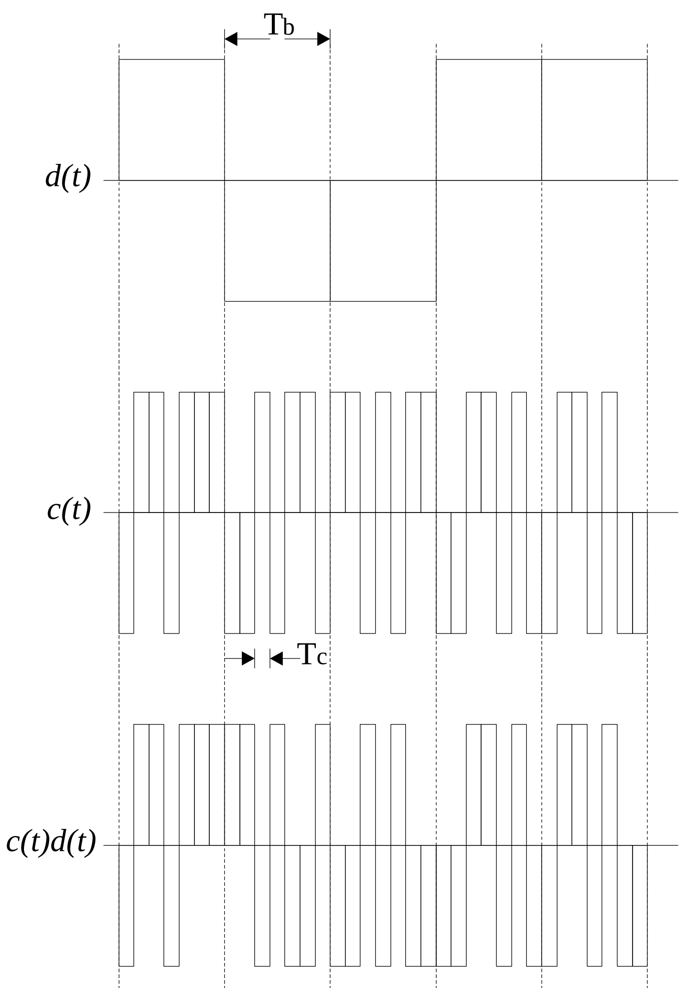

Figure 20. $d(t)$, $c(t)$ **and** $c(t)d(t)$ **example with** $N=7$**.**

From Figure 19, the equivalent form of $x(t)$ is given by:

$$x(t) = c(t)s(t) \tag{41}$$

where

$$s(t) = d(t)\sqrt{2S}\,\cos(\boldsymbol{w}_0 t) \tag{42}$$

This is the original BPSK signal. The property:

$$c^2(t) = 1 \quad \text{for all } t \tag{43}$$

is the key point exploited to "recover" the original BPSK signal:

$$c(t)x(t) = s(t) \tag{44}$$

If the receiver possesses a copy of the PN sequence and can synchronize the local copy with the received signal *x(t)*, it is able to de-spread the signal and recover the transmitted data.

3.2.1.1 Constant Power Broadband Noise Jammer

A jammer, *J(t)*, with constant power *J* is shown in Figure 21:

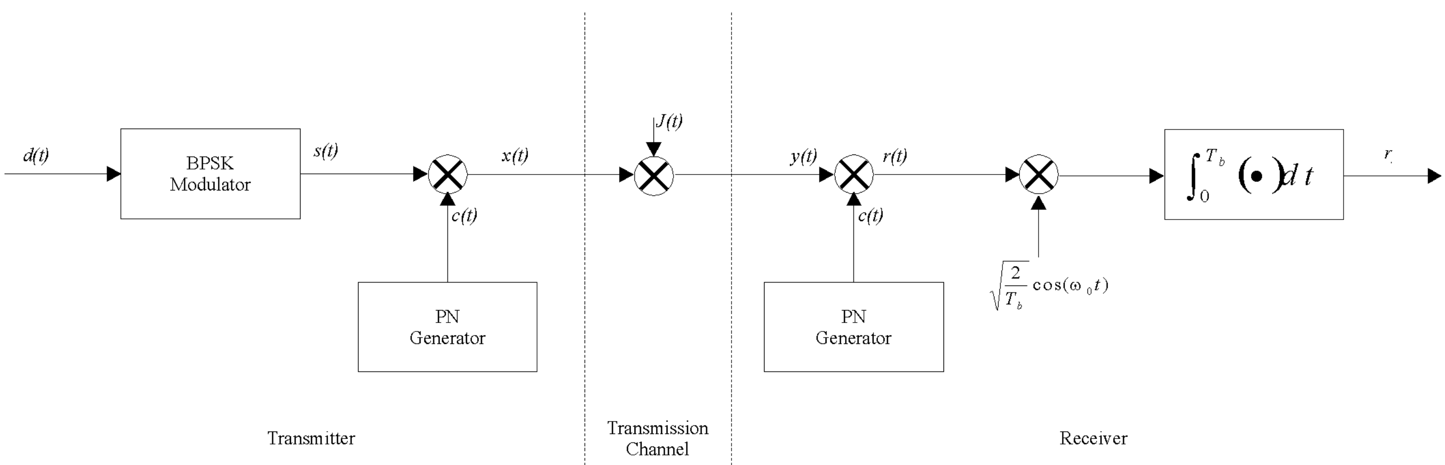

Figure 21. Uncoded DS/BPSK

The system is also assumed to have no noise from the transmission channel. An ideal BPSK demodulator is assumed after the received signal *y(t)* is multiplied by the PN sequence. The channel output is:

$$y(t) = x(t) + J(t) \tag{45}$$

This is multiplied by the PN sequence *c(t)*:

$$\begin{aligned} r(t) &= c(t)y(t) \\ &= c(t)x(t) + c(t)J(t) \\ &= s(t) + c(t)J(t) \end{aligned} \tag{46}$$

This term shows the original BPSK signal plus a noise given by *c(t)J(t).* The output of the conventional BPSK detector is then:

$$r = d\sqrt{E_b} + n \tag{47}$$

where:

d is the data bit for the actual T_b second interval.
$E_b = ST_b$ is the bit energy.
n *is* the equivalent noise component.

n is further defined as:

$$n = \sqrt{\frac{2}{T_b}} \int_0^{T_b} c(t)J(t)\cos(\boldsymbol{w}_0 t)dt \tag{48}$$

The usual decision rule for BPSK is:

$$\hat{d} = \begin{cases} 1, & \text{if } r > 0 \\ -1, & \text{if } r \le 0 \end{cases} \tag{49}$$

Therefore the bit error probability can be found. Assume *d=-1* (without losing the generality):

$$\begin{aligned} P_b &= \Pr\{r > 0 \mid d = -1\} \\ &= \Pr\{n > \sqrt{E_b}\} \end{aligned} \tag{50}$$

This bit error probability depends on the random variable n from Eq. (48). The noise component is computed during N bits of the PN code, where a single data bit is transmitted, using:

$$n = \sum_{k=0}^{N-1} c_k \sqrt{\frac{2}{T_b}} \int_{kTc}^{(k+1)Tc} J(t)\cos(\boldsymbol{w}_0 t)dt \qquad (51)$$

where $\{c_0, c_1, \ldots c_{N-1}\}$ are the N PN bits during the transmitted data bit.

It can be defined:

$$n_k = \sqrt{\frac{2}{T_c}} \int_{kTc}^{(k+1)Tc} J(t)\cos(\boldsymbol{w}_0 t)dt \qquad (52)$$

If it is assumed that the jammer is transmitting broadband Gaussian noise of one-sided power spectral density, this leads the terms n_k to be independent Gaussian random variables with zero mean and variance $N_J/2$. Hence, the n defined on Eq. (51) can be rewritten as:

$$n = \sum_{k=0}^{N-1} c_k \sqrt{\frac{T_c}{T_b}} n_k \qquad (53)$$

Eq. (53) is a zero mean Gaussian random variable with variance $N_J/2$. For a jammer with continuous broadband noise of constant power J, the uncoded bit error probability is:

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_J}}\right) \qquad (54)$$

with:

$$Q(x) \stackrel{\Delta}{=} \int_x^{\infty} \frac{1}{\sqrt{2\boldsymbol{p}}} e^{\frac{-t^2}{2}} dt \qquad (55)$$

Eq. (55) is the Gaussian probability integral [5]. Applying Eq. (24) and Eq. (37):

$$
\begin{aligned}
\frac{E_b}{N_J} &= \frac{PG}{(J/S)} \\
&= \frac{(W_{SS}/R_b)}{(J/S)} \qquad (56) \\
\frac{E_b}{N_J} &= \frac{N}{(J/S)}
\end{aligned}
$$

This bit error probability is based on the assumption that the PN sequence $\{c\}$ is approximated as an independent binary sequence.

3.2.1.2 Pulse Jammer

Suppose that the jammer is not on all the time, but transmits pulses with higher power for a fraction of the time. It can be defined that $\boldsymbol{r}$ is the fraction of the time that the jammer is on, and $N_J/\boldsymbol{r}$ is the jammer power spectral density. It is important to note that the average power of the jammer is still J, although the power during a pulse is higher, and equal to $J/\boldsymbol{r}$.

In this analysis it is also assumed that the jammer is on a period of time that is equal or greater than T_b (the data bit time). The particular data bit will encounter the jammer on or off during the whole T_b with probabilities equal to $\boldsymbol{r}$ and $1-\boldsymbol{r}$ respectively. The cases where the jammer is on a period of time less than T_b are ignored. The output of the detector in Figure 21 is represented now by:

$$r = d\sqrt{E_b} + Zn \qquad (57)$$

where n is the noise term injected by the jammer. This is a Gaussian random variable with zero-mean and variance $N_J/2\boldsymbol{r}$. Z is a random variable independent of n with a probability of occurrence equal to:

$$
\begin{aligned}
\Pr\{Z = 1\} &= \boldsymbol{r} \\
\Pr\{Z = 0\} &= 1 - \boldsymbol{r}
\end{aligned}
\qquad (58)
$$

This random variable Z specifies if the jammer is on during a particular data bit T_b. The bit error probability is then:

$$
\begin{aligned}
P_b &= \Pr\{Zn > \sqrt{E_b}\} \\
&= \Pr\{Zn > \sqrt{E_b} \mid Z = 1\}\Pr\{Z = 1\} \\
&= \Pr\{Zn > \sqrt{E_b} \mid Z = 0\}\Pr\{Z = 0\} \\
&= \Pr\{n > \sqrt{E_b}\}\boldsymbol{r} \\
&= \boldsymbol{r}Q\left(\sqrt{\frac{2E_b}{N_J}\boldsymbol{r}}\right)
\end{aligned}
\qquad (59)
$$

When the jammer is off, the probability is reduced to:

$$
P_b = \Pr\{Zn > \sqrt{E_b} \mid Z = 0\} = 0 \qquad (60)
$$

Figure 22 shows the values of P_b versus E_b/N_J for some values of $\boldsymbol{r}$.

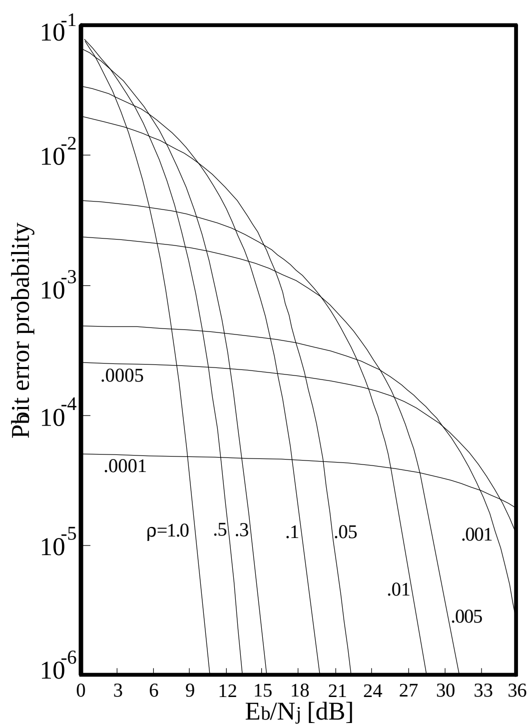

Figure 22. DS/BPSK pulse jammer [5]

The value that maximizes P_b is the derivative of Eq. (59) with respect to $\boldsymbol{r}$. This value is called $\boldsymbol{r}*$ and is given by:

$$\boldsymbol{r}* = \begin{cases} \dfrac{0.709}{E_b / N_J} & E_b / N_J > 0.709 \\ 1 & E_b / N_J \leq 0.709 \end{cases} \tag{61}$$

The maximum bit error probability is then:

$$\begin{aligned} P_b* &= \max_{0 \leq \boldsymbol{r} \leq 1} \boldsymbol{r} Q\left(\sqrt{\frac{2E_b}{N_J} \boldsymbol{r}} \right) \\ &= \boldsymbol{r} * Q\left(\sqrt{\frac{2E_b}{N_J} \boldsymbol{r}} \right) \\ &= \begin{cases} \dfrac{0.083}{E_b / N_J} & E_b / N_J > 0.709 \\ Q\left(\sqrt{\dfrac{2E_b}{N_J}} \right) & E_b / N_J \leq 0.709 \end{cases} \end{aligned} \tag{62}$$

The difference between a constant power jammer ***(r=1)*** and the worst case jammer ***(r=r*)*** can be seen in Figure 23. These results apply only to uncoded DS/BPSK systems. Note that for a bit error probability of 10^{-5} there is almost a 30 dB difference in E_b/N_J. A pulse jammer with the same average power than a constant power jammer can be more disruptive to the DS/BPSK system.

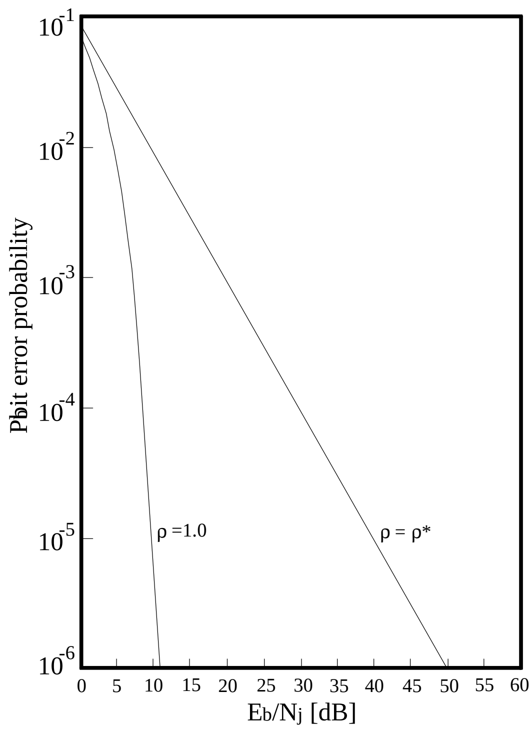

Figure 23. Constant power and worst case pulse jammer for DS/BPSK [5]

3.2.2 CODED DIRECT-SEQUENCE SPREAD BINARY PHASE-SHIFT-KEYING

To overcome the damage done with pulse jamming, several types of coding techniques can be used. The techniques provide extra gain and force the worst case jammer to be a constant power jammer. Coding techniques usually require the data rate to be decreased or the bandwidth increased because of the redundancy inherent to the coding. In spread spectrum systems, coding does not require an increase of the bandwidth or decrease of the bit rate. These properties can be seen in a simple example. If $k=2$ (constant length) the rate is $R=1/2$ bits per coded symbol of convolutional code. For each data bit of the sequence $\{d\}$, the encoder generates two coded bits. For the k^{th} transmission interval, the two data bits are:

$$a_k = (a_{k1}, a_{k2}) \tag{63}$$

where:

$$\begin{aligned} a_{k1} &= d_k \\ a_{k2} &= \begin{cases} 1 & d_k = d_{k-1} \\ -1 & d_k \neq d_{k-1} \end{cases} \end{aligned} \tag{64}$$

If T_b is the data bit time, each coded bit time is given by:

$$T_s = \frac{T_b}{2} \tag{65}$$

Defining:

$$\begin{aligned} a(t) &= \begin{cases} a_{k1} & kT_b \leq t < (k+1/2)T_b \\ a_{k2} & (k+1/2)T_b \leq t < (k+1)T_b \end{cases} \\ k &= \text{integer} \end{aligned} \tag{66}$$

In Figure 24 the uncoded data signal *d(t)*, the PN sequence *c(t)* and the coded signal *a(t)* are shown for *N=6*. In Figure 25 the multiplicated signals *d(t)c(t)* and *a(t)c(t)* are shown. With ordinary BPSK, the coded signal *a(t)* would have twice the bandwidth of the uncoded signal; but after spreading with the PN sequence, the final bandwidth is the same as the original.

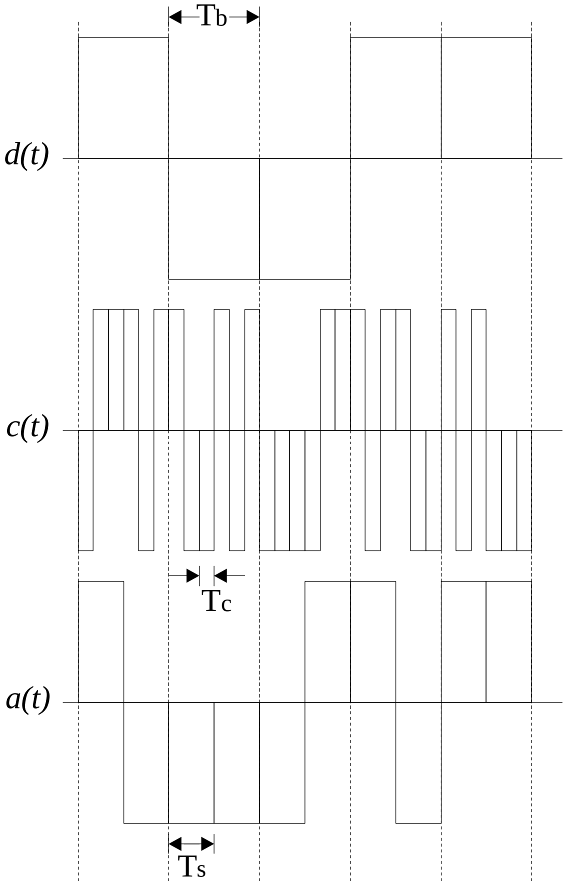

Figure 24. Coded and uncoded signals before spreading

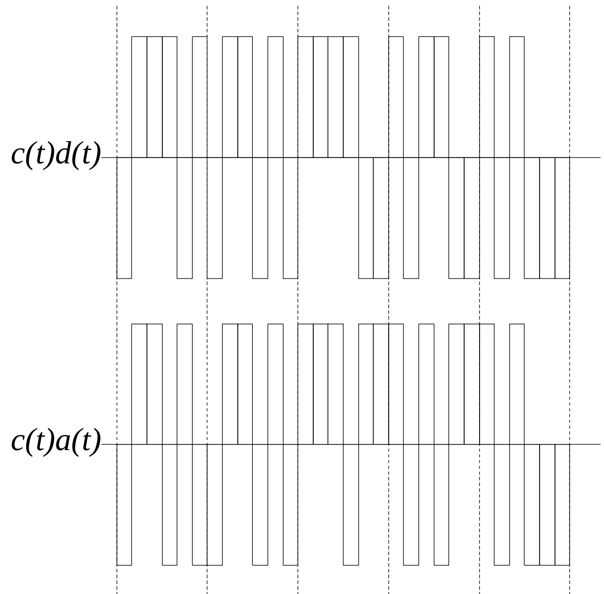

Figure 25. Coded and uncoded signals after spreading

One of the simplest coding schemes is the "repeat code." It sends *m* bits with the same value, *d*, for each data bit. The rate is then *R=1/m* bits per coded symbol. In this case, the resulting coded bits are:

$$a = (a_1, a_2, \ldots, a_m) \qquad \textbf{(67)}$$

where:

$$a_i = d \quad i = 1,2,\ldots,m \qquad \textbf{(68)}$$

Also, each coded bit a_i has a transmission time of:

$$T_s = \frac{T_b}{m} \qquad \textbf{(69)}$$

It is very important to note that if $m<N$, the bandwidth of the spread signal does not change. The complete coded DS/BPSK system is shown in Figure 26.

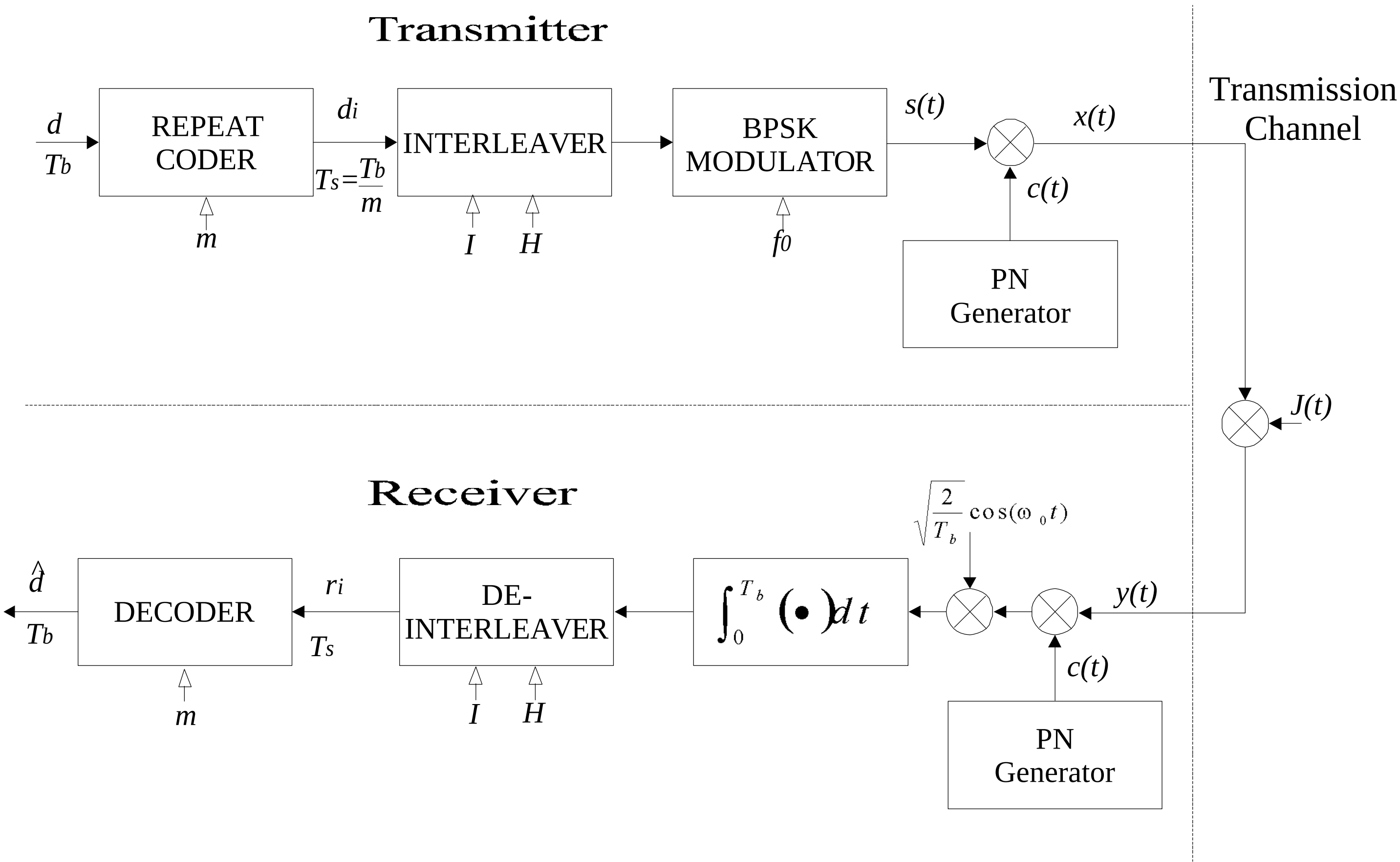

Figure 26. Repeat code DS/BPSK system

The interleaver scrambles the bits in time at the transmission, and the deinterleaver reconstructs the data sequence at the receiver. After the interleaver, the signal is BPSK modulated and then multiplied by the PN sequence. At this point the transmitted DS/BPSK signal looks like the one in Eq. (41).

$$x(t) = c(t)s(t)$$

where *s(t)* is the common BPSK (with coding). The input at the receiver is the same as that in Eq. (45):

$$y(t) = x(t) + J(t)$$

After multiplication with *c(t)* (de-spreading), it becomes Eq. (46):

$$r(t) = s(t) + c(t)J(t)$$

The output of the detector after the de-interleaver is given by:

$$r_i = a_i \sqrt{\frac{E_b}{m}} + Z_i n_i \qquad \text{(70)}$$
$$i = 1,2,\ldots,m$$

where $n_1,n_2,\ldots n_m$ are independent zero mean Gaussian random variables with variance $N_J/(2\boldsymbol{r})$. $\boldsymbol{r}$ is the fraction of time that the pulse jammer is on, and Z_i is the jammer state:

$$Z_i = \begin{cases} 1 & \text{jammer on during } a_i \text{ transmission} \\ 0 & \text{jammer off during } a_i \text{ transmission} \end{cases} \qquad \text{(71)}$$

With probability equal to:

$$\Pr\{Z_i = 1\} = \boldsymbol{r}$$
$$\Pr\{Z_i = 0\} = 1 - \boldsymbol{r} \qquad \text{(72)}$$

3.2.2.1 Interleaver and Deinterleaver

The idea of using an interleaver to scramble the data bits at transmission and a deinterleaver to unscramble the bits at reception causes the pulse jamming interference on each affected data bit to be independent from each other.

In the ideal interleaving and deinterleaving process, the variables $Z_1,Z_2,\ldots,Z_m$ become independent random variables. Assume that there is no interleaver and/or deinterleaver in the system shown in Figure 26. The output of the channel is given by:

$$r_i = d \sqrt{\frac{E_b}{m}} + Z n_i \qquad \text{(73)}$$
$$i = 1,2,\ldots,m$$

and because there is no interleaver/deinterleaver:

$$\begin{aligned} a_i &= d \\ Z_i &= Z \\ i &= 1,2,\ldots,m \end{aligned} \tag{74}$$

Also, it is assumed that the jammer was on during the whole data bit transmission T_b.

Because there is no interleaver/deinterleaver, the optimum decision rule is:

$$\begin{aligned} r &= \sum_{i=1}^{m} r_i \\ &= d\sqrt{mE_b} + Z\sum_{i=1}^{m} n_i \end{aligned} \tag{75}$$

Eq. (49) is used as a decision rule:

$$\hat{d} = \begin{cases} 1, & \text{if } r > 0 \\ -1, & \text{if } r \le 0 \end{cases}$$

If it is assumed that *d=-1*, the error probability is:

$$\begin{aligned} P_b &= \Pr\{r > 0 \mid d = -1\} \\ &= \Pr\left\{ Z\sum_{i=1}^{m} n_i > \sqrt{mE_b} \right\} \\ &= \Pr\left\{ Z\frac{1}{\sqrt{m}}\sum_{i=1}^{m} n_i > \sqrt{E_b} \right\} \\ &= \boldsymbol{r}\Pr\left\{ Z\frac{1}{\sqrt{m}}\sum_{i=1}^{m} n_i > \sqrt{E_b} \mid Z = 1 \right\} \\ &= \boldsymbol{r}Q\left(\sqrt{\frac{2E_b}{N_J}\boldsymbol{r}} \right) \end{aligned} \tag{76}$$

This bit error probability is the same for uncoded DS/BPSK; this means that without a interleaver/deinterleaver, there is no difference between uncoded systems and simple repeat code systems. Therefore, the use of a interleaver/deinterleaver is mandatory in order to achieve a good error probability measure against a pulse jammer.

Selection of the decision technique that determines the value of the coded bits $\{r\}$ requires knowledge about the state of the channel. With an ideal interleaver/deinterleaver, the output of the channel is given by Eq. (70):

$$r_i = a_i\sqrt{\frac{E_b}{m}} + Z_i n_i$$
$$i = 1,2,\ldots,m$$

where $Z_1,Z_2,\ldots,Z_m$ and $n_1,n_2,\ldots,n_m$ are considered to be independent random variables. The decoder takes $r_1,r_2,\ldots,r_m$ and finds $d_1,d_2,\ldots,d_m$ with possible values of *1* or *–1*. This analysis is valid only for the instances where the state of the channel is unknown (there is no information regarding the state of the jammer signal).

3.2.2.1.1 Hard Decision Decoder

The hard decision decoder performs a binary decision on each coded bit received:

$$\hat{d}_i = \begin{cases} 1 & r_i > 0 \\ -1 & r_i \le 0 \end{cases} \qquad i = 1,2,\ldots,m \tag{77}$$

The final decision in decoding the transmitted bit is:

$$\hat{d}_k = \begin{cases} 1 & \sum_{i=1}^{m} \hat{d}_i > 0 \\ -1 & \sum_{i=1}^{m} \hat{d}_i \le 0 \end{cases} \tag{78}$$

If m is an odd integer, the probability of error is the probability that $(m+1)/2$ or more of the m symbol decisions are in error. The probability that a single coded symbol is in error is:

$$
\begin{aligned}
\boldsymbol{e} &= \Pr\{r_i > 0 \mid d = -1\} \\
&= \Pr\left\{Z_i n_i > \sqrt{\frac{E_b}{m}}\right\} \\
&= \boldsymbol{r} Q\left(\sqrt{\frac{2E_b}{mN_J}\boldsymbol{r}}\right)
\end{aligned}
\tag{79}
$$

The overall bit error probability is given with the probability that more than half of the *m* coded symbol decisions are in error.

$$
P_b = \sum_{k=\frac{m+1}{2}}^{m} \binom{m}{k} \boldsymbol{e}^{k} (1-\boldsymbol{e})^{m-k}
\tag{80}
$$

Figure 27 shows the plot of P_b versus E_b/N_J for a repeat length $m=3$. Figure 28 and Figure 29 show the same plot for $m=5$ and $m=9$, respectively.

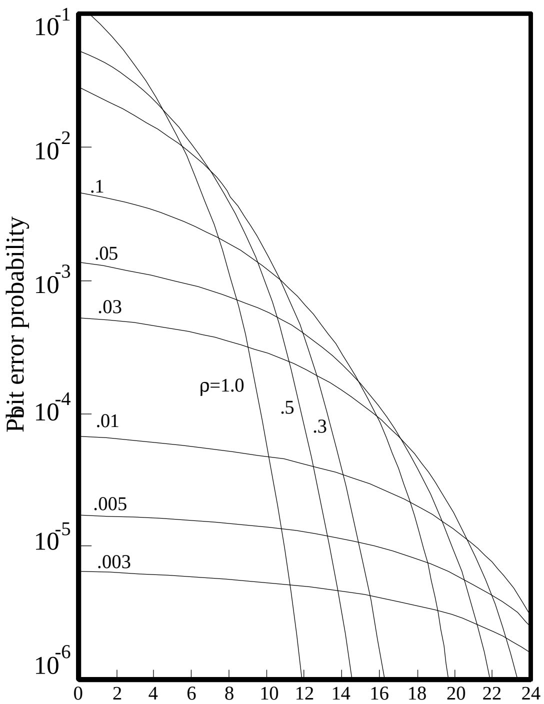

Figure 27. Repeat code *m=3* **with unknown jammer state/hard decision [5]**

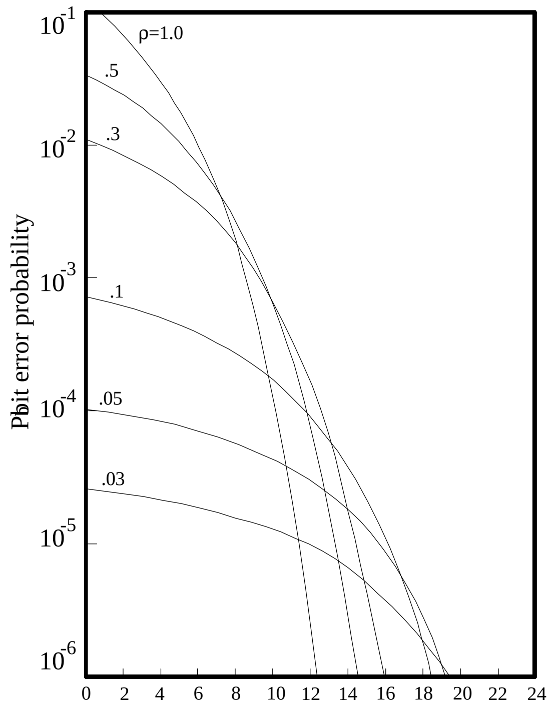

Figure 28. Repeat code *m=5* **with unknown jammer state/hard decision [5]**

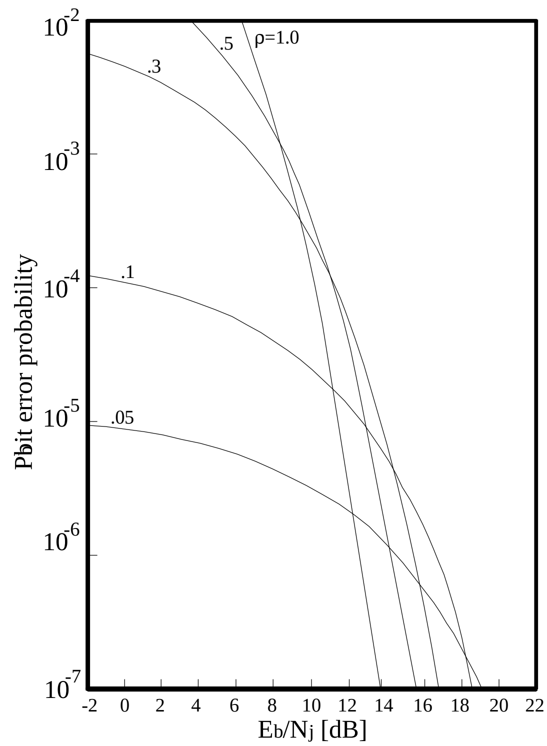

Figure 29. Repeat code $m=9$ with unknown jammer state/hard decision [5]

3.2.2.1.2 Interleaver Matrix

The interleaving techniques will improve the performance in pulse jammer environments because it makes the noise components become statistically independent variables. A block interleaver with depth $I=5$ and interleaver span $H=15$ is shown in Figure 30. The coded symbols are written to the interleaver matrix along columns, while the transmitted symbols are read out of the matrix along rows. If the coded symbol sequence is $x_1,x_2,x_3\ldots$ the sequence that comes out of the interleaver matrix is $x_1,x_{16},x_{31},x_{46},x_{61},\ldots$. At the receiver, the deinterleaver performs the inverse process, writing symbols into rows and reading them by columns. A jamming pulse of duration b symbols, with $b \leq I$ will result in these jammed symbols at the deinterleaver output to be separated at least by H symbols.

X_1	X_{16}	X_{31}	X_{46}	X_{61}
X_2	X_{17}	X_{32}	X_{47}	X_{62}
X_3	X_{18}	X_{33}	X_{48}	X_{63}
X_4	X_{19}	X_{34}	X_{49}	X_{64}
X_5	X_{20}	X_{35}	X_{50}	X_{65}
X_6	X_{21}	X_{36}	X_{51}	X_{66}
X_7	X_{22}	X_{37}	X_{52}	X_{67}
X_8	X_{23}	X_{38}	X_{53}	X_{68}
X_9	X_{24}	X_{39}	X_{54}	X_{69}
X_{10}	X_{25}	X_{40}	X_{55}	X_{70}
X_{11}	X_{26}	X_{41}	X_{56}	X_{71}
X_{12}	X_{27}	X_{42}	X_{57}	X_{72}
X_{13}	X_{28}	X_{43}	X_{58}	X_{73}
X_{14}	X_{29}	X_{44}	X_{59}	X_{74}
X_{15}	X_{30}	X_{45}	X_{60}	X_{75}

Figure 30. Interleaver matrix with $I=5$ **and** $H=15$

3.3 SYNCHRONIZATION OF SPREAD-SPECTRUM SYSTEMS

Because a pseudorandom sequence PN is used at the transmitter to modulate the signal, the first requirement at the receiver is to have a local copy of this PN sequence. The copy is needed to de-spread the incoming signal. This is done by multiplying the incoming signal by the local PN sequence copy. To accomplish a good de-spreading, the local copy has to be synchronized with the incoming signal and the PN sequence that was used in the spreading process.

The process of synchronization is usually performed in two steps: first, a coarse alignment of the PN sequence is done with a precision of less than a “chip.” This is called “PN acquisition.” After this, a fine synchronization takes care of the final alignment and corrects the small differences in the clock during transmission. This is called “PN tracking.” Theoretically, acquisition and tracking can be done in the same step with a structure of matched filters or correlators searching with high resolution the incoming signal and comparing it with the local PN sequence.

3.3.1 FAST FOURIER TRANSFORM (FFT) SCALAR FILTERS

These filters are implemented in the frequency domain, and they use the Fast Fourier Transform (forward and backward). They work over a set of N samples (usually in the frequency domain) [7]. The block diagram of an adaptive digital filter is shown in Figure 31

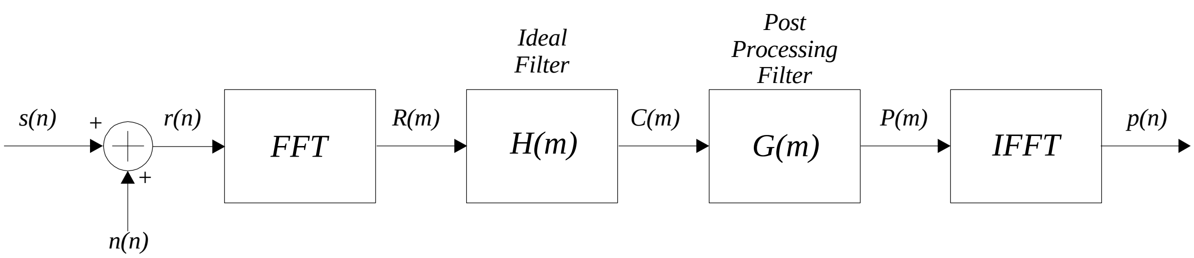

Figure 31. FFT filter assuming additive signal and noise

where:

s(n) is the input signal
n(n) is the noise (unwanted) signal
r(n) is the input to the filter
R(m) is the frequency representation of the signal (n)
H(m) is the transfer function of the filter
C(m) is the output (in frequency domain) after the filter is applied
G(m) is the transfer function of the post-processing filter
P(m) is the output after the post-processing filter
p(n) is the output signal in the time domain

The following relationships are given:

$$\begin{aligned} r(n) &= s(n) + n(n) \\ R(m) &= \mathrm{FFT}(r(n)) \\ C(m) &= H(m)R(m) \\ P(m) &= G(m)C(m) \\ p(n) &= \mathrm{FFT}^{-1}(P(m)) \end{aligned} \qquad \textbf{(81)}$$

3.3.1.1 High-resolution Detection FFT Scalar Filter

The high-resolution detection filter outputs a peak when the desired signal *s(n)* and noise *n(n)* are applied to it. The transfer function is given by:

$$H(m) = \frac{S^*(m)}{\left|S(m)\right|^2 + \left|N(m)\right|^2} \tag{82}$$

This version of high-resolution detection assumes that the noise and the signal are uncorrelated (orthogonal). The output of this filter *C(m)* must be transformed to the time domain to detect the level and the position of the peak on the output vector *c(n)*. This position can be interpreted as the exact point where the desired signal starts within the processed set of samples *N*.

3.3.1.2 Adaptive Filtering

Adaptive filters require a learning process and use adaption techniques to form the transfer function of the desired filter *H(m)*. The components of the transfer function are updated periodically with actual values taken from the signal or with estimates made using stored data. The class 1/3 high-resolution detection filter is given by [7]:

$$H(m) = \frac{S^*(m)}{\left\langle \left|R(m)\right|^2 \right\rangle} \tag{83}$$

where S*(m) is the conjugate of the spectrum of the desired signal to detect and $\left|R(m)\right|$ is the magnitude of the spectrum of the actual input of the system.

The expression $\left\langle R(m) \right\rangle$ is used to denote the “smoothing” process. This process is done to estimate the average spectrum of the signal plus noise from the actual input of the system. The smoothing used is called “inner block averaging” or “frequency domain averaging” and it is defined as:

$$\langle R(j\boldsymbol{w})\rangle = \frac{1}{2\boldsymbol{p}} R(j\boldsymbol{w}) * B(j\boldsymbol{w})$$
$$\text{or } r_b = r(t)b(t) \qquad (84)$$

The frequency averaging window *B(j**w**)* is convolved with the spectrum of the input signal. This is equivalent to a temporal weighting of the input *r(t)* by *b(t)* in the time domain. The window is usually selected to be a percentage of the input vector length.

3.4 DISCUSSION

In this chapter the spread spectrum communication technique was introduced and the Coded Direct-Sequence Spread Binary Phase-Shift-Keying (coded DS/BPSK) was explained in detail. The characteristics that make the coded DS/BPSK fit the requirements of the watermarking algorithm were also emphasized. It is important to note that literature about the communication systems is usually focused on high-speed data channels. In the watermarking algorithm, the same principles are applied, but adapted to audio frequencies and low-speed data channels.

In the proposed watermarking algorithm, the watermark bit stream is considered the data information to be spread, to form a “watermarked audio signal.” The actual “audio signal” (i.e. music) is considered as the jammer signal. A special characteristic of this system is that the jammer is stronger than the data and eventually “more important.” The recovery of the watermark bit stream requires knowledge of some of the parameters used in its generation.

4 PROPOSED SYSTEM

Different systems have been applied to watermarking of audio signals. All of them are classified as "steganographic systems" because they deal with the concept of hiding data within the signal. Boney et al. [23] proposed a system where a PN sequence was filtered using a filter that approached the masking characteristics of the human auditory system in the frequency and time domains. Some other techniques have been imported from the fields of video and still image watermarking. Cox [24] proposes a multiplatform system capable of extract a pseudorandom sequence without the use of the original unwatermarked data.

The watermarking algorithm proposed in this paper 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 is an audio waveform obtained by the use of spread spectrum techniques with a digital bit stream. The input audio signal (i.e. music) is analyzed using the psychoacoustic auditory model to find the final masking threshold. This information is then used to shape and embed the watermark into the audio signal, retaining its perceptual quality. All the audio waveforms are assumed to be in the same sound digital format (i.e., PCM).

4.1 WATERMARK GENERATION AND EMBEDDING

4.1.1 WATERMARK GENERATION

The objective of the watermark generation is to generate a watermark audio signal *x(t)* that contains the watermark bit stream data. This watermark signal can be transmitted and then processed for data recovery. The technique used to generate the watermark signal *x(t)* is the “coded DS/BPSK spread spectrum.” The process is condensed in Figure 32.

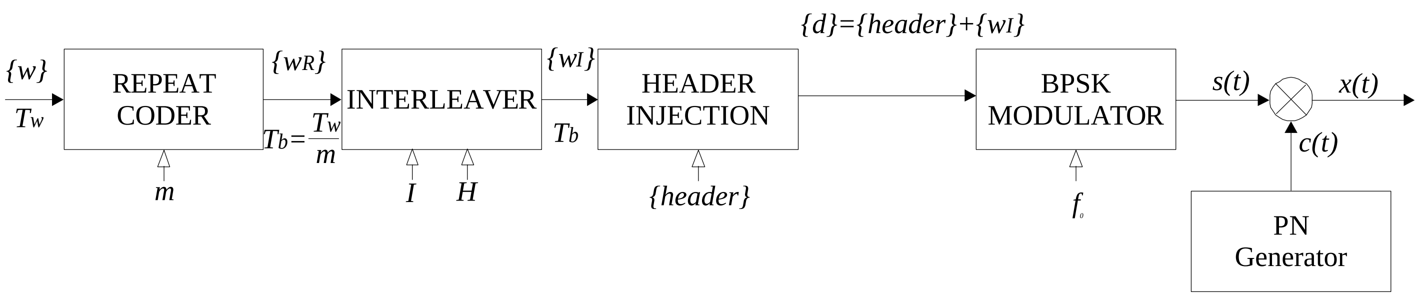

Figure 32. Watermark generation system

where:

$\{w\}$ is the original digital bit stream(watermark)
m is the repetition code factor
$\{w_R\}$ is the watermark after the coding process (repeat code)
I,H = width and length of the interleaver matrix
$\{w_I\}$ is the watermark after the interleaver process
$\{header\}$= is the header sequence
$\{d\}=\{header\}+\{w_I\}$= sequence to be spread and transmitted
f_0 = frequency used by the BPSK modulator

The process can be explained with a simple example: Let $\{w\}$ be the watermark bit stream. All the bit streams used are bipolar (value 1 or –1). Defining $\{w\}$with a length of 16 bits as the sequence:

$$\{w\} = \{\ 1 \quad 1 \quad -1 \quad 1 \quad -1 \quad -1 \quad 1 \quad -1 \mid 1 \quad 1 \quad -1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1\}$$

Using Eq. (67) to generate the repeat code, and choosing *m=3*, the $\{w_R\}$ sequence is:

$$\{w_R\} = \{ \begin{array}{ccc:ccc:ccc:ccc} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{array} \}$$

The next step is to perform interleaving. To do this, the values of the interleaving matrix are chosen; in this case, *I=5, H=10,* (see Figure 30). The resulting matrix is shown in Figure 33:

1	*1*	*1*	*-1*	*1*
1	*1*	*-1*	*-1*	*1*
1	*-1*	*-1*	*-1*	*-1*
1	*-1*	*-1*	*1*	*-1*
1	*-1*	*1*	*1*	*-1*
1	*-1*	*1*	*1*	*-1*
-1	*-1*	*1*	*1*	*-1*
-1	*-1*	*1*	*1*	*-1*
-1	*1*	*1*	*1*	*1*
1	*1*	*1*	*1*	*1*

Figure 33. Interleaver matrix

The last two spaces are padded with 1's. Using the interleaving matrix, the output sequence $\{w_I\}$ is:

$$\{w_I\} = \{ \begin{array}{cccccccccccc} 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & \} & & & & & & & & & \end{array}$$

The selected header is a sequence usually composed by 1's.

$$\{header\} = \{1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1\}$$

The final data sequence $\{d\}$ is obtained concatenating the $\{header\}$ and the $\{w_I\}$:

$$\{d\} = \{header\} + \{w_I\}$$

$$\begin{array}{llllllllllllll} \{d\} = \{ & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \\ & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & \\ & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & \\ & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & \\ & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \} \end{array}$$

The PN sequence $\{c\}$ can be generated by any means. Usually this is done using a pseudorandom number generator. In this case, the PN sequence is assumed to be long enough to spread a complete bit stream (header and data) without repeating any portion of it. The important factor is that the transmitter and the receiver must have a copy of the whole PN sequence $\{c\}$. This sequence is ideally uncorrelated with the $\{d\}$ sequence, and has the form:

$$\{c\} = \{ \ 1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1 \quad 1 \quad \ldots\}$$

4.1.1.1 Spread Spectrum Parameter Selection

Audio signals are usually considered to be baseband signals [21]. The described spread spectrum technique can be applied to passband systems (with $f_0 > 0$) or baseband systems ($f_0 = 0$) without losing generality.

The selection of all the parameters is based on the considerations of how the overall watermarked audio signal will be transmitted or stored. The frequency response of those systems determines which frequencies are likely to be present at the receiver.

Let a baseband bandlimited signal, with no modulation ($f_0=0$) have the magnitude spectrum shown in Figure 34:

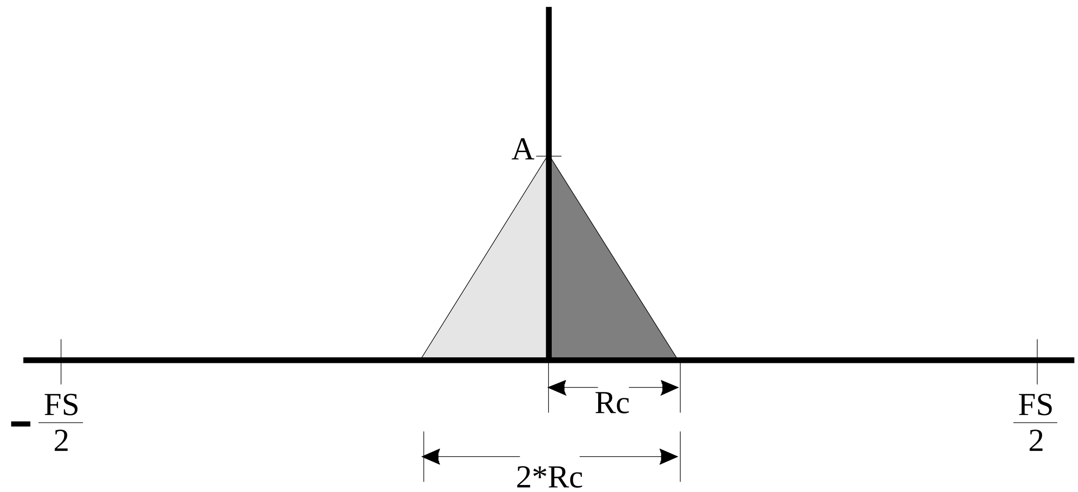

Figure 34. Baseband System Parameters

With amplitude modulation ($f_0>0$), the spectrum will have the form shown in Figure 35:

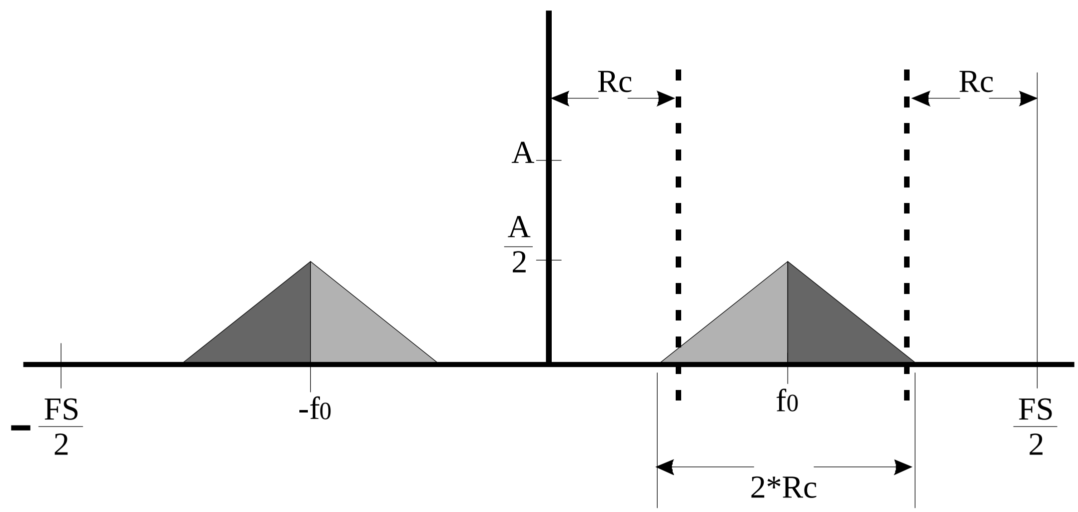

Figure 35. Passband System Parameters (anti-aliasing)

where FS is the sampling frequency of the system. To avoid aliasing because of the use of modulation, the modulation frequency should be:

$$Rc \le f_0 \le \frac{FS}{2} - Rc \qquad \textbf{(85)}$$

If a system possesses a lower frequency limit *LF* and/or an upper frequency limit *HF*, the modulation frequency f_0 have to be selected in a way that the sidebands fall between the lower an upper limits, as shown in Figure 36.

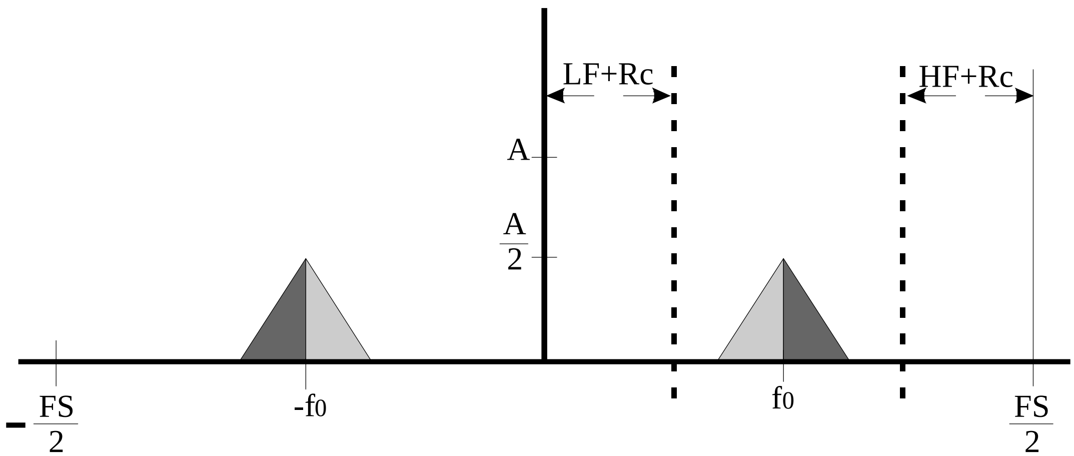

Figure 36. Passband System with Frequency Limits *LF* and *HF*

If a sideband falls outside of these limits, aliasing or data loss could result. Taking into account, the selection of parameters should be done using:

$$\begin{aligned} & LF + Rc \le f_0 \le HF - Rc \\ & LF \ge 0, HF \le \frac{FS}{2} \end{aligned} \qquad \textbf{(86)}$$

The parameters selected must satisfy Eq. (85) and Eq. (86), along with the following relationships:

R_d = is the data bits per second
m = is the repetition code factor
N = is the spreading factor, Eq. *(37)*
$R_b = R_d*m$ is the coded bits per second
$T_b = 1/R_b$ is the time of each coded bit
$R_c=N*R_b$ is thePN sequence bits per second
$T_c=Tb/N$ is the time of each PN bit or “chip”

Assuming a frequency response similar to FM Radio [22] with *LF = 50 Hz* and *HF = 15000 Hz*, for the actual example, a set of spread spectrum parameters that satisfy all the requirements is:

$N = 3$
$m = 3$
$R_d = 100$ bits/sec
$R_b = 300$ bits/sec
$R_c = 900$ bits/sec
$f_0 = 3500$ Hz

Note that N and m are selected with small values for this example. The modulation is done using Eq. (42):

$$s(t) = d(t)\sqrt{2S}\cos(\boldsymbol{w}_0 t)$$

The spreading is done using Eq. (41):

$$x(t) = c(t)s(t)$$

The output of the system is the watermarked audio waveform *x(t)* shown in Figure 37:

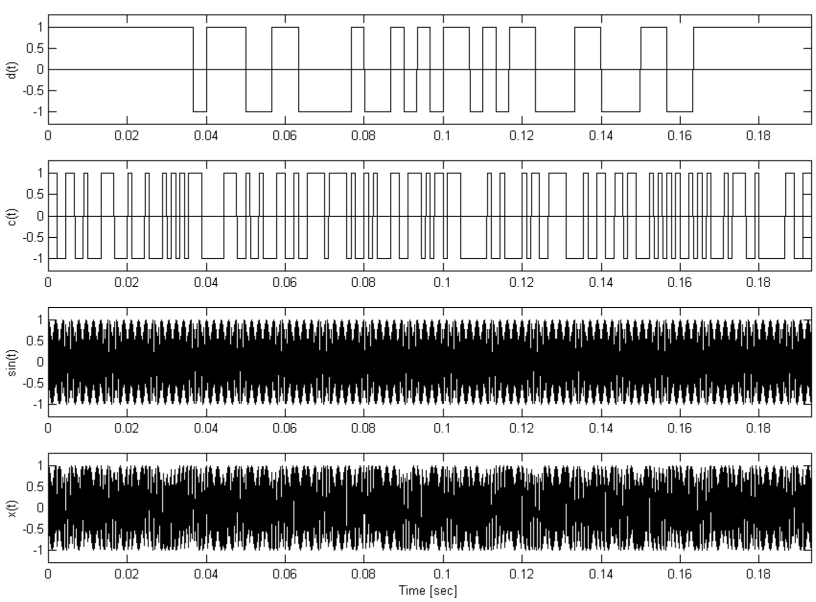

Figure 37 Time domain signals: data bit stream, $d(t)$**; PN sequence,** $c(t)$**; BPSK modulator,** $sin(t)$**; and watermark audio signal,** $x(t)$

4.1.2 FRAME SEGMENTATION

To overcome the potential problem of the audio signal or the watermark signal being too long to be processed using a single FFT, the signal is segmented in short overlapping segments, processed and added back together [8]. Another consideration for the watermark algorithm is that the audio signal has to be longer than the watermark signal. Therefore, the watermark can be repeated several times during the duration of the audio signal. This redundancy is one of the important features in the watermarking algorithm. Figure 38 shows audio and watermark signals that will be segmented. The watermark is repeated several times.

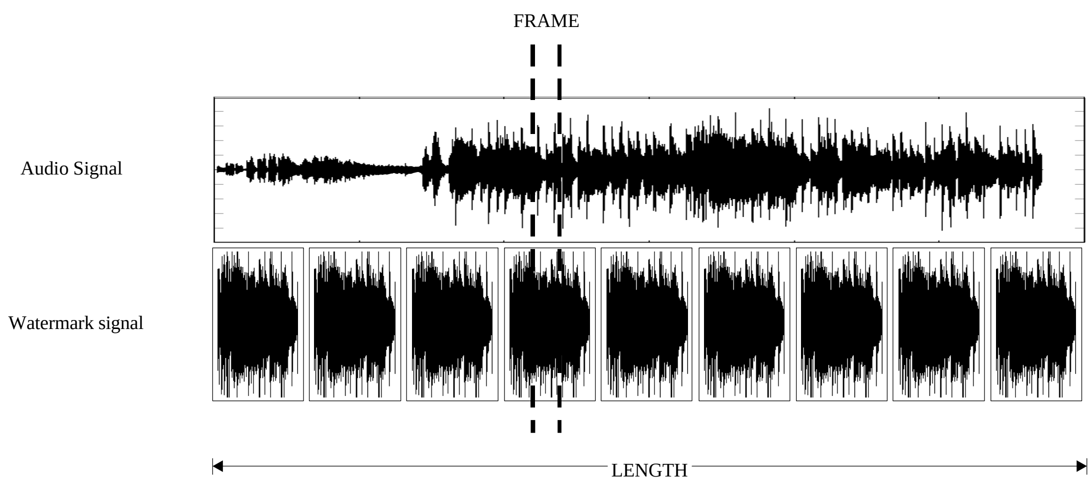

Figure 38. Frame segmentation and watermark redundancy

If the total length of the audio signal is *LENGTH* samples, the desired length of the analysis frame is *BLOCK* samples, and the overlap between consecutive frames is *OVERLAP* samples, the total number of *FRAMES* is given by:

$$FRAMES = \frac{LENGTH - OVERLAP}{BLOCK - OVERLAP} \qquad (87)$$

In Figure 38 two equal length frames were selected to be processed. One from the audio signal and the other from the respective point in the watermark signal. The last frame is zero-padded if it is shorter than *BLOCK* samples. These padded samples are discarded in post processing. From this point on, all processes described are applied to the audio or watermark signal frames, not the entire signal.

4.1.3 FREQUENCY REPRESENTATION

The Short Time Fourier Transform (STFT) discussed in section 2.1 is used to acquire a frequency representation of the actual frames. Before doing the STFT, a Hamming window is applied to both signals [7], [8]. This improves the representation of the signal in the frequency domain reducing the leakage. If *s(t)* is the actual audio signal frame and *x(t)* the actual watermark signal frame, then the windowing is done using:

$$sw(t) = s(t)w(t) \tag{88}$$

$$xw(t) = x(t)w(t) \tag{89}$$

The Hamming window is defined as:

$$\begin{aligned} w(n) &= 0.54 + 0.46\cos\left(\frac{2n\boldsymbol{p}}{BLOCK}\right) \\ n &= 1,2\ldots BLOCK \\ w(t) &= w(nT) \\ T &= \text{sampling period} \end{aligned} \tag{90}$$

For the frequency representation of the audio frame, Eq. (1) is used:

$$Sw(j\boldsymbol{w}) = \mathrm{FT}\{sw(t)\} \tag{91}$$

and the watermark frame:

$$Xw(j\boldsymbol{w}) = \mathrm{FT}\{xw(t)\} \tag{92}$$

The power spectra is found using Eq. (5):

$$Sp(j\boldsymbol{w}) = \left|Sw(j\boldsymbol{w})\right|^2 \tag{93}$$

The indices of the actual frequency representations have to be mapped to the Bark scale. Once this index mapping is done, the representation in the critical band scale is formed by mapping the components to the respective position on the critical band axis. The relationship between each component index, i, and the corresponding frequency, $f_{i,}$ that it represents is given by:

$$\begin{aligned} f_i &= \frac{(i-1)*FS}{BLOCK} \\ i &= 1,2\ldots\frac{BLOCK}{2} \\ FS &= \text{Sampling Frequency} \end{aligned} \tag{94}$$

The relationship between each frequency f_i and the bark scale or critical band scale z_I is found using Eq. (3):

$$z_i = 13\tan^{-1}\left(\frac{0.76*f_i}{1000}\right) + 3.5\tan^{-1}\left(\left(\frac{f_i}{7500}\right)^2\right)$$

This relationship between each component index i and the frequency f_i or critical band z_i that it represents can be calculated at the beginning of the algorithm and stored in a table.

The energy per critical band is calculated using Eq. (6):

$$Spz(z) = \sum_{\boldsymbol{w}=LBZ}^{HBZ} Sp(j\boldsymbol{w})$$

with:

$$\begin{aligned} \text{z} &= 1,2,\ldots,\text{Zt} \\ \text{LBZ} &= \text{Lower frequency in critical band z} \\ \text{HBZ} &= \text{Higher frequency in critical band z} \end{aligned}$$

Figure 39 (a) shows the original audio frame *s(t)* in the time domain and the shape of the Hamming window *w(t);* (b) shows the *sw(t)* frame after the windowing process; (c) shows the magnitude of $Sw(jw)$, and (d) shows the power spectrum $Sp(jw)$ and the energy per critical band $Spz(z)$.

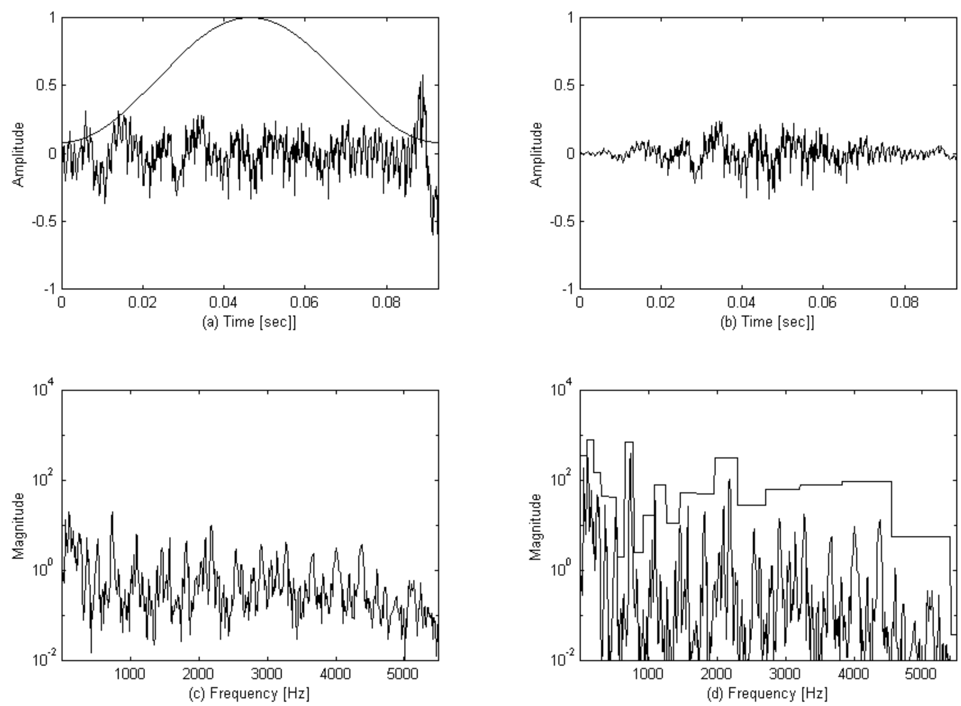

Figure 39. (a) Audio signal *s(t)* **and window signal** *w(t)*, **(b) windowed signal** *sw(t)*, **(c) magnitude of frequency representation** *Sw(jw)*, **and (d) power spectrum** *Sp(jw)* **and energy per critical band** *Spz(z)*

4.1.4 BASILAR MEMBRANE SPREADING FUNCTION

The basilar membrane spreading function determines how much of the energy of each critical band is contributed to the neighboring bands. The spreading function *B(z)* is calculated using Eq. (7):

$$B_k = 15.91 + 7.5(k + 0.474) - 17.5\sqrt{1 + (k + 0.474)^2}$$
$$k = \ldots -2, -1, 0, 1, 2 \ldots$$

The spreading across bands is computed by the convolution of the spreading function *B(z)* and the energy per critical band $Spz(z)$,using Eq. (8):

$$Sm(z) = Spz(z) * B(z)$$

Figure 40 (a) shows the energy per critical band $Spz(z)$, (b) shows the spreading function *B(z)* for 9 points, and (c) shows the spread energy per critical band *Sm(z)*.

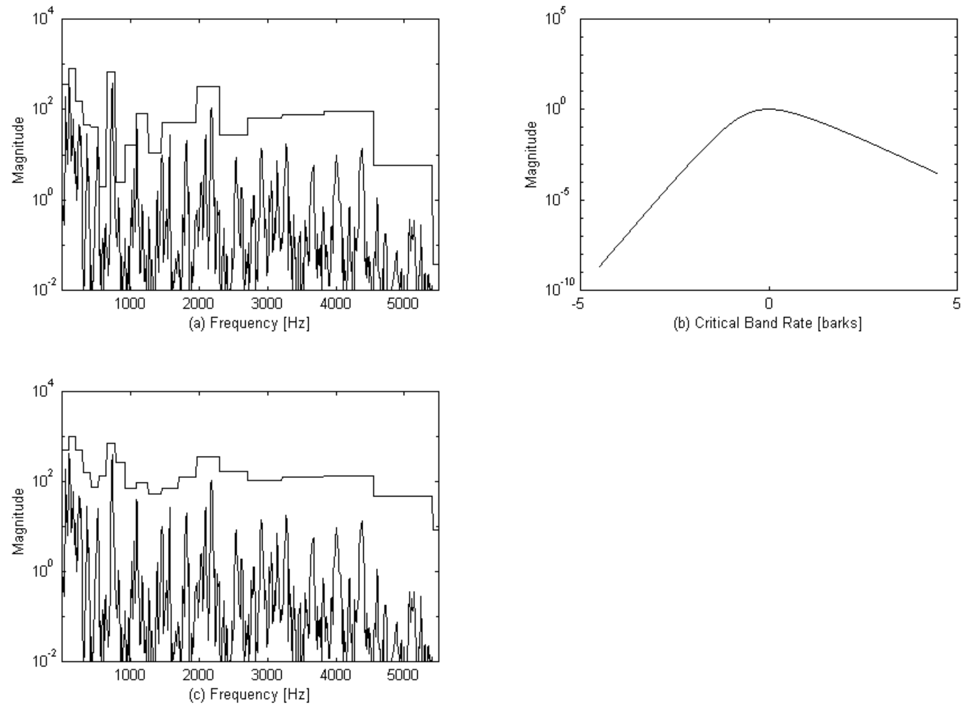

Figure 40. (a) Energy per critical band *Spz(z)*, **(b) spreading function** *B(z)*, **and (c) Spread energy per critical band** *Sm(z)*

4.1.5 MASKING THRESHOLD ESTIMATE

The Spectral Flatness Measure (SFM) of the actual audio frame $Sw(j\boldsymbol{w})$ is taken using Eq. (9):

$$SFM_{dB} = 10\log 10 \left\{ \frac{\prod_{z=1}^{Zt} Spz(z)}{\frac{1}{Zt}\sum_{z=1}^{Zt} Spz(z)} \right\}^{\frac{1}{Zt}}$$

with Zt = total number of critical bands in each frame

The energy per critical band $Spz(z)$ is used rather than spread energy per critical band *Sm(z)* to avoid false results due to smoothing of the signal. The tonality factor α is then calculated using Eq. (10):

$$\alpha = \min\left(\frac{SFM_{dB}}{SFM_{dB\max}}, 1 \right)$$

with $SFM_{dB\max} = -60dB$.

The masking energy offset *O(z)* is then calculated using Eq. (11):

$$O(z) = \alpha(14.5 + z) + (1 - \alpha)5.5$$

The raw masking threshold, *Traw(z)*, is calculated with Eq. (12):

$$Traw(z) = 10^{\left(\log_{10}(Sm(z)) - \frac{O(z)}{10}\right)}$$

The raw masking threshold is normalized using Eq. (13):

$$Tnorm(z) = \frac{Traw(z)}{P_z}$$

where:

$$\mathrm{P_z} = \text{number of points in each band z}$$
$$\mathrm{z} = 1,2,\ldots.\mathrm{Zt}$$

To calculate the final masking threshold *T* it is necessary to first calculate the hearing threshold (or threshold in quiet) *TH.* It is defined as a sinusoidal tone of 4000 Hz with one bit of dynamic range. Using Eq. (14):

$$TH = \max\left(\left|Pp(j\omega)\right|\right)$$

where:

$$Pp(j\omega) = \text{power spectrum of the probe signal } p(t)$$
$$p(t) = \sin(2\pi\, 4000t)$$

Then the final masking threshold T is calculated using Eq.(15):

$$T(z) = \max(Tnorm(z), TH)$$

with: z=1,2,….Zt

Figure 41 (a) shows the raw masking threshold *Traw(z)* and (b) shows the normalized threshold *Tnorm(z).*

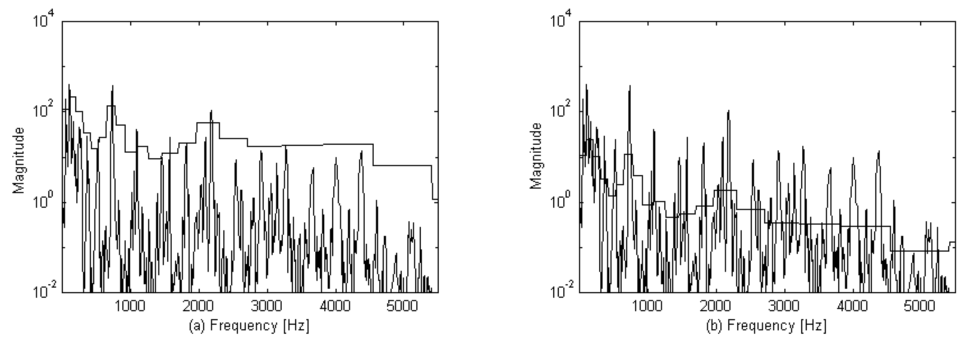

Figure 41. (a) raw masking threshold *Traw(z)*, **and (b) normalized masking threshold** *Tnorm(z)*

4.1.6 WATERMARK SPECTRAL SHAPING

The final masking threshold T is used to determine which components of the audio signal $Sw(j\boldsymbol{w})$ can be removed without affecting the perceptual quality of the signal. The power spectrum $Sp(j\boldsymbol{w})$ is compared against the final masking threshold T. The components that fall below it are removed in $Sw(j\boldsymbol{w})$. The new frame with only the components above the threshold is called $Swnew(j\boldsymbol{w})$. Eq.

(16) is used:

$$Swnew_i(j\boldsymbol{w}) = \begin{cases} Sw_i(j\boldsymbol{w}) & Sp_i(j\boldsymbol{w}) \geq T(z) \\ 0 & Sp_i(j\boldsymbol{w}) < T(z) \end{cases}$$
$$i = 1,2\ldots \text{number of components}$$
$$\text{z},\boldsymbol{w} \text{ according to component } i$$

Then the unneeded components of the watermark signal $Xw(j\boldsymbol{w})$ are removed. These components correspond to the non-removed components in $Sw(j\boldsymbol{w})$. Eq.

(17) is used:

$$Xwnew_i(j\boldsymbol{w}) = \begin{cases} 0 & Sp_i(j\boldsymbol{w}) \geq T(z) \\ Xw_i(j\boldsymbol{w}) & Sp_i(j\boldsymbol{w}) < T(z) \end{cases}$$
$$i = 1,2\ldots \text{number of components}$$
$$\text{z},\boldsymbol{w} \text{ according to component } i$$

The factors that will shape the new watermark, $Xwnew(j\boldsymbol{w})$, are found using Eq.

(20):

$$F_z = A\frac{\sqrt{T(j\boldsymbol{w})}}{\max\left(\left|Xwnew(j\boldsymbol{w})\right|\right)}$$
$$z = 1,2\ldots Zt$$
$$\boldsymbol{w} = LBZ \text{ to } HBZ \text{ for each band } z$$

The square root of the final threshold is divided by the maximum magnitude component found in the energy of the new watermark in each critical band. Each one of these factors is scaled using the gain *A*, that varies from 0 to 1, and controls the overall magnitude of the watermark signal in relation with the audio signal.

Each one of the components in each critical band *k* is scaled by the corresponding factor using Eq. (21):

$$Xfinal(j\boldsymbol{w}) = Xwnew(j\boldsymbol{w})F_z$$
$$z = 1,2\ldots Zt$$
$$\boldsymbol{w} = LBZ \text{ to } HBZ \text{ for each band } z$$

Figure 42 shows the final masking threshold and the watermark signal before shaping (a) and after shaping (b). Note that the watermark falls below the threshold of masking. The factor *A* gives control of "how much gain" will have the watermark related with the masking threshold (*A* is a value from 0 to 1).

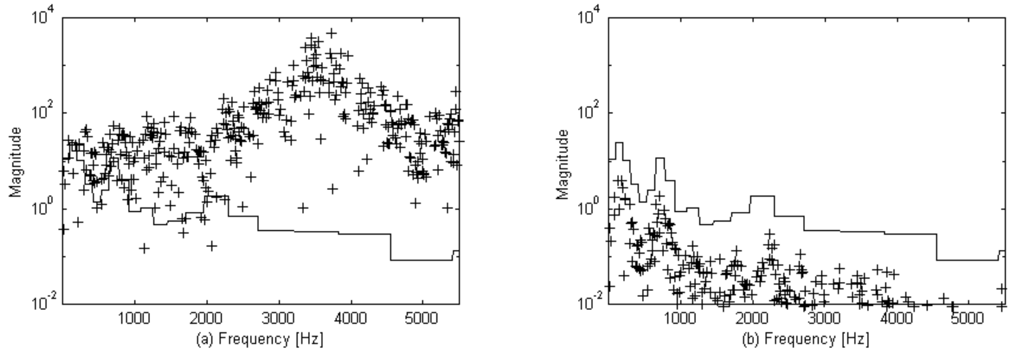

Figure 42. (a) *Xwnew(z)* **before shaping, (b) after shaping with** $A = 0.4$

4.1.7 AUDIO AND WATERMARK SIGNAL COMBINATION

The final output $OUT(j\boldsymbol{w})$ is the sum of the new audio, $Swnew(j\boldsymbol{w})$, and the final watermark $Xfinal(j\boldsymbol{w})$. This is given by the Eq. (22):

$$OUT(j\boldsymbol{w}) = Swnew(j\boldsymbol{w}) + Xfinal(j\boldsymbol{w})$$

Figure 43 shows the final masking threshold *Tfinal(z)*, and the power spectrum of (a) *Swnew(j**w**)*, (b) *Xfinal(j**w**)*, and (c) *OUT(j**w**)*.

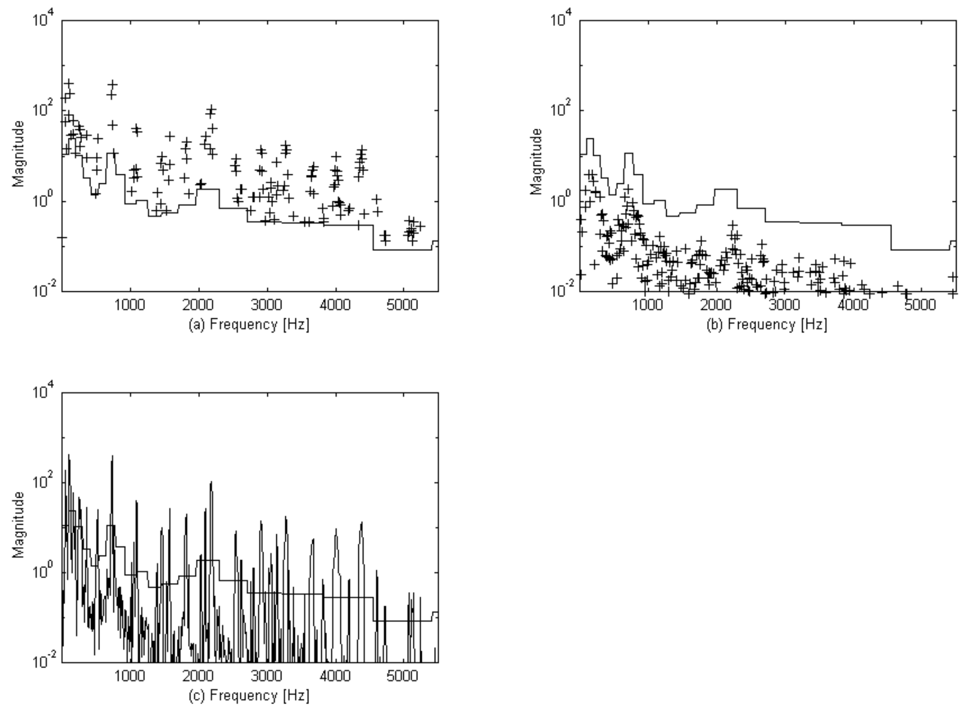

Figure 43. Final masking threshold *Tfinal(z)*, **and the power spectrum of (a)** *Swnew(j**w**)*, **(b)** *Xfinal(j**w**)*, **and (c)** *OUT(j**w**)*.

4.1.8 TRANSFORMATION TO THE TIME DOMAIN

The Inverse Fourier Transform is used to convert the frequency domain information back to the time domain. Using Eq. (2):

$$out(t) = \mathrm{IFT}\{OUT(j\boldsymbol{w})\}$$

This output frame *out(t)* is added to the correspondent point at the total time domain output *output(t)*. The next frames of audio and watermark signals are taken, and the process is repeated.

4.2 DATA RECOVERY

The watermarked audio signal is intended to be transmitted through a diverse number of channels. In some cases, the channel will introduce noise, convert several times from digital to analog and analog to digital, or even use a psychoacoustic auditory model to process the audio signal. The watermark bit stream should survive the transmission and be recoverable.

A very important characteristic is that the developed system does not require access to the original audio signal (before watermark) to extract the watermark at the receiving. The process of recovery uses the psychoacoustic auditory model, but in this case the goal is to remove all the audio components that have less probability of belonging to the watermark signal. This means that the masking threshold is calculated and the components above it are removed. The final signal is the "residual." This residual is then analyzed to find the possible points where the watermark is present. If some criterion is applied, the majority of the false points detected can be eliminated (i.e. rejecting points too close to fit a watermark). Synchronization and recovery of the watermark bit stream are then performed.

4.2.1 MASKING THRESHOLD AND RESIDUAL SIGNAL

The watermarked audio signal after the transmission is symbolized as $s_2(t)$. The process described in sections 4.1.2 to 4.1.5 is used to calculate the frames *sw2(t)*, frequency representation $Sw_2(j\boldsymbol{w})$,and masking threshold T_2, respectively.

The residual signal $R(j\boldsymbol{w})$ is defined as the signal composed of the components below the masking threshold. Eq. (16) can be changed to:

$$R_i(j\boldsymbol{w}) = \begin{cases} Sw_{2i}(j\boldsymbol{w}) & Sp_{2i}(j\boldsymbol{w}) \le T_2(z) \\ 0 & Sp_{2i}(j\boldsymbol{w}) > T_2(z) \end{cases}$$
$$i = 1,2\ldots \text{number of components} \qquad (95)$$
$$\text{z}, \boldsymbol{w} \text{ according to component } i$$

4.2.2 RESIDUAL EQUALIZATION

The spectrum of the residual $R(j\boldsymbol{w})$ is then shaped to be flat. Eq. (20) can be modified to shape all the maximum components of each band to be at equal levels. The factors are found using:

$$F_z = \frac{1}{\max(|R(j\boldsymbol{w})|)}$$
$$z = 1,2\ldots Zt \qquad (96)$$
$$\boldsymbol{w} = LBZ \text{ to } HBZ \text{ for each band } z$$

Each one of the components in each critical band z is scaled by the corresponding factor F_z using Eq. (21):

$$Rfinal(j\boldsymbol{w}) = R(j\boldsymbol{w})F_z$$
$$z = 1,2\ldots Zt$$
$$\boldsymbol{w} = LBZ \text{ to } HBZ \text{ for each band } z$$

4.2.3 TIME DOMAIN RESIDUAL

The residual is taken back to the time domain using the Inverse Short Time Fourier Transform IFT. Using Eq. (2):

$$r(t) = \text{IFT}\{Rfinal(j\boldsymbol{w})\}$$

The time domain *r(t)* frame is added to the total time domain residual signal *residual(t)* at the point specified by the frame segmentation step. The next frame is then processed.

4.2.4 SYNCHRONIZATION WITH WATERMARK HEADER

To be able to synchronize and to have a good de-spreading of the watermark signal, it is necessary to have knowledge of the parameters used at the generation of the watermark signal, such as f_0, T_b, *m, H, I, N,* $\{header\}$,*{c}, etc.*

4.2.4.1 header(t) Signal Generation

The first step is to generate a *header(t)* waveform signal using the process of section 4.1.1, except that only the $\{header\}$ sequence is used as the input sequence. This audio signal will be used to locate the exact positions of the watermark signals in the *residual(t)* signal. Frame segmentation as explained in section 4.1.2 is also required in order to analyze the whole *residual(t)* signal. The parameters for the frame segmentation are chosen to have up to two *header(t)* signals in each frame. Therefore, BLOCK is equal to twice the number of samples in *header(t)*, and OVERLAP is equal to one half the number of samples in header(t). The resulting frame taken from *residual(t)* with BLOCK length is called *r(t)*.

4.2.4.2 header(t) Position Detection

Eq. (83) describes an adaptive high-resolution filter that can be used to detect the presence of *header(t)* in the *r(t)* frame and therefore, all the occurrences of *header(t)* in the *residual(t)* audio signal.

$$H(j\boldsymbol{w}) = \frac{HEADER^*(j\boldsymbol{w})}{\left\langle \left| R(j\boldsymbol{w}) \right|^2 \right\rangle}$$

Where:

$$R(j\boldsymbol{w}) = \text{FFT}(r(t))$$
$$HEADER(j\boldsymbol{w}) = \text{FFT}(header(t))$$

The denominator of the filter is the smoothed version of $|R(j\boldsymbol{w})|^2$. Smoothing is done using Eq. (84), where *w(t)* is a Hanning window of width 10%. The output of the filter applied to $R(j\boldsymbol{w})$ is:

$$DET(j\boldsymbol{w}) = R(j\boldsymbol{w}) \frac{HEADER^*(j\boldsymbol{w})}{\left\langle |R(j\boldsymbol{w})|^2 \right\rangle}$$

This result is transformed to the time domain to be analyzed.

$$det(t) = \text{real}(\text{IFFT}(DET(j\boldsymbol{w})))$$

A typical output of the filter, *det(t)*, is shown in Figure 44. The peak shows the position in samples where the *header(t)* signal starts in the frame *r(t)*.

This detection is done for all the frames in the *residual(t)* signal, and all the positions of the peaks are stored for further analysis. A proposed criterion of analysis is to determine the minimum distance between peaks to decide which ones have more probability to represent the start of a watermark signal.

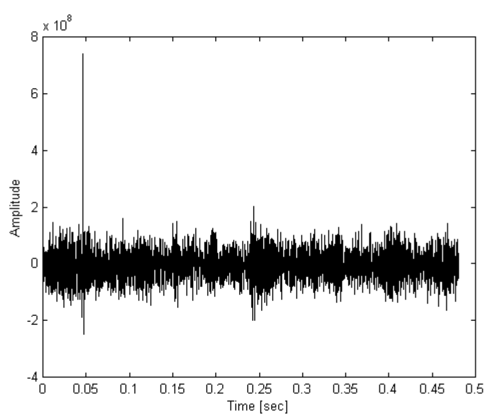

Figure 44. Detection peak in *det(t)*

4.2.5 WATERMARK DE-SPREADING

For each peak position found in the *residual(t)*, a selected frame *y(t)* with the same length as the watermark signal is processed. This process is shown in Figure 45:

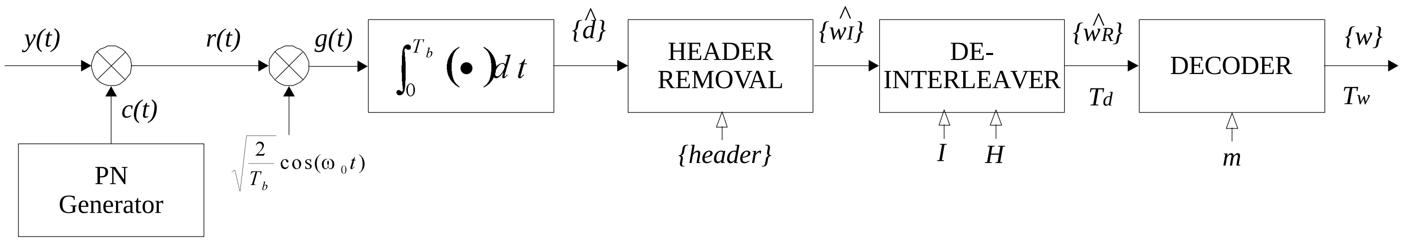

Figure 45. Watermark recovery system

Using Eq. (46):

$$r(t) = c(t)y(t)$$

Demodulation is performed using Eq. (42):

$$g(t) = r(t)\sqrt{\frac{2}{T_b}}\cos(2\pi f_0 t)$$

To estimate the bit stream:

$$r_i = \int_{(i-1)T_s}^{iT_s} g(t)dt \qquad i = 1,2\ldots\text{total bits in bit stream} \qquad \textbf{(97)}$$

The decision rule, to form a recovered bit stream $\{\hat{d}\}$, is given by Eq. (49),

$$\hat{d}_i = \begin{cases} 1, & \text{if } r_i > 0 \\ -1, & \text{if } r_i \le 0 \end{cases} \qquad i = 1,2\ldots\text{total bits in bit stream}$$

After this decision, the $\{header\}$ sequence is discarded from the $\{\hat{d}\}$bit stream. This produces the bit stream, $\{\hat{w}_I\}$.

4.2.6 WATERMARK DE-INTERLEAVING AND DECODING

The de-interleaving process is done using the same matrix used in the watermark generation in section 4.1.1 and shown in Figure 30. The bits are written into rows and read by columns to accomplish the de-interleaving process. The de-interleaved sequence is called $\{\hat{w}_R\}$. The decoding of the repeat code of value *m* is done using Eq. (78):

$$\hat{w}_k = \begin{cases} 1 & \sum_{i=1}^{m} \hat{w}_{Ri} > 0 \\ -1 & \sum_{i=1}^{m} \hat{w}_{Ri} \leq 0 \end{cases}$$
$$k = 1,2\ldots \text{total bits in data sequence}$$

The final recovered sequence $\{\hat{w}\}$ is the recovered watermark.

5 SYSTEM PERFORMANCE

5.1 SURVIVAL OVER DIFFERENT CHANNELS

A watermarking system was implemented using the software *Matlab* of MathWorks, Inc. The system was composed of two modules: watermark generation and embedding, and watermark recovery. The watermark was first generated and embedded in an audio signal. The watermarked signal was then tested for recovery of the watermark after transmission by different channels, such as sub-band encoding, digital to analog – analog to digital conversions and radio transmission. The parameters used for the generation of the watermarks were:

Left channel:

Fw = 100 bits/sec
N = 15
m = 5
I = 9
H = 27
f0 = 3500 Hz
{header} = {1 1 1 1 …} (24 one's)
watermark = '123456' (The ASCII representation of the six character string.)
A = -2 dB, -4 dB, -6 dB and -8 dB respectively.

Right channel:

Fw = 100 bits/sec
N = 20
m = 5
I = 9
H = 27
f0 = 3500 Hz
{header} = {1 1 1 1 …} (24 one's)
watermark = '123456' (The ASCII representation of the six character string)
A = -2 dB, -4 dB, -6 dB and -8 dB respectively.

The music used was a 26 second excerpt of the song “In the Midnight Hour” (W. Pickett & S. Cropper) performed by *The Commitments*. A sampling frequency of 44.1 KHz was used. Each of the watermarked audio signals was labeled to reflect the level of the watermark below the masking threshold (the A value), i.e. W2, W4, W6 and W8. With these parameters, a total of 35 watermarks were embedded during the duration of each signal. The four watermarked music signals and the original signal were recorded digitally on a compact disc using Sonic Foundry CD Architect music authoring software and a CD recorder unit RICOH CD-RW MP 6200S. The computer was also equipped with a full duplex digital sound card MALIBU made by Turtle Beach. All the radio systems were simulated using a multiplex stereo modulator VP – 7633A (Panasonic), FM/AM signal generator VP – 8119A (Panasonic), and ordinary consumer CD player and FM/AM radio receiver.

The percentage of correct bits recovered per watermark is measured before and after transmission. This is shown in a graphic where the percentage of correct bits before transmission is the continuous line, and the percentage of correct bit after transmission is the dotted line. Also, the offset from the expected starting point of each watermark after transmission is measured (in samples), as well as the total of watermarks recovered and the average recovery percentage. The tables with all the values for each test can be found in the section 8.1 of the appendix. Discussion of these system performances will be done in section 5.4.

5.1.1 MPEG LAYER 3

All the digital audio editing was done using the software *Sound Forge 4.0d* of Sonic Foundry. The encoding of the signals was done using the MPEG layer 3 codec at 64 KBps, 44100 Hz mono (left chanel). After the MPEG encoding, the same codec was used to convert back to PCM audio and recover the watermarks. The results are shown in Figure 46.

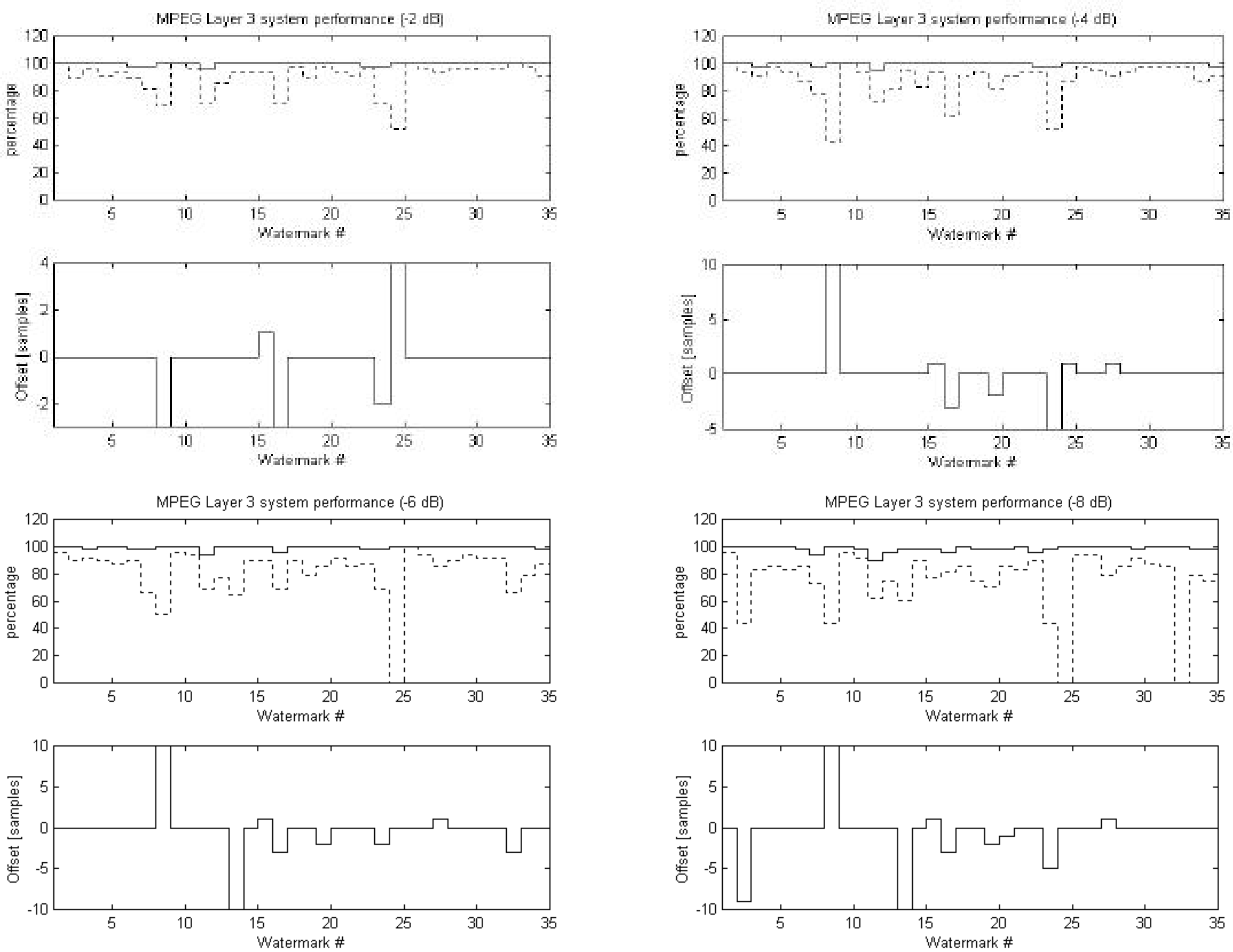

Figure 46. MPEG Layer 3 system performance

5.1.2 MINI DISC

A Sony portable Mini Disc Recorder MZ-1 was used for the following test. The analog output of the computer’s sound card was routed to the line input of the Mini Disc recorder. The AGC option was disabled. All the tracks were played from the Mini Disc and recorded into the computer using the analog input of the sound card. The results are shown in Figure 47.

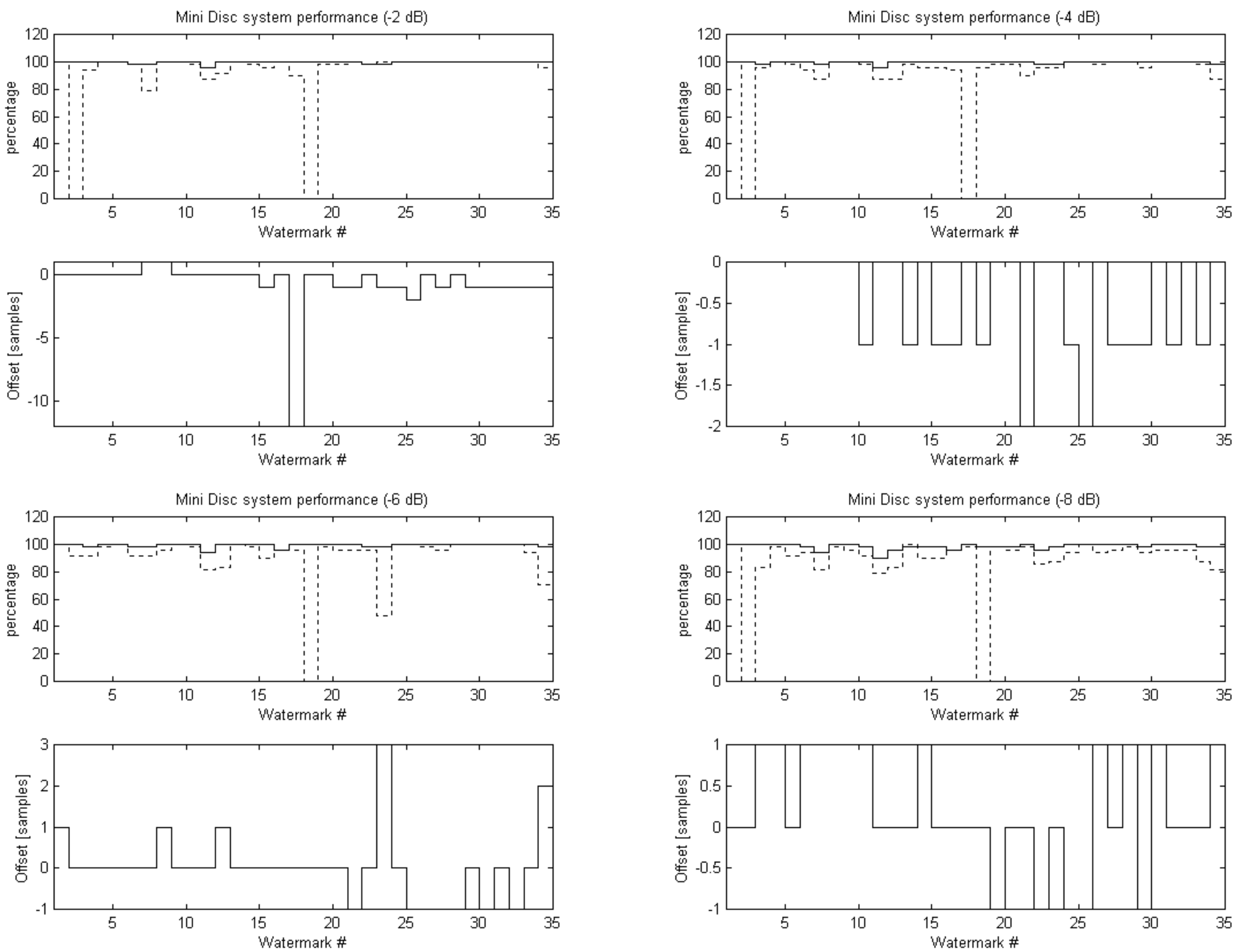

Figure 47. Mini Disc system performance

5.1.3 D/A A/D

For the digital to analog – analog to digital test, two sets of conversions were made. Each conversion used the computer's sound card analog output and input to loop back the audio into the computer. This process was done twice. The results are shown in Figure 48.

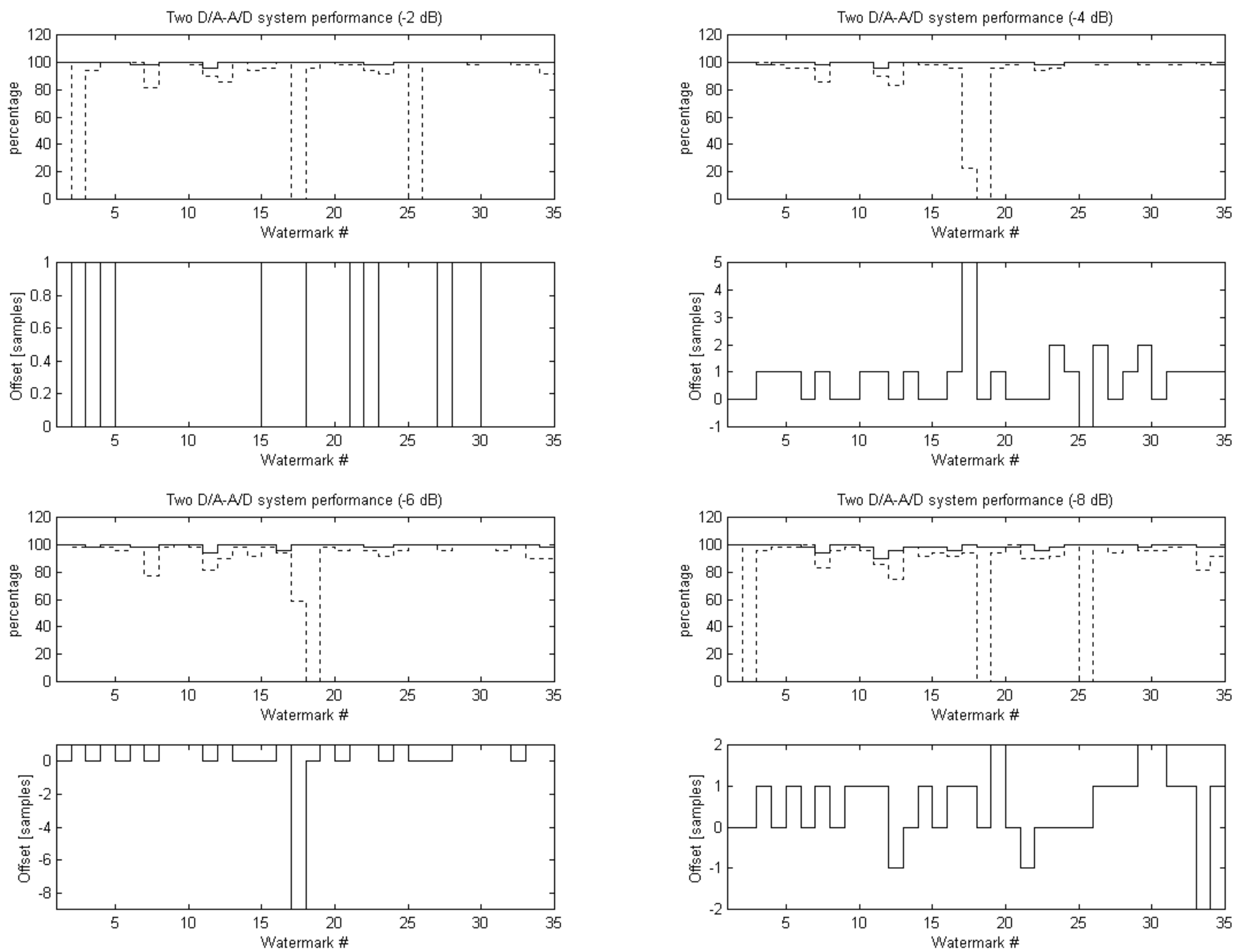

Figure 48. Two D/A – A/D system performance

5.1.4 ANALOG TAPE

The output of the sound card was fed in a four track *Tascam* Portastudio 464. The tape speed selected was "high," and DBX noise reduction was enabled. The tape was played back into the computer. The results are shown in Figure 49.

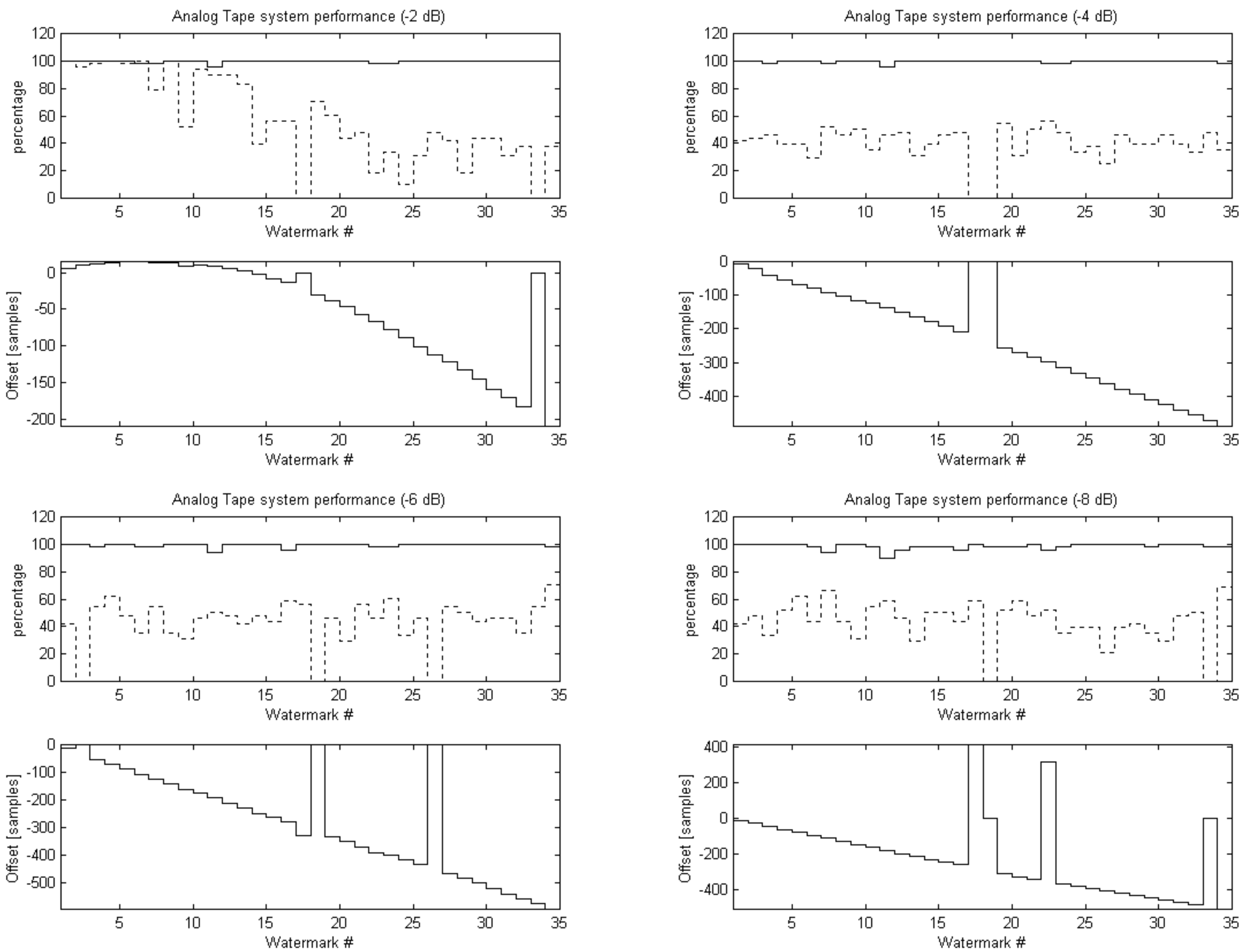

Figure 49. Analog Tape system performance

5.1.5 FM STEREO RADIO

For the FM Stereo transmission the carrier frequency was 97.7 MHZ and the power was 65 dBF. The results are shown in Figure 50 for the left channel and in Figure 51 for the right channel.

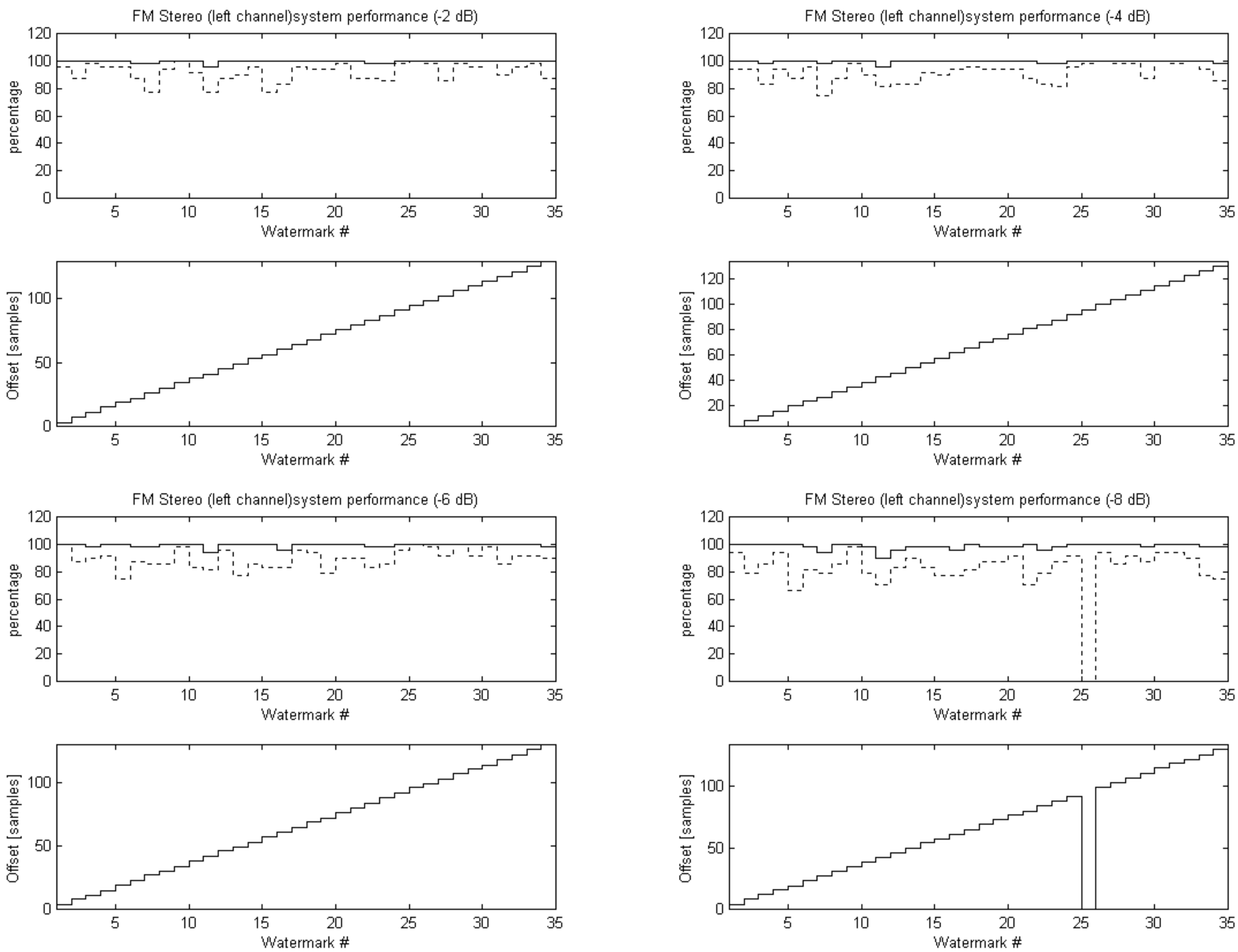

Figure 50. FM Stereo (left channel) system performance

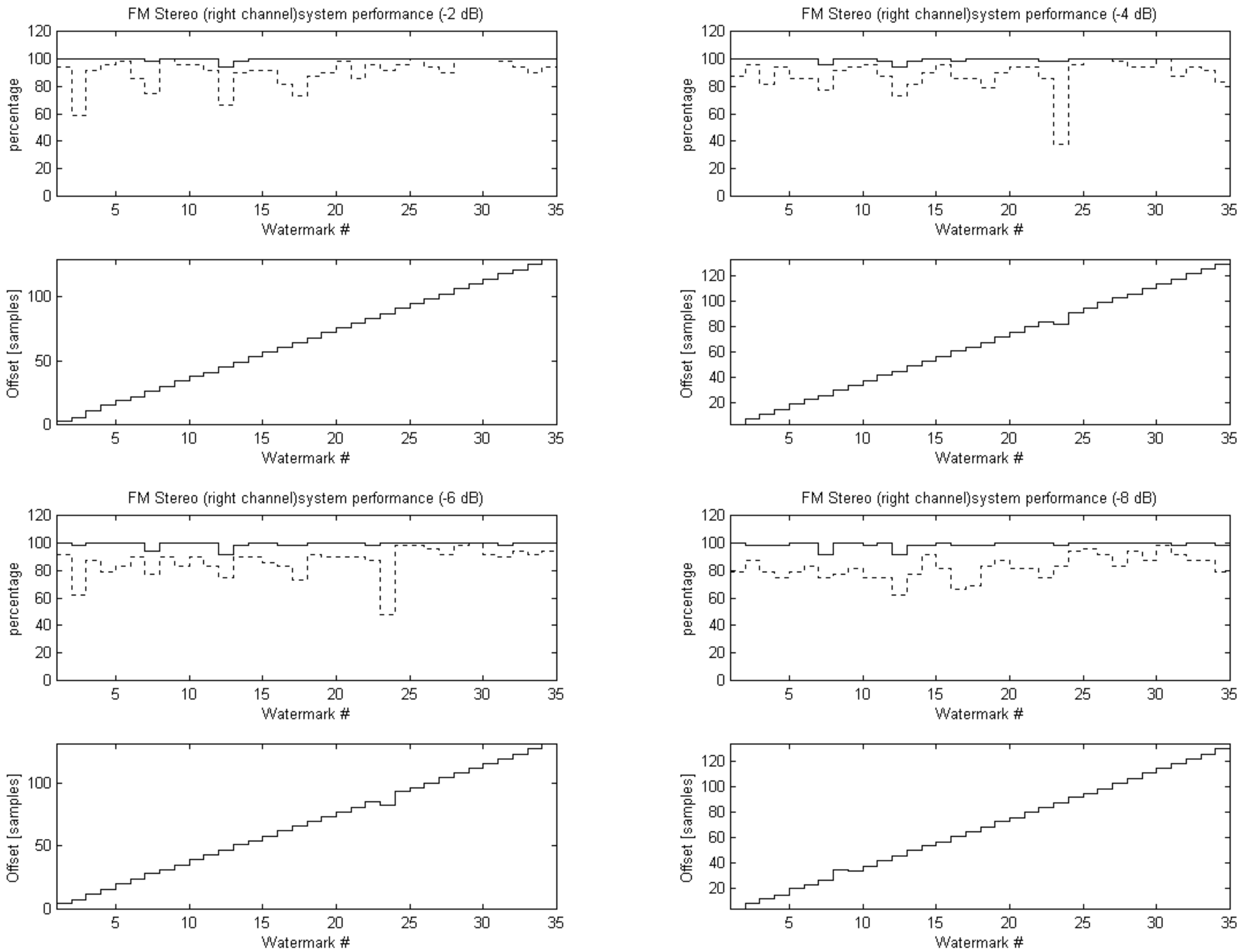

Figure 51. FM Stereo (right channel) system performance

5.1.6 FM RADIO MONO

This test uses the same configuration as the FM Stereo test, but the modulation was done monophonically using just the left channel. The results are shown in Figure 52.

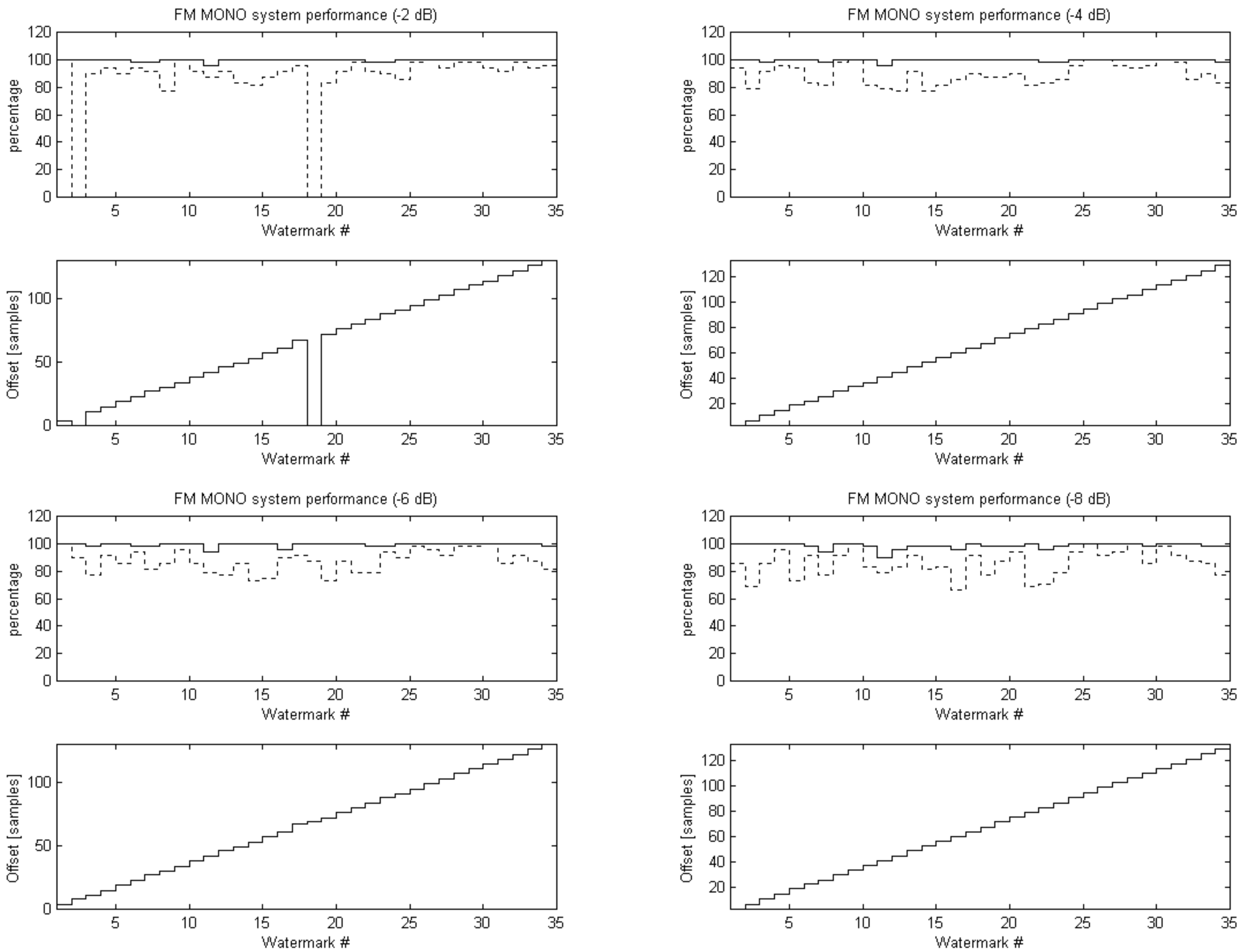

Figure 52. FM MONO system performance

5.1.7 FM RADIO MONO WEAK SIGNAL

The same setup used for the FM Mono test was employed, but the power was reduced to 10 dBF. The results are shown in Figure 53.

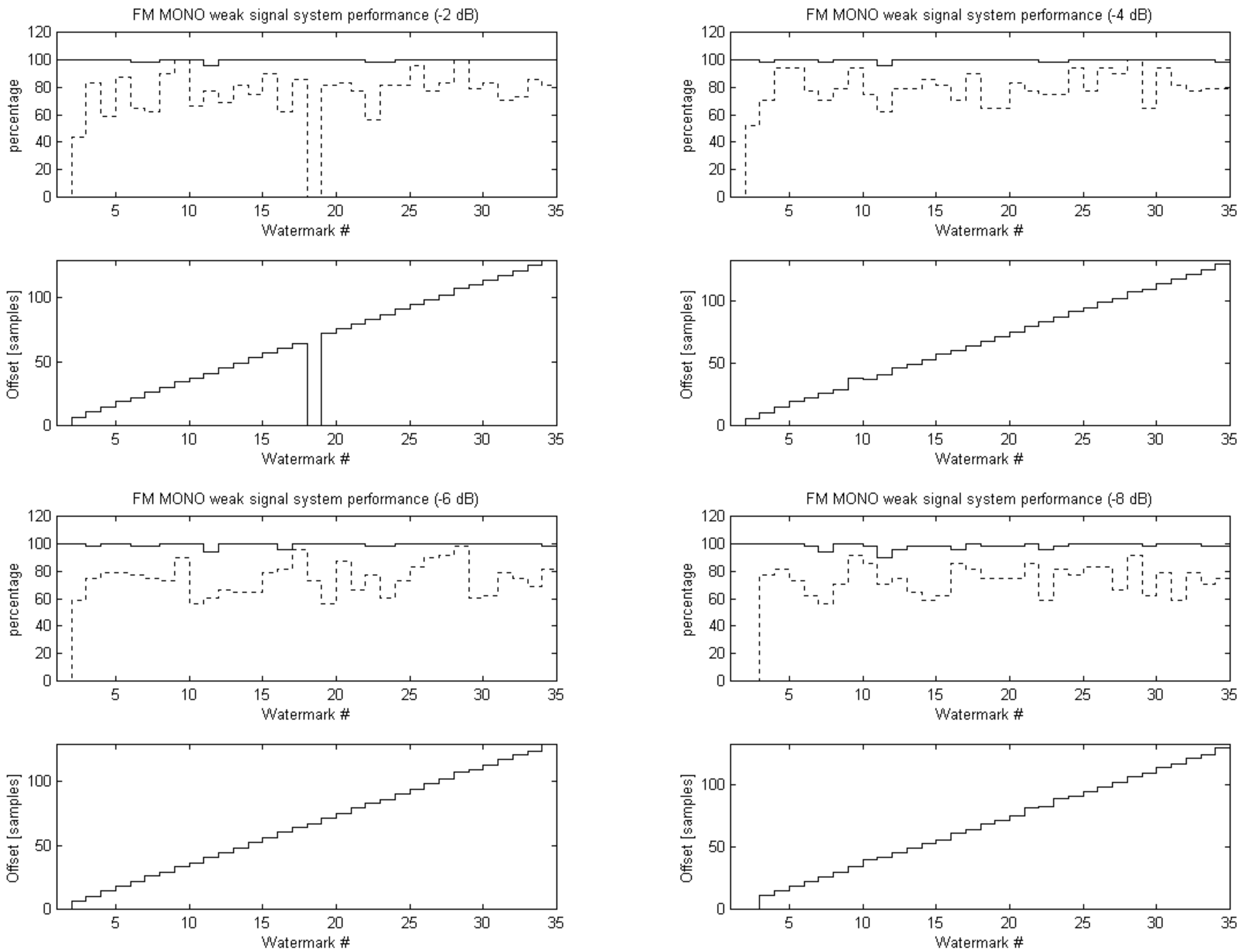

Figure 53. FM MONO weak signal system performance

5.1.8 AM RADIO

For this test the same setup for the FM Stereo test was used. The carrier frequency was 1 MHz, and the power was 60 dBF. The results are shown in Figure 54.

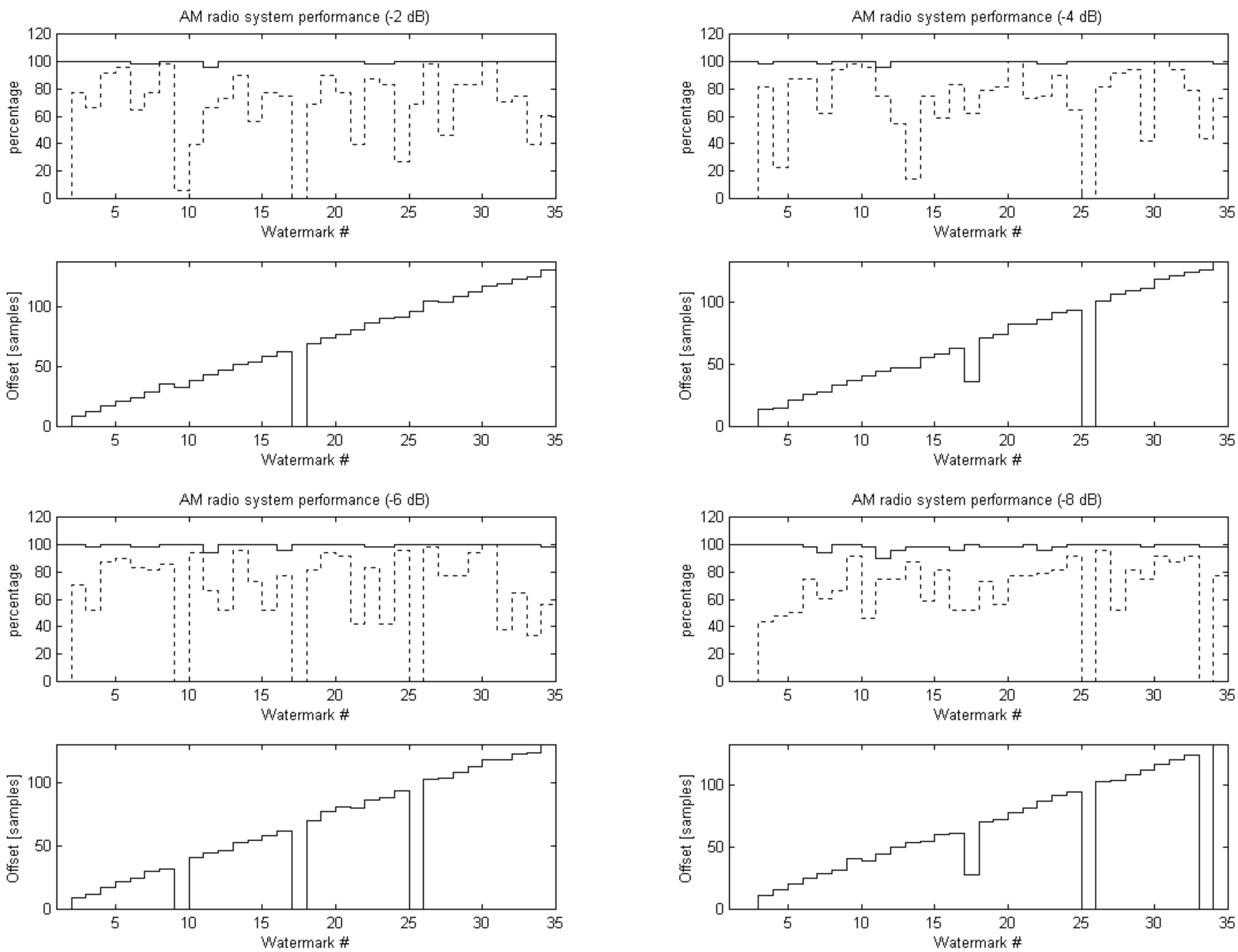

Figure 54. AM radio system performance

5.2 ENERGY INJECTED/REMOVED RATIO

An energy ratio *E* is now defined that measured the difference at the audio signal before and after the watermarking process. It uses the removed energy per critical band and the injected energy per critical band.

$$E = \frac{\frac{1}{FRAMES}\sum_{1}^{FRAMES}\left(\frac{1}{Zt}\sum_{1}^{Zt}\mathrm{EI}\right)}{\frac{1}{FRAMES}\sum_{1}^{FRAMES}\left(\frac{1}{Zt}\sum_{1}^{Zt}\mathrm{ER}\right)} \tag{98}$$

Where:

FRAMES is the number of frames in each watermark
Zt is the number of critical bands
EI is the energy injected per critical band
ER is the energy removed per critical band

The effect of the value of *E* is summarized in Table 3:

$E = 0$	No energy injected. There is no watermark information in the signal.
$0 < E < 1$	Less energy injected than removed.
$E = 1$	Equal amount of energy injected and removed.
$E > 1$	More energy injected than removed

Table 3. Effect of the value of *E*

The average of all *E* calculated on all frames is the total energy *Etotal*.

5.2.1 E ratio in W2 (watermark –2 dB below the masking threshold)

Figure 56 shows the spectral analysis of energy removed from the audio signal over a period equal to 1154 frames (35 watermarks). The x-axis shows the frame number, the y-axis shows the critical bands and the intensity of each mark shows the actual energy removed per critical band (see the scale in the colorbar). Figure 56 uses the same axes and colorbar as to display the energy injected in each critical band. Figure 57 shows the ratio *E* per frame and per watermark. The *Etotal* of W2 is 1.0376, W4 is 0.6718, W6 is 0.4345 and W8 is 0.2837.

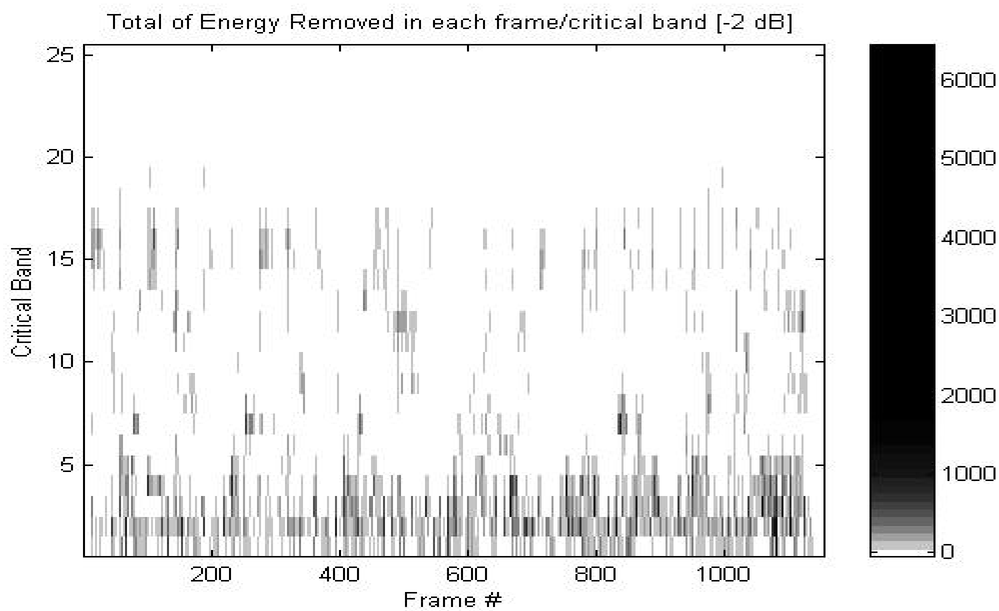

Figure 55. Energy removed per critical band (W2 –2dB)

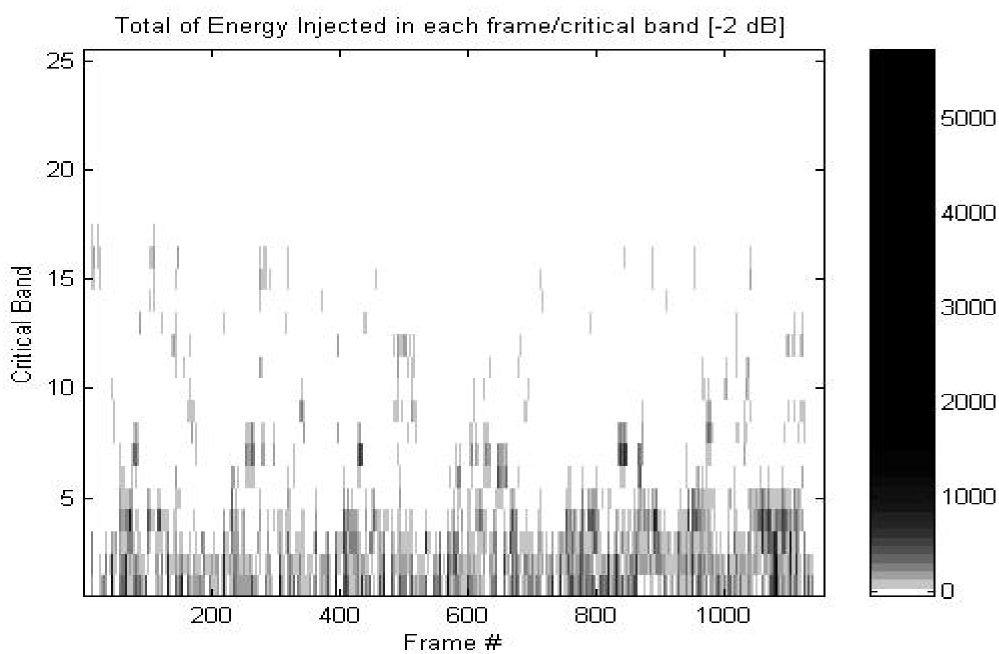

Figure 56. Energy injected per critical band (W2 –2 dB)

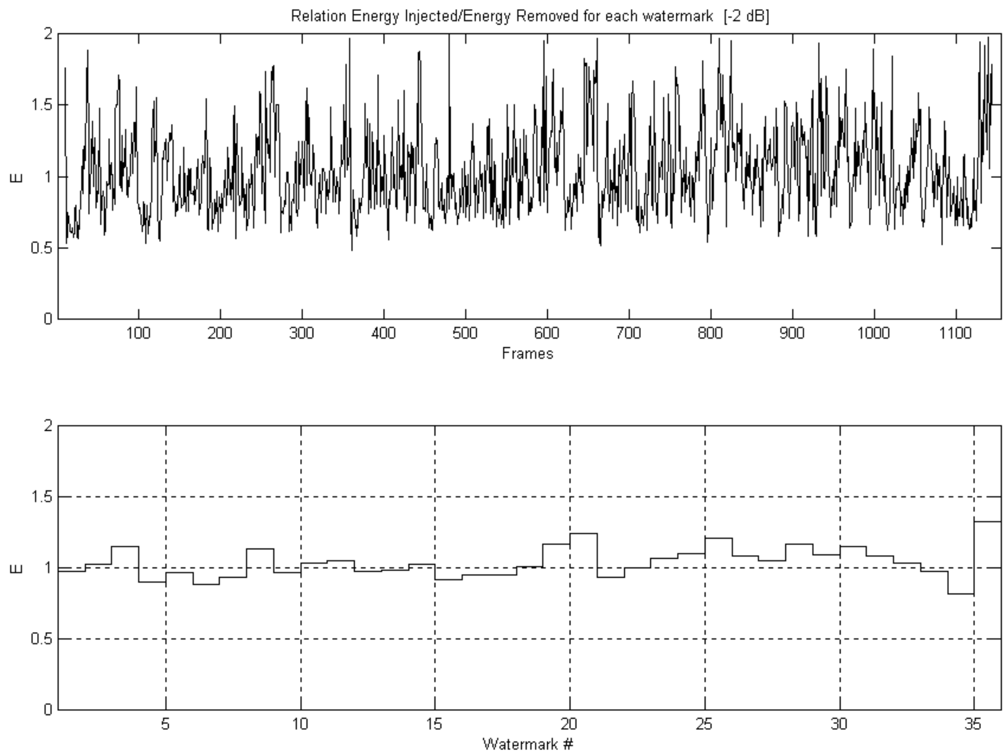

Figure 57. (a) *E* ratio for each frame and, (b) average *E* ratio for each watermark (W2 –2dB)

5.2.2 E ratio in W4 (watermark –4 dB below the masking threshold)

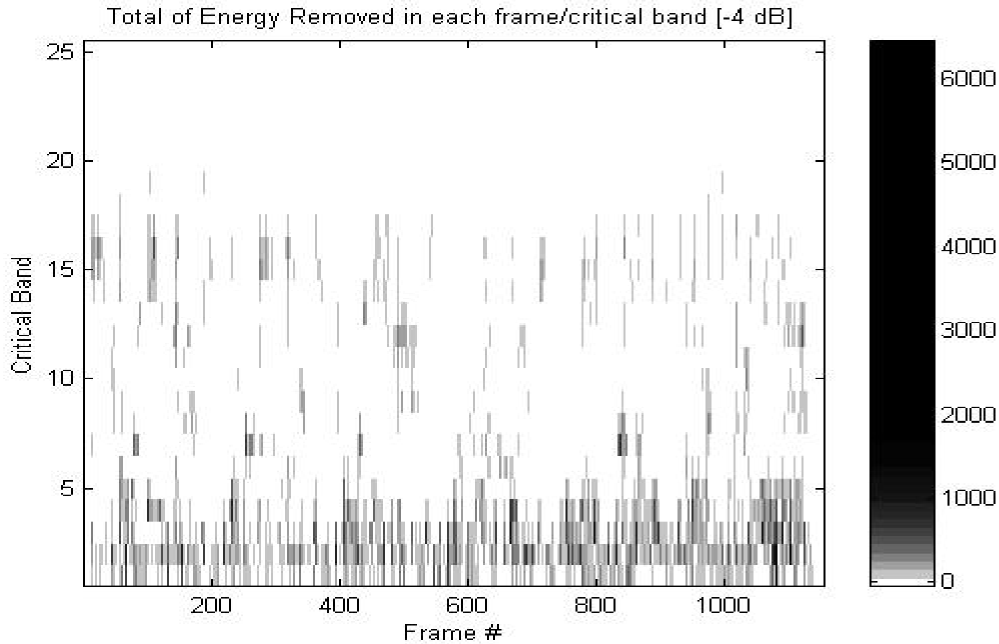

Figure 58. Energy removed per critical band (W4 –4dB)

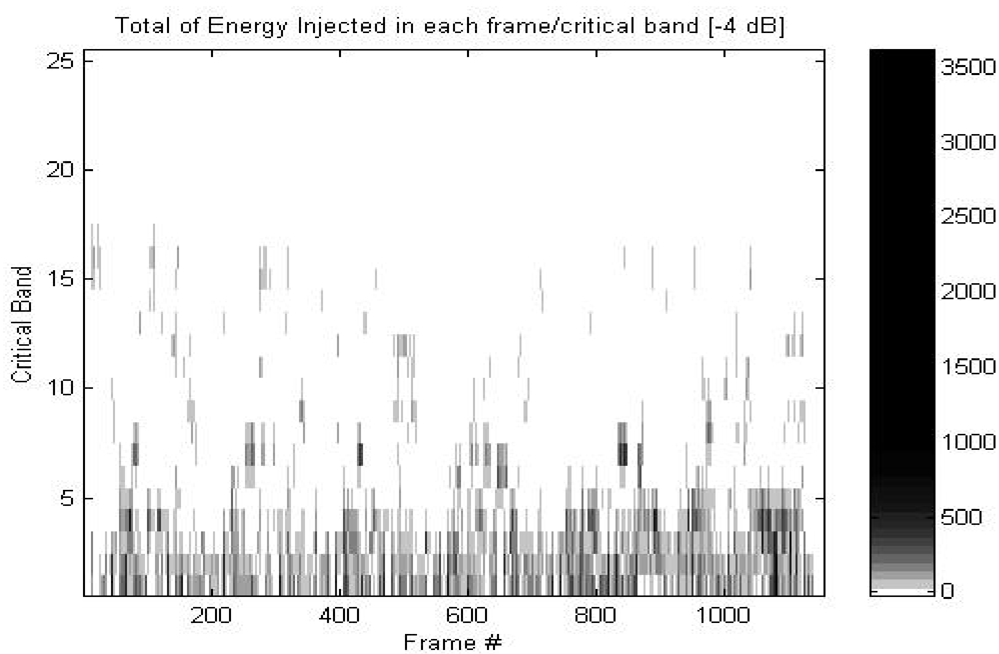

Figure 59. Energy injected per critical band (W4 –4 dB)

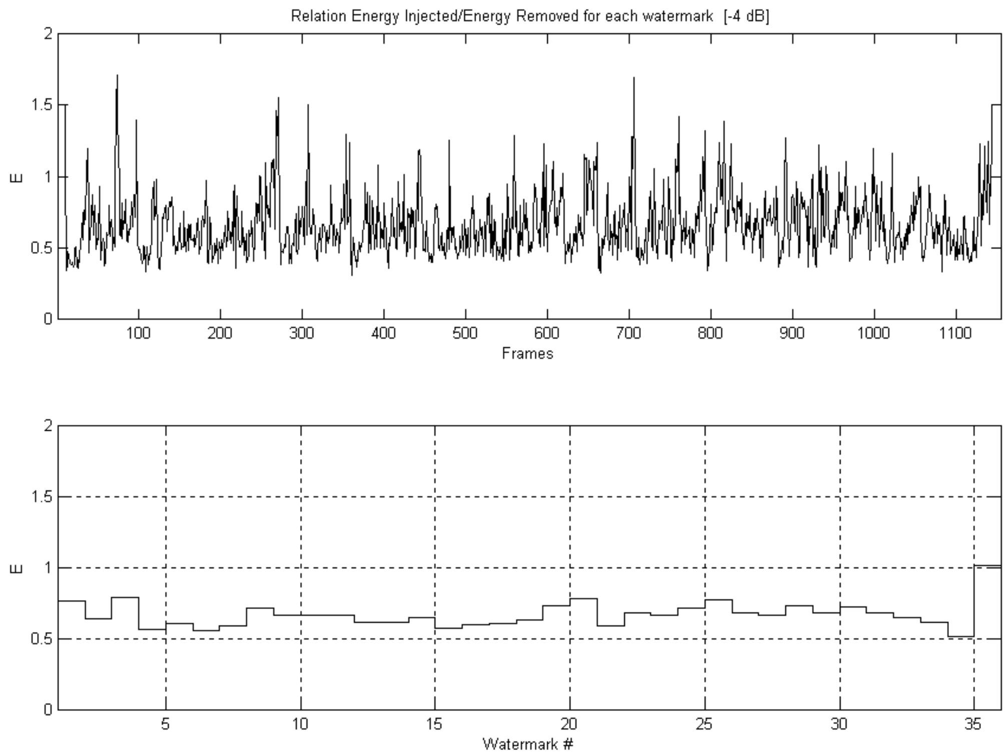

Figure 60. (a) *E* ratio for each frame and, (b) average *E* ratio for each watermark (W4 –4dB)

5.2.3 E ratio in W6 (watermark –6 dB below the masking threshold)

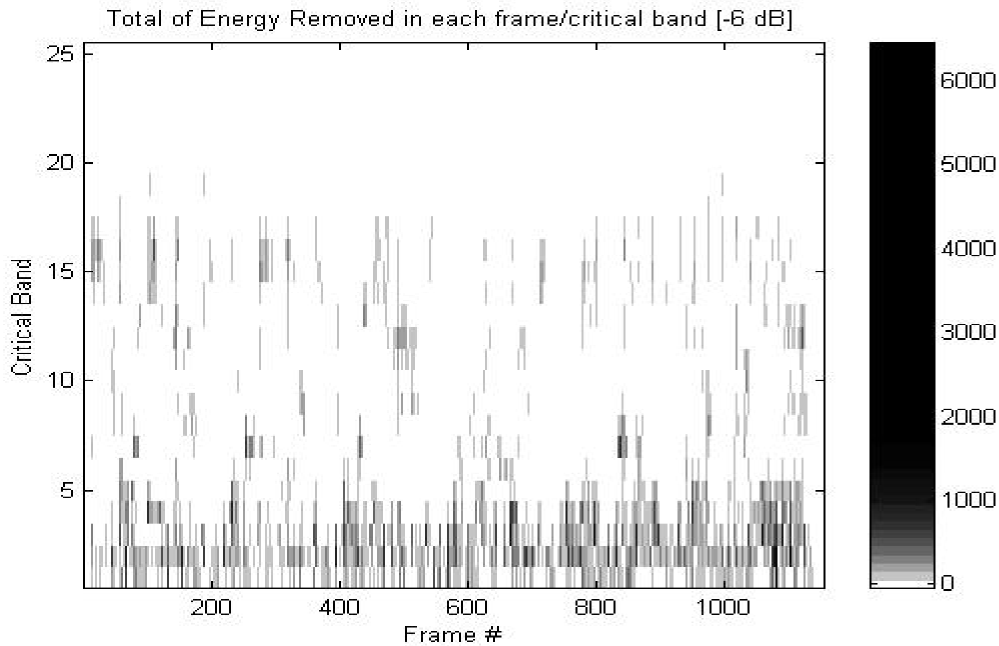

Figure 61. Energy removed per critical band (W6 –6dB)

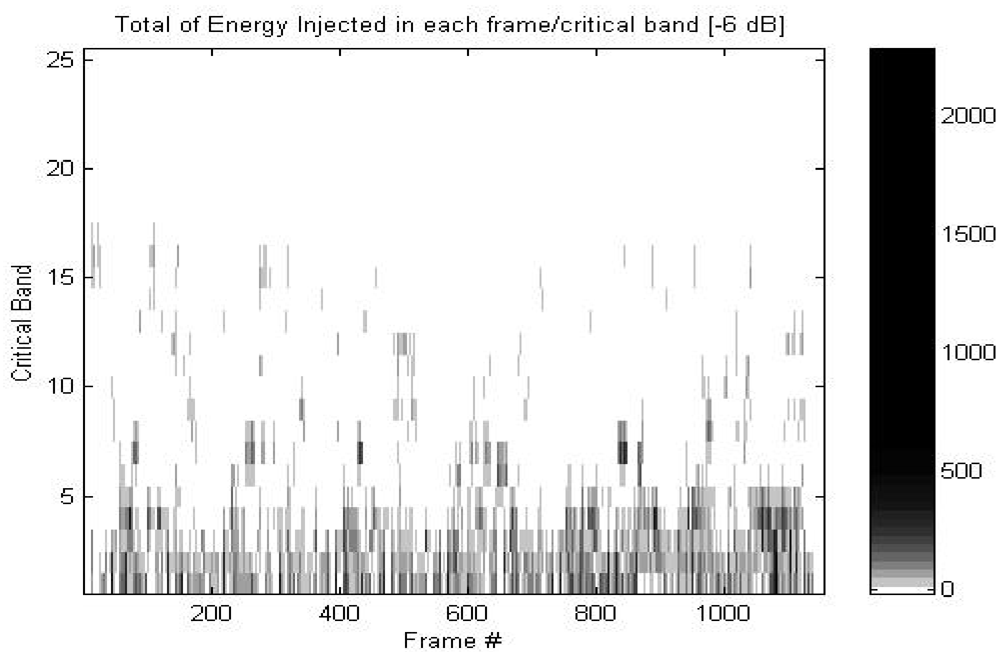

Figure 62. Energy injected per critical band (W6 –6 dB)

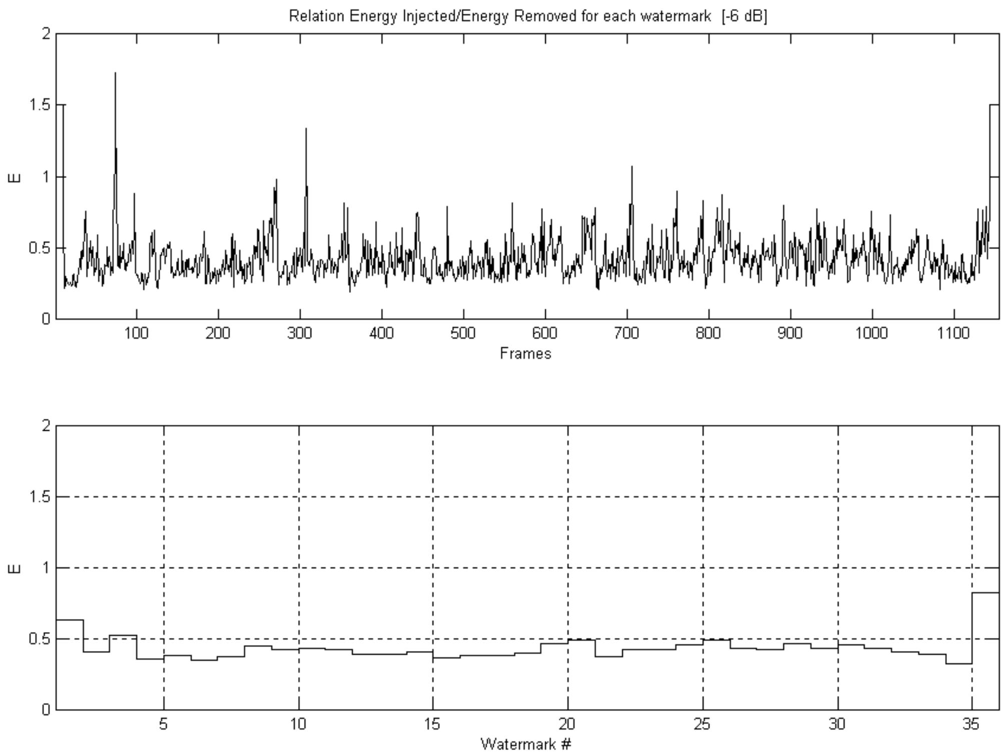

Figure 63. (a) *E* ratio for each frame and, (b) average *E* ratio for each watermark (W6 –6dB)

5.2.4 E ratio in W8 (watermark –8 dB below the masking threshold)

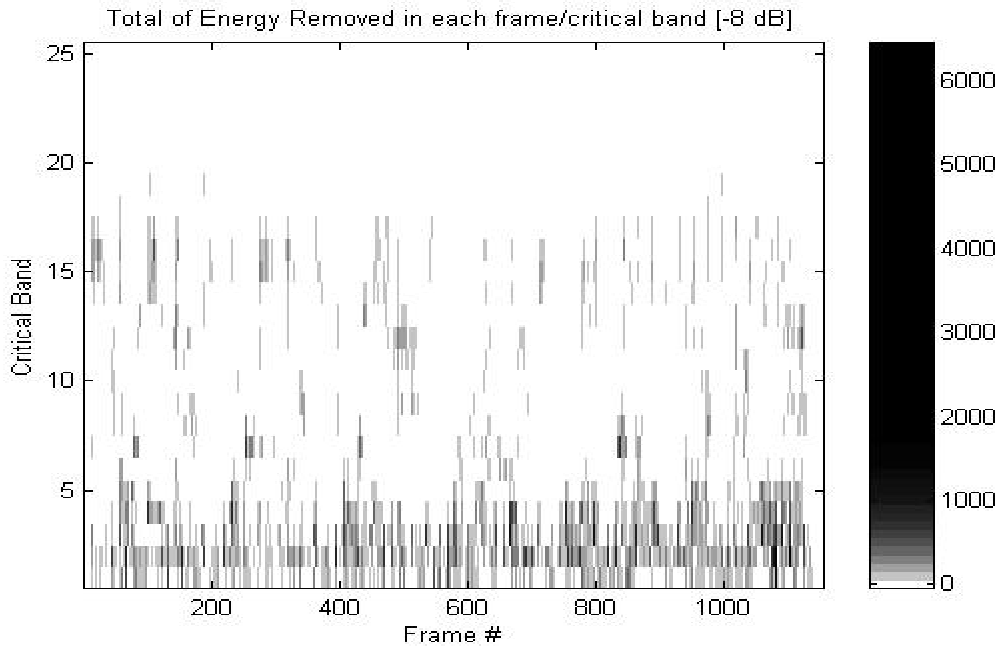

Figure 64. Energy removed per critical band (W8 –8dB)

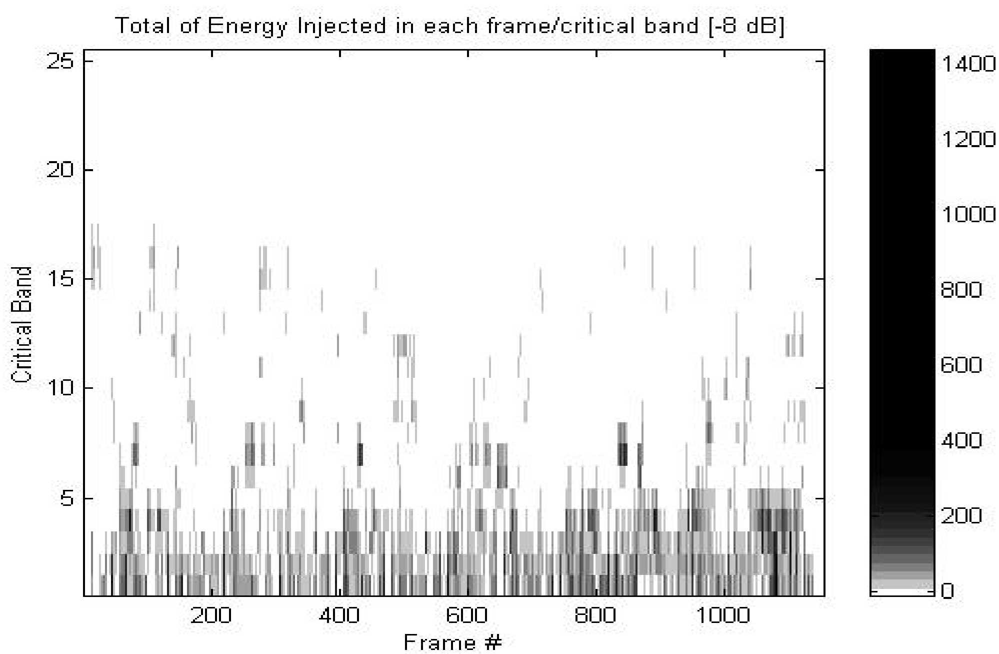

Figure 65. Energy injected per critical band (W8 –8 dB)

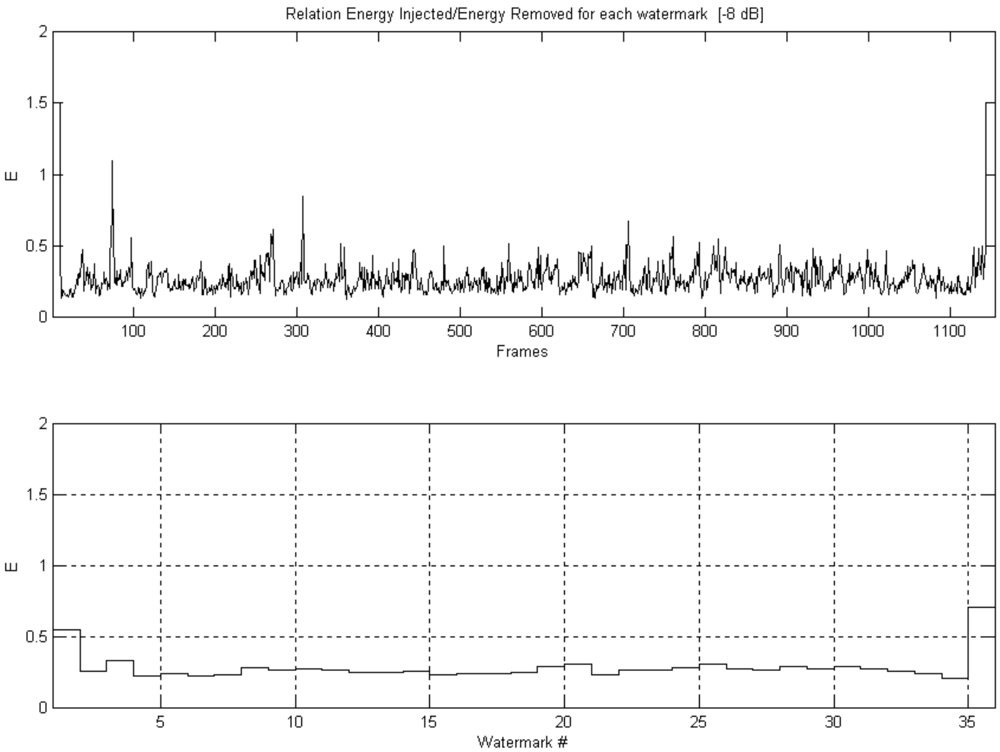

Figure 66. (a) *E* ratio for each frame and, (b) average *E* ratio for each watermark (W8 –8dB)

5.3 LISTENING TEST

One of the requirements of the watermarking system is to retain the perceptual quality of the signal. This is often referred to as "transparency." The transparency of the watermarking algorithm was tested using three of the four watermarked audio signals (W2, W4 and W6) used in section 5.1.

An ABX listening test was used as the testing mechanism. In an ABX test the listener can hear selection A (in this case the non-watermarked audio), selection B (the watermarked audio) and X (either the watermarked or non-watermarked audio). The listener is then asked to decide if selection X is equal to A or B. The number of correct answers is the basis to decide if the watermarked audio is perceptually different than the original audio and would, therefore, declare the watermarking algorithm as "non-transparent." In the other case, if the watermarked audio is perceptually equal to the original audio, the watermarking algorithm will be declared as "transparent."

Using the theory explained in Burstein [19], [20], different parameters were selected to find an appropriate sample size. A criterion of significance $\alpha'=0.1$ is selected (also known as Error Type 1). The Type 2 error risk is assumed $\beta'=0.1$. The probability $p1$ that a listener finds the right answer by chance is 0.5 in an ABX system. The effect size is selected as $p2=0.7$. With these parameters, the approximated required sample size that meets the specifications is 37.61 samples. The sample size is selected as $n=40$. (40 listeners per ABX set). The critical c (c') is the minimum number of correct samples which, together with n and $p1$, can produce a significance level α equal to or less than the specified criterion of significance α'. The calculated c' is 24.55 and can be rounded off to

25. This is the minimum number of correct answers to accept the hypothesis that the listener perceives differences between audio A and B. With *c'*=25, the criterion of significance becomes α'=0.78, which is below the required level. The type 2 error risk β'=1.11 and does not exceed desired level. The results and their approximate significance level are shown in Table 4.

	Total samples	Correct Identifications	α
W2	40	24	0.14
W4	40	19	0.50
W6	40	19	0.50

Table 4. Listening test results

5.4 DISCUSSION

The survival over different channels showed that after encoding, 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.

The theoretical position of the watermark and the offset of the actual watermark represent the starting position of the *{header}* of each watermark. This position will not affect the recovery of the watermark because each watermark is embedded independently of the others. In the actual tests three different cases are seen: almost no offset, linearly increasing offset and varying offset. When no offset is seen, the original signal and the recorded signal after transmission where played at the same speed. This is the case in 5.1.1 to 5.1.3. In the cases where the offset is linearly increasing (5.1.5 to 5.1.8), it is assumed that the speed of the playback device (in this case an ordinary consumer CD player was different (slightly slower) than the recording device. The last case, 5.1.4, shows the unstable speed variations of the tape device. If the speed of the playback device is close enough to the original speed, the de-spreading can be successful because the difference in alignment between the watermarked audio and the de-spreading signals (PN sequence, demodulator and *{header}*) will not greatly affect the final result.

Finally, the percentage of correct bits recovered measures quality of the recovery for each watermark. Notice that not all the watermarks are recovered (*%bits* = 0.0), and not all the watermarks are recovered in their totality but many of them were recovered with more than 80% of the bits. A good bit error detection/correction algorithm or averaging technique could substantially improve the recovery of the watermark. A very strong point in the watermarking system is the redundancy of watermarks embedded into the audio stream. In this case, each watermark lasts approximately 600 ms. Even if just a few watermarks are recovered, the goal of transmitting the watermark information within the audio signal and recovering it afterwards is accomplished.

The listening test showed that the watermark at –2dB below the masking threshold (W2) is the most likely to be heard, but it can not be ensured that people actually noticed the difference. For all the other watermarked signals, the results show that the process is “transparent.”

6 CONCLUSIONS

The proposed digital watermarking method for audio signals is based on a psychoacoustic auditory model to shape an audio watermark signal that is generated using spread spectrum techniques. The method retains the perceptual quality of the audio signal, while being resistant to diverse removal attacks, either intentional or unintentional. The recovery of the watermark is accomplished without knowledge of the original audio signal. The only information used includes the watermarked audio signal, and the parameters used for the watermark generation.

The psychoacoustic auditory model retrieves the necessary information about the masking threshold of the input audio signal. This model is a good approach that can be used for several applications such: perceptual coding, masking analysis, or watermark embedding. The spread spectrum theory describes two important Direct Sequence techniques, but the employed technique is Coded Direct-Sequence Spread Binary Phase-Shift-Keying (coded DS/BPSK). Because the normal literature about this topic is reserved for communication theory, some assumptions were made to use the theory in an audio bandwidth environment. Specifically in this case, the audio information was considered the “noise” or “jammer” signal that interferes with the watermark.

Future research could be performed in different aspects of this proposed algorithm such as:

- System performance with different types of music.
- Experimenting with different spread spectrum encoding parameters.
- Changes in the playback speed of the signal.

- Crosstalk interference.
- Multiple watermark embedding.
- Use of techniques to enhance recovery of the watermark (i.e., bit error detection/ correction, averaging, etc).
- Real - time implementation.
- Investigate different signal schemes for the generation of the PN sequence.

7 REFERENCES

[1] 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)**

[2] T. Sporer and K. Brandenburg, "Constraints of Filter Banks Used for Perceptual Measurement," ***J. Audio Eng. Soc.*****, vol. 43, pp. 107 - 115 (1995 March)**

[3] J. Mourjopoulos and D. Tsoukalas, "Neural Network Mapping to Subjective Spectra of Music Sounds," ***J. Audio Eng. Soc.*****, vol. 40, pp. 253 - 259 (1992 April)**

[4] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," ***IEEE Journal on Selected Areas in Communications,*** **vol. 6, pp. 314 – 323 (1988 Feb.)**

[5] M. K. Simon, J. K Omura, R A. Scholtz and B. K. Levitt, ***Spread Spectrum Communications Handbook*** **(McGraw-Hill, New York, 1994)**

[6] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of Spread-Spectrum Communications – A Tutorial," ***IEEE Transactions on Communications,*** **vol. COM-30, pp. 855 – 884 (1982 May)**

[7] C. S. Lindquist, ***Adaptive & Digital Signal Processing with Digital Filtering Applications*** **(Steward & Sons, Miami, 1989)**

[8] L. R. Rabiner, and R. W. Schafer, ***Digital Processing of Speech Signals*** **(Prentice Hall, New Jersey, 1978)**

[9] E. Zwicker, and h. Fastl, ***Psychoacoustics Facts and Models*** **(Springer-Verlag, Berlin, 1990)**

[10] D. L. Nicholson, ***Spread Spectrum Signal Design. LPE & AJ Systems*** **(Computer Science Press, Rockville, Maryland, 1988)**

[11]C. Neubauer and J. Herre, "Digital Watermarking and Its Influence on Audio Quality," presented at the 105th Convention of the Audio Engineering Society, ***J. Audio Eng. Soc. (Abstracts),*** **vol. 46, pp. 1041 (1998 November), preprint 4823.**

[12] J. G. Roederer, ***The Physics and Psychophysics of Music*** **(Springer-Verlag, New York, 1995)**

[13] J. G. Beerends and J. A. Stemerdink, "A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation," ***J. Audio Eng. Soc.*****, vol. 42, pp. 115 - 123 (1994 March)**

[14] J. G. Beerends and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation," ***J. Audio Eng. Soc.*****, vol. 40, pp. 963 - 978 (1992 December)**

[15] C. Colomes, M. Lever, J. B. Rault, Y. F. Dehery and G. Faucon, "A Perceptual Model Applied to Audio Bit-Rate Reduction," ***J. Audio Eng. Soc.*****, vol. 43, pp. 233 - 239 (1995 April)**

[16] T. Sporer, G. Gbur, J. Herre and R. Kapust, "Evaluating a Measurement System," ***J. Audio Eng. Soc.*****, vol. 43, pp. 353 - 362 (1995 May)**

[17] M. R. Schroeder, B. S. Atal and J. L. Hall, "Optimizing Digital Speech Coders by Exploiting

Masking Properties of the Human Ear," ***J. Acoust. Soc. Am.*****, vol. 66, pp. 1647 – 1652 (1979 Dec.)**

[18] B. Paillard, P. Mabilleau, S. Morissette and J. Soumagne, "PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals," ***J. Audio Eng. Soc.*****, vol. 40, pp. 21 - 31 (1992 Jan./Feb.)**

[19] H. Burstein, "By the Numbers," ***Audio*****, vol. 74, pp. 43 – 48 (1990 Feb.)**

[20] H. Burstein, "Approximation Formulas for Error Risk and Sample Size in ABX Testing," ***J. Audio Eng. Soc.*****, vol. 36, pp. 879 - 883 (1988 Nov.)**

[21] S. Haykin, ***Communication Systems 3rd ed.*** **(Wiley, New York, 1994)**

[22] R. L. Shrader, ***Electronic Communication 5th ed.*** **(McGraw Hill, New York, 1985)**

[23] L. Boney, A. H. Tewfik and K. N. Hamdy, "Digital Watermarks for Audio Signals," ***IEEE Int.Conf. on Multimedia Computing and Systems*****, Hiroshima, Japan (June 1996)**

[24] I. J. Cox, "Spread Spectrum Watermark for Embedded Signalling", ***United States Patent 5,848,155*** **(1998 Dec)**

8 APPENDIX

8.1 TRANSMISSION OVER DIFFERENT CHANNELS (RESULTS)

Explanation of the fields in each table:

position	Theoretical position of each watermark measured in samples from the beginning of the audio signal.
%orig	Percentage of bits per watermark recovered after watermark embedding. (before transmission).
Off	Offset of the actual watermark with respect its theoretical position.
%bits	Percentage of bits recovered after transmission by the actual channel.
watermarks detected	Number of watermarks recovered (maximum = 35).
average % bits recovered	The sum of all the *%bits* over the number of watermarks detected

8.1.1 MPEG LAYER 3

	Original Position	MPEG -2 dB %orig.	Off.	%bits	MPEG -4 dB %orig.	Off.	%bits	MPEG -6 dB %orig.	Off.	%bits	MPEG -8 dB %orig.	Off.	%bits
[1]	2049	100.00	0	100.00	100.00	0	100.00	100.00	0	95.83	100.00	0	95.83
[2]	35841	100.00	0	89.58	100.00	0	93.75	100.00	0	89.58	100.00	-9	43.75
[3]	69633	100.00	0	95.83	97.92	0	91.67	97.92	0	91.67	100.00	0	83.33
[4]	103425	100.00	0	91.67	100.00	0	97.92	100.00	0	89.58	100.00	0	85.42
[5]	137217	100.00	0	93.75	100.00	0	93.75	100.00	0	87.50	100.00	0	83.33
[6]	171009	97.92	0	89.58	100.00	0	87.50	97.92	0	89.58	97.92	0	85.42
[7]	204801	97.92	0	81.25	97.92	0	77.08	97.92	0	66.67	93.75	0	72.92
[8]	238593	100.00	-3	68.75	100.00	10	43.75	100.00	10	50.00	100.00	10	43.75
[9]	272385	100.00	0	100.00	100.00	0	100.00	100.00	0	95.83	100.00	0	95.83
[10]	306177	100.00	0	95.83	100.00	0	93.75	100.00	0	93.75	97.92	0	91.67
[11]	339969	95.83	0	70.83	95.83	0	72.92	93.75	0	68.75	89.58	0	62.50
[12]	373761	100.00	0	85.42	100.00	0	81.25	100.00	0	77.08	95.83	0	75.00
[13]	407553	100.00	0	93.75	100.00	0	95.83	100.00	-10	64.58	97.92	-10	60.42
[14]	441345	100.00	0	93.75	100.00	0	83.33	100.00	0	89.58	97.92	0	89.58
[15]	475137	100.00	1	93.75	100.00	1	93.75	100.00	1	89.58	97.92	1	77.08
[16]	508929	100.00	-3	70.83	100.00	-3	62.50	95.83	-3	68.75	95.83	-3	81.25
[17]	542721	100.00	0	97.92	100.00	0	91.67	100.00	0	89.58	100.00	0	85.42
[18]	576513	100.00	0	89.58	100.00	0	93.75	100.00	0	79.17	97.92	0	75.00
[19]	610305	100.00	0	97.92	100.00	-2	81.25	100.00	-2	85.42	97.92	-2	70.83
[20]	644097	100.00	0	93.75	100.00	0	91.67	100.00	0	91.67	97.92	-1	85.42
[21]	677889	100.00	0	91.67	100.00	0	93.75	100.00	0	85.42	100.00	0	83.33
[22]	711681	97.92	0	95.83	97.92	0	93.75	97.92	0	87.50	95.83	0	89.58
[23]	745473	97.92	-2	70.83	97.92	-5	52.08	97.92	-2	68.75	97.92	-5	43.75
[24]	779265	100.00	4	52.08	100.00	1	87.50	100.00	0	0.00	100.00	0	0.00
[25]	813057	100.00	0	100.00	100.00	0	97.92	100.00	0	100.00	100.00	0	93.75
[26]	846849	100.00	0	95.83	100.00	0	95.83	100.00	0	93.75	100.00	0	93.75
[27]	880641	100.00	0	93.75	100.00	1	91.67	100.00	1	85.42	100.00	1	79.17
[28]	914433	100.00	0	95.83	100.00	0	93.75	100.00	0	89.58	100.00	0	85.42
[29]	948225	100.00	0	95.83	100.00	0	97.92	100.00	0	93.75	97.92	0	91.67
[30]	982017	100.00	0	100.00	100.00	0	97.92	100.00	0	91.67	100.00	0	87.50
[31]	1015809	100.00	0	95.83	100.00	0	97.92	100.00	0	91.67	100.00	0	85.42
[32]	1049601	100.00	0	100.00	100.00	0	97.92	100.00	-3	66.67	100.00	0	0.00
[33]	1083393	100.00	0	97.92	100.00	0	87.50	100.00	0	79.17	97.92	0	79.17
[34]	1117185	100.00	0	91.67	97.92	0	91.67	97.92	0	87.50	97.92	0	75.00
[35]	1150977	100.00	1	91.67	97.92	1	89.58	97.92	1	91.67	97.92	1	91.67
Watermark detected:				35			35			34			33
average % bits recovered				90.36			88.39			84.31			79.48

Table 5. MPEG layer 3 system performance

8.1.2 MINI DISC

	Original	MINI DISC -2 dB			MINI DISC -4 dB			MINI DISC -6 dB			MINI DISC -8 dB		
	Position	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	0	100.00	100.00	0	100.00	100.00	1	100.00	100.00	0	100.00
[2]	35841	100.00	0	0.00	100.00	0	0.00	100.00	0	91.67	100.00	0	0.00
[3]	69633	100.00	0	93.75	97.92	0	95.83	97.92	0	91.67	100.00	1	83.33
[4]	103425	100.00	0	100.00	100.00	0	100.00	100.00	0	97.92	100.00	1	97.92
[5]	137217	100.00	0	100.00	100.00	0	97.92	100.00	0	100.00	100.00	0	91.67
[6]	171009	97.92	0	97.92	100.00	0	93.75	97.92	0	91.67	97.92	1	93.75
[7]	204801	97.92	1	79.17	97.92	0	87.50	97.92	0	91.67	93.75	1	81.25
[8]	238593	100.00	1	100.00	100.00	0	100.00	100.00	1	95.83	100.00	1	97.92
[9]	272385	100.00	0	100.00	100.00	0	100.00	100.00	0	100.00	100.00	1	95.83
[10]	306177	100.00	0	97.92	100.00	-1	97.92	100.00	0	97.92	97.92	1	91.67
[11]	339969	95.83	0	87.50	95.83	0	87.50	93.75	0	81.25	89.58	0	79.17
[12]	373761	100.00	0	91.67	100.00	0	87.50	100.00	1	83.33	95.83	0	83.33
[13]	407553	100.00	0	100.00	100.00	-1	97.92	100.00	0	100.00	97.92	0	100.00
[14]	441345	100.00	0	97.92	100.00	0	95.83	100.00	0	97.92	97.92	1	89.58
[15]	475137	100.00	-1	95.83	100.00	-1	95.83	100.00	0	89.58	97.92	0	89.58
[16]	508929	100.00	0	100.00	100.00	-1	93.75	95.83	0	95.83	95.83	0	95.83
[17]	542721	100.00	-12	89.58	100.00	0	0.00	100.00	0	95.83	100.00	0	97.92
[18]	576513	100.00	0	0.00	100.00	-1	95.83	100.00	0	0.00	97.92	0	0.00
[19]	610305	100.00	0	97.92	100.00	0	97.92	100.00	0	97.92	97.92	-1	97.92
[20]	644097	100.00	-1	97.92	100.00	0	97.92	100.00	0	95.83	97.92	0	95.83
[21]	677889	100.00	-1	100.00	100.00	-2	89.58	100.00	-1	95.83	100.00	0	97.92
[22]	711681	97.92	0	97.92	97.92	0	95.83	97.92	0	95.83	95.83	-1	85.42
[23]	745473	97.92	-1	100.00	97.92	0	95.83	97.92	3	47.92	97.92	0	87.50
[24]	779265	100.00	-1	100.00	100.00	-1	100.00	100.00	0	100.00	100.00	-1	93.75
[25]	813057	100.00	-2	100.00	100.00	-2	100.00	100.00	-1	100.00	100.00	-1	100.00
[26]	846849	100.00	0	100.00	100.00	0	97.92	100.00	-1	97.92	100.00	1	93.75
[27]	880641	100.00	-1	100.00	100.00	-1	100.00	100.00	-1	95.83	100.00	0	95.83
[28]	914433	100.00	0	100.00	100.00	-1	100.00	100.00	-1	100.00	100.00	1	97.92
[29]	948225	100.00	-1	100.00	100.00	-1	95.83	100.00	0	100.00	97.92	-1	93.75
[30]	982017	100.00	-1	100.00	100.00	0	100.00	100.00	-1	100.00	100.00	1	95.83
[31]	1015809	100.00	-1	100.00	100.00	-1	100.00	100.00	0	100.00	100.00	0	95.83
[32]	1049601	100.00	-1	100.00	100.00	0	100.00	100.00	-1	100.00	100.00	0	95.83
[33]	1083393	100.00	-1	100.00	100.00	-1	97.92	100.00	0	93.75	97.92	0	87.50
[34]	1117185	100.00	-1	95.83	97.92	0	87.50	97.92	2	70.83	97.92	1	81.25
[35]	1150977	100.00	0	0.00	97.92	0	0.00	97.92	0	0.00	97.92	0	0.00
Watermark detected:				32			32			33			32
average % bits recovered				97.53			96.35			93.75			92.64

Table 6. Mini Disc system performance

8.1.3 D/A A/D

Original		D/A - A/D -2 dB			D/A - A/D -4 dB			D/A - A/D -6 dB			D/A - A/D -8 dB		
Position		%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	1	100.00	100.00	0	100.00	100.00	0	100.00	100.00	0	100.00
[2]	35841	100.00	0	0.00	100.00	0	100.00	100.00	1	97.92	100.00	0	0.00
[3]	69633	100.00	1	93.75	97.92	1	100.00	97.92	0	97.92	100.00	1	95.83
[4]	103425	100.00	0	100.00	100.00	1	97.92	100.00	1	97.92	100.00	0	97.92
[5]	137217	100.00	1	100.00	100.00	1	95.83	100.00	0	95.83	100.00	1	97.92
[6]	171009	97.92	1	100.00	100.00	0	95.83	97.92	1	97.92	97.92	0	100.00
[7]	204801	97.92	1	81.25	97.92	1	85.42	97.92	0	77.08	93.75	1	83.33
[8]	238593	100.00	1	100.00	100.00	0	100.00	100.00	1	97.92	100.00	0	95.83
[9]	272385	100.00	1	100.00	100.00	0	100.00	100.00	1	100.00	100.00	1	97.92
[10]	306177	100.00	1	97.92	100.00	1	100.00	100.00	1	97.92	97.92	1	95.83
[11]	339969	95.83	1	89.58	95.83	1	89.58	93.75	0	81.25	89.58	1	85.42
[12]	373761	100.00	1	85.42	100.00	0	83.33	100.00	1	89.58	95.83	-1	75.00
[13]	407553	100.00	1	100.00	100.00	1	100.00	100.00	0	97.92	97.92	0	97.92
[14]	441345	100.00	1	93.75	100.00	0	97.92	100.00	0	91.67	97.92	1	91.67
[15]	475137	100.00	0	95.83	100.00	0	97.92	100.00	0	97.92	97.92	0	93.75
[16]	508929	100.00	0	100.00	100.00	1	95.83	95.83	1	93.75	95.83	1	91.67
[17]	542721	100.00	0	0.00	100.00	5	22.92	100.00	-9	58.33	100.00	1	93.75
[18]	576513	100.00	1	95.83	100.00	0	0.00	100.00	0	0.00	97.92	0	0.00
[19]	610305	100.00	1	100.00	100.00	1	95.83	100.00	1	97.92	97.92	2	93.75
[20]	644097	100.00	1	97.92	100.00	0	97.92	100.00	0	95.83	97.92	0	100.00
[21]	677889	100.00	0	97.92	100.00	0	100.00	100.00	1	100.00	100.00	-1	89.58
[22]	711681	97.92	1	93.75	97.92	0	93.75	97.92	1	95.83	95.83	0	89.58
[23]	745473	97.92	0	91.67	97.92	2	95.83	97.92	0	91.67	97.92	0	91.67
[24]	779265	100.00	0	100.00	100.00	1	100.00	100.00	1	95.83	100.00	0	100.00
[25]	813057	100.00	0	0.00	100.00	-1	100.00	100.00	0	100.00	100.00	0	0.00
[26]	846849	100.00	0	100.00	100.00	2	97.92	100.00	0	100.00	100.00	1	100.00
[27]	880641	100.00	1	100.00	100.00	0	100.00	100.00	0	95.83	100.00	1	93.75
[28]	914433	100.00	0	100.00	100.00	1	100.00	100.00	1	100.00	100.00	1	100.00
[29]	948225	100.00	0	97.92	100.00	2	97.92	100.00	1	100.00	97.92	2	95.83
[30]	982017	100.00	1	100.00	100.00	0	100.00	100.00	1	100.00	100.00	2	95.83
[31]	1015809	100.00	1	100.00	100.00	1	97.92	100.00	1	95.83	100.00	1	97.92
[32]	1049601	100.00	1	97.92	100.00	1	100.00	100.00	0	100.00	100.00	1	100.00
[33]	1083393	100.00	1	97.92	100.00	1	97.92	100.00	1	89.58	97.92	-2	81.25
[34]	1117185	100.00	1	91.67	97.92	1	100.00	97.92	1	89.58	97.92	1	91.67
[35]	1150977	100.00	0	0.00	97.92	0	0.00	97.92	0	0.00	97.92	0	0.00
Watermark detected:				31			33			33			31
average % bits recovered				96.77			95.08			94.51			94.02

Table 7. Two D/A – A/D system performance

8.1.4 ANALOG TAPE

Original		TAPE -2dB			TAPE -4dB			TAPE -6dB			TAPE -8dB		
Position		%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	6	100.00	100.00	-7	41.67	100.00	-13	41.67	100.00	-10	41.67
[2]	35841	100.00	10	95.83	100.00	-21	43.75	100.00	0	0.00	100.00	-26	47.92
[3]	69633	100.00	12	97.92	97.92	-40	45.83	97.92	-55	54.17	100.00	-44	33.33
[4]	103425	100.00	14	100.00	100.00	-54	39.58	100.00	-72	62.50	100.00	-63	52.08
[5]	137217	100.00	15	97.92	100.00	-67	39.58	100.00	-88	47.92	100.00	-77	62.50
[6]	171009	97.92	15	100.00	100.00	-79	29.17	97.92	-106	35.42	97.92	-94	43.75
[7]	204801	97.92	14	79.17	97.92	-91	52.08	97.92	-123	54.17	93.75	-110	66.67
[8]	238593	100.00	13	100.00	100.00	-103	45.83	100.00	-142	35.42	100.00	-125	43.75
[9]	272385	100.00	8	52.08	100.00	-116	50.00	100.00	-162	31.25	100.00	-148	31.25
[10]	306177	100.00	11	93.75	100.00	-125	35.42	100.00	-176	45.83	97.92	-161	54.17
[11]	339969	95.83	8	89.58	95.83	-137	45.83	93.75	-193	50.00	89.58	-177	58.33
[12]	373761	100.00	5	89.58	100.00	-150	47.92	100.00	-211	47.92	95.83	-195	45.83
[13]	407553	100.00	3	83.33	100.00	-165	31.25	100.00	-228	41.67	97.92	-211	29.17
[14]	441345	100.00	-2	39.58	100.00	-179	39.58	100.00	-248	47.92	97.92	-227	50.00
[15]	475137	100.00	-8	56.25	100.00	-191	45.83	100.00	-260	43.75	97.92	-240	50.00
[16]	508929	100.00	-14	56.25	100.00	-209	47.92	95.83	-280	58.33	95.83	-258	43.75
[17]	542721	100.00	0	0.00	100.00	0	0.00	100.00	-329	56.25	100.00	416	58.33
[18]	576513	100.00	-30	70.83	100.00	0	0.00	100.00	0	0.00	97.92	0	0.00
[19]	610305	100.00	-39	60.42	100.00	-257	54.17	100.00	-332	45.83	97.92	-307	52.08
[20]	644097	100.00	-47	43.75	100.00	-269	31.25	100.00	-348	29.17	97.92	-325	58.33
[21]	677889	100.00	-58	47.92	100.00	-285	50.00	100.00	-371	56.25	100.00	-340	47.92
[22]	711681	97.92	-67	18.75	97.92	-299	56.25	97.92	-390	45.83	95.83	322	52.08
[23]	745473	97.92	-78	33.33	97.92	-315	47.92	97.92	-399	60.42	97.92	-367	35.42
[24]	779265	100.00	-89	10.42	100.00	-333	33.33	100.00	-417	33.33	100.00	-378	39.58
[25]	813057	100.00	-101	31.25	100.00	-347	37.50	100.00	-433	45.83	100.00	-392	39.58
[26]	846849	100.00	-112	47.92	100.00	-363	25.00	100.00	0	0.00	100.00	-404	20.83
[27]	880641	100.00	-122	41.67	100.00	-379	45.83	100.00	-467	54.17	100.00	-417	39.58
[28]	914433	100.00	-133	18.75	100.00	-395	39.58	100.00	-483	50.00	100.00	-430	41.67
[29]	948225	100.00	-145	43.75	100.00	-410	39.58	100.00	-500	43.75	97.92	-442	35.42
[30]	982017	100.00	-159	43.75	100.00	-426	45.83	100.00	-521	45.83	100.00	-455	29.17
[31]	1015809	100.00	-170	31.25	100.00	-441	39.58	100.00	-539	45.83	100.00	-467	47.92
[32]	1049601	100.00	-184	37.50	100.00	-457	33.33	100.00	-556	35.42	100.00	-481	50.00
[33]	1083393	100.00	0	0.00	100.00	-472	47.92	100.00	-575	54.17	97.92	0	0.00
[34]	1117185	100.00	-210	37.50	97.92	-490	35.42	97.92	-594	70.83	97.92	-508	68.75
[35]	1150977	100.00	0	0.00	97.92	0	0.00	97.92	0	0.00	97.92	0	0.00
Watermark detected:				32			32			31			32
average % bits recovered				60.94			41.99			47.45			45.96

Table 8. Analog Tape system performance

8.1.5 FM STEREO RADIO

Left channel.

	Original Position	FM ST. LEFT -2 dB			FM ST. LEFT -4 dB			FM ST. LEFT -6 dB			FM ST. LEFT -8 dB		
		%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	3	95.83	100.00	4	93.75	100.00	4	100.00	100.00	4	93.75
[2]	35841	100.00	7	87.50	100.00	9	93.75	100.00	8	87.50	100.00	8	79.17
[3]	69633	100.00	11	97.92	97.92	12	83.33	97.92	11	89.58	100.00	12	85.42
[4]	103425	100.00	15	95.83	100.00	16	93.75	100.00	15	91.67	100.00	16	93.75
[5]	137217	100.00	19	95.83	100.00	20	87.50	100.00	19	75.00	100.00	19	66.67
[6]	171009	97.92	22	87.50	100.00	24	95.83	97.92	23	87.50	97.92	23	81.25
[7]	204801	97.92	26	77.08	97.92	27	75.00	97.92	27	85.42	93.75	27	79.17
[8]	238593	100.00	30	93.75	100.00	31	87.50	100.00	30	85.42	100.00	31	85.42
[9]	272385	100.00	34	100.00	100.00	35	97.92	100.00	34	97.92	100.00	35	97.92
[10]	306177	100.00	38	91.67	100.00	39	89.58	100.00	38	83.33	97.92	38	79.17
[11]	339969	95.83	41	77.08	95.83	43	81.25	93.75	42	81.25	89.58	42	70.83
[12]	373761	100.00	45	87.50	100.00	46	83.33	100.00	46	95.83	95.83	46	83.33
[13]	407553	100.00	49	89.58	100.00	50	83.33	100.00	49	77.08	97.92	50	89.58
[14]	441345	100.00	53	95.83	100.00	54	91.67	100.00	53	85.42	97.92	54	83.33
[15]	475137	100.00	56	77.08	100.00	58	89.58	100.00	57	83.33	97.92	57	77.08
[16]	508929	100.00	60	83.33	100.00	62	93.75	95.83	61	83.33	95.83	61	77.08
[17]	542721	100.00	64	95.83	100.00	66	95.83	100.00	65	95.83	100.00	65	81.25
[18]	576513	100.00	68	93.75	100.00	70	93.75	100.00	69	93.75	97.92	69	87.50
[19]	610305	100.00	72	93.75	100.00	73	93.75	100.00	72	79.17	97.92	73	87.50
[20]	644097	100.00	76	97.92	100.00	77	93.75	100.00	76	89.58	97.92	77	91.67
[21]	677889	100.00	79	87.50	100.00	81	87.50	100.00	80	89.58	100.00	80	70.83
[22]	711681	97.92	83	87.50	97.92	84	83.33	97.92	84	83.33	95.83	84	79.17
[23]	745473	97.92	87	85.42	97.92	88	81.25	97.92	88	85.42	97.92	88	87.50
[24]	779265	100.00	91	97.92	100.00	92	95.83	100.00	92	95.83	100.00	92	91.67
[25]	813057	100.00	95	100.00	100.00	96	97.92	100.00	96	100.00	100.00	0	0.00
[26]	846849	100.00	98	97.92	100.00	100	100.00	100.00	99	97.92	100.00	99	93.75
[27]	880641	100.00	102	85.42	100.00	104	97.92	100.00	103	91.67	100.00	103	85.42
[28]	914433	100.00	106	97.92	100.00	108	97.92	100.00	107	100.00	100.00	107	91.67
[29]	948225	100.00	110	95.83	100.00	111	87.50	100.00	111	91.67	97.92	111	87.50
[30]	982017	100.00	114	100.00	100.00	115	100.00	100.00	114	97.92	100.00	115	93.75
[31]	1015809	100.00	117	89.58	100.00	119	97.92	100.00	118	85.42	100.00	119	93.75
[32]	1049601	100.00	121	95.83	100.00	123	100.00	100.00	122	91.67	100.00	122	89.58
[33]	1083393	100.00	125	97.92	100.00	127	93.75	100.00	126	91.67	97.92	126	77.08
[34]	1117185	100.00	129	87.50	97.92	130	85.42	97.92	130	89.58	97.92	130	75.00
[35]	1150977	100.00	0	0.00	97.92	134	64.58	97.92	0	0.00	97.92	134	70.83
Watermark detected:				34			35			34			34
average % bits recovered				91.79			90.54			89.40			84.07

Table 9. FM Stereo (left channel) system performance

Right channel.

	Original	FM ST.RIGHT -2 dB			FM ST.RIGHT -4 dB			FM ST.RIGHT -6 dB			FM ST.RIGHT -8 dB		
	Position	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	3	93.75	100.00	3	87.50	100.00	5	91.67	100.00	4	79.17
[2]	35841	100.00	5	58.33	100.00	8	95.83	97.92	7	62.50	97.92	9	87.50
[3]	69633	100.00	11	91.67	100.00	11	81.25	100.00	12	87.50	97.92	12	79.17
[4]	103425	100.00	15	95.83	100.00	15	93.75	100.00	16	79.17	97.92	15	75.00
[5]	137217	100.00	19	97.92	100.00	19	85.42	100.00	20	83.33	100.00	20	79.17
[6]	171009	100.00	22	85.42	100.00	23	85.42	100.00	24	89.58	100.00	23	83.33
[7]	204801	97.92	26	75.00	95.83	26	77.08	93.75	28	77.08	91.67	27	75.00
[8]	238593	100.00	30	100.00	100.00	30	91.67	100.00	31	89.58	100.00	35	77.08
[9]	272385	100.00	34	95.83	100.00	34	93.75	100.00	35	83.33	100.00	34	81.25
[10]	306177	100.00	38	95.83	100.00	38	95.83	100.00	39	89.58	97.92	38	75.00
[11]	339969	100.00	41	91.67	97.92	42	87.50	100.00	43	83.33	100.00	42	75.00
[12]	373761	93.75	45	66.67	93.75	45	72.92	91.67	47	75.00	91.67	46	62.50
[13]	407553	97.92	49	89.58	97.92	49	81.25	97.92	51	89.58	97.92	50	77.08
[14]	441345	100.00	53	91.67	100.00	53	89.58	100.00	54	89.58	97.92	54	91.67
[15]	475137	100.00	57	91.67	100.00	57	95.83	100.00	58	85.42	100.00	57	81.25
[16]	508929	100.00	60	81.25	97.92	61	85.42	97.92	62	83.33	97.92	61	66.67
[17]	542721	100.00	64	72.92	100.00	64	85.42	97.92	66	72.92	97.92	65	68.75
[18]	576513	100.00	68	87.50	100.00	68	79.17	100.00	70	91.67	97.92	69	83.33
[19]	610305	100.00	72	89.58	100.00	72	89.58	100.00	73	89.58	100.00	73	87.50
[20]	644097	100.00	76	97.92	100.00	76	93.75	100.00	77	89.58	100.00	76	81.25
[21]	677889	100.00	79	85.42	100.00	80	93.75	100.00	81	89.58	100.00	80	81.25
[22]	711681	100.00	83	95.83	97.92	84	85.42	97.92	85	87.50	100.00	84	75.00
[23]	745473	100.00	87	91.67	97.92	82	37.50	100.00	82	47.92	97.92	88	83.33
[24]	779265	100.00	91	95.83	100.00	91	95.83	100.00	93	97.92	100.00	92	93.75
[25]	813057	100.00	95	100.00	100.00	95	100.00	100.00	96	97.92	100.00	95	95.83
[26]	846849	100.00	98	93.75	100.00	99	100.00	100.00	100	95.83	100.00	99	91.67
[27]	880641	100.00	102	89.58	100.00	103	97.92	100.00	104	91.67	100.00	103	83.33
[28]	914433	100.00	106	100.00	100.00	106	93.75	100.00	108	97.92	100.00	107	93.75
[29]	948225	100.00	110	100.00	100.00	110	93.75	100.00	112	100.00	100.00	111	87.50
[30]	982017	100.00	114	100.00	100.00	114	100.00	100.00	115	91.67	100.00	115	97.92
[31]	1015809	100.00	118	97.92	100.00	118	87.50	97.92	119	89.58	97.92	119	91.67
[32]	1049601	100.00	121	93.75	100.00	122	93.75	100.00	123	93.75	100.00	122	87.50
[33]	1083393	100.00	125	89.58	100.00	126	91.67	100.00	127	91.67	100.00	126	87.50
[34]	1117185	100.00	129	93.75	100.00	129	83.33	100.00	131	93.75	97.92	130	79.17
[35]	1150977	100.00	0	0.00	100.00	133	64.58	100.00	0	0.00	100.00	134	68.75
Watermark detected:				34			35			34			35
average % bits recovered				90.50			87.62			86.76			81.85

Table 10. FM Stereo (right channel) system performance

8.1.6 FM RADIO MONO

Original		FM MONO -2 dB			FM MONO -4 dB			FM MONO -6 dB			FM MONO -8dB		
Position		%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	4	100.00	100.00	3	93.75	100.00	4	100.00	100.00	3	85.42
[2]	35841	100.00	0	0.00	100.00	7	79.17	100.00	8	89.58	100.00	7	68.75
[3]	69633	100.00	11	89.58	97.92	11	91.67	97.92	11	77.08	100.00	11	85.42
[4]	103425	100.00	15	93.75	100.00	15	95.83	100.00	15	91.67	100.00	15	95.83
[5]	137217	100.00	19	89.58	100.00	19	93.75	100.00	19	85.42	100.00	19	72.92
[6]	171009	97.92	23	93.75	100.00	22	83.33	97.92	23	93.75	97.92	23	91.67
[7]	204801	97.92	27	91.67	97.92	26	81.25	97.92	27	81.25	93.75	26	77.08
[8]	238593	100.00	30	77.08	100.00	30	97.92	100.00	30	85.42	100.00	30	91.67
[9]	272385	100.00	34	100.00	100.00	34	100.00	100.00	34	95.83	100.00	34	100.00
[10]	306177	100.00	38	91.67	100.00	37	81.25	100.00	38	85.42	97.92	38	83.33
[11]	339969	95.83	42	87.50	95.83	41	79.17	93.75	42	79.17	89.58	41	79.17
[12]	373761	100.00	46	91.67	100.00	45	77.08	100.00	46	77.08	95.83	45	83.33
[13]	407553	100.00	49	83.33	100.00	49	91.67	100.00	49	85.42	97.92	49	91.67
[14]	441345	100.00	53	81.25	100.00	53	77.08	100.00	53	72.92	97.92	53	81.25
[15]	475137	100.00	57	87.50	100.00	57	81.25	100.00	57	75.00	97.92	57	83.33
[16]	508929	100.00	61	91.67	100.00	60	85.42	95.83	61	89.58	95.83	60	66.67
[17]	542721	100.00	67	95.83	100.00	64	89.58	100.00	67	91.67	100.00	64	91.67
[18]	576513	100.00	0	0.00	100.00	68	87.50	100.00	69	87.50	97.92	68	77.08
[19]	610305	100.00	72	83.33	100.00	72	87.50	100.00	72	72.92	97.92	72	87.50
[20]	644097	100.00	76	91.67	100.00	76	89.58	100.00	76	87.50	97.92	76	93.75
[21]	677889	100.00	80	97.92	100.00	79	81.25	100.00	80	79.17	100.00	79	68.75
[22]	711681	97.92	84	91.67	97.92	83	83.33	97.92	84	79.17	95.83	83	70.83
[23]	745473	97.92	88	89.58	97.92	87	85.42	97.92	88	93.75	97.92	87	79.17
[24]	779265	100.00	91	85.42	100.00	91	95.83	100.00	91	89.58	100.00	91	93.75
[25]	813057	100.00	95	97.92	100.00	95	100.00	100.00	95	97.92	100.00	95	100.00
[26]	846849	100.00	99	100.00	100.00	99	100.00	100.00	99	95.83	100.00	99	91.67
[27]	880641	100.00	103	93.75	100.00	103	95.83	100.00	103	91.67	100.00	103	93.75
[28]	914433	100.00	107	97.92	100.00	106	93.75	100.00	107	97.92	100.00	107	100.00
[29]	948225	100.00	111	97.92	100.00	110	95.83	100.00	111	97.92	97.92	110	85.42
[30]	982017	100.00	114	93.75	100.00	114	100.00	100.00	115	100.00	100.00	114	97.92
[31]	1015809	100.00	118	91.67	100.00	118	97.92	100.00	118	85.42	100.00	118	91.67
[32]	1049601	100.00	122	97.92	100.00	121	85.42	100.00	122	91.67	100.00	121	87.50
[33]	1083393	100.00	126	93.75	100.00	125	89.58	100.00	126	87.50	97.92	126	85.42
[34]	1117185	100.00	130	95.83	97.92	129	83.33	97.92	130	81.25	97.92	129	77.08
[35]	1150977	100.00	0	0.00	97.92	133	79.17	97.92	0	0.00	97.92	133	72.92
Watermark detected:				32			35			34			35
average % bits recovered				92.06			88.87			87.44			85.24

Table 11. FM MONO system performance

8.1.7 FM RADIO MONO WEAK SIGNAL

	Original	FM MONO WEAK -2dB			FM MONO WEAK -4dB			FM MONO WEAK -6dB			FM MONO WEAK -8dB		
	Position	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00
[2]	35841	100.00	6	43.75	100.00	6	52.08	100.00	6	58.33	100.00	0	0.00
[3]	69633	100.00	11	83.33	97.92	10	70.83	97.92	10	75.00	100.00	11	77.08
[4]	103425	100.00	14	58.33	100.00	15	93.75	100.00	14	79.17	100.00	15	81.25
[5]	137217	100.00	19	87.50	100.00	19	93.75	100.00	18	79.17	100.00	19	72.92
[6]	171009	97.92	22	64.58	100.00	22	77.08	97.92	22	77.08	97.92	22	62.50
[7]	204801	97.92	26	62.50	97.92	26	70.83	97.92	26	75.00	93.75	26	56.25
[8]	238593	100.00	30	89.58	100.00	29	79.17	100.00	29	72.92	100.00	30	70.83
[9]	272385	100.00	34	100.00	100.00	38	93.75	100.00	33	89.58	100.00	34	91.67
[10]	306177	100.00	37	66.67	100.00	37	75.00	100.00	36	56.25	97.92	40	85.42
[11]	339969	95.83	41	77.08	95.83	41	62.50	93.75	41	60.42	89.58	42	70.83
[12]	373761	100.00	45	68.75	100.00	46	79.17	100.00	44	66.67	95.83	46	75.00
[13]	407553	100.00	49	81.25	100.00	49	79.17	100.00	48	64.58	97.92	49	64.58
[14]	441345	100.00	53	75.00	100.00	53	85.42	100.00	52	64.58	97.92	53	58.33
[15]	475137	100.00	57	89.58	100.00	57	81.25	100.00	56	79.17	97.92	56	62.50
[16]	508929	100.00	60	62.50	100.00	60	70.83	95.83	60	81.25	95.83	61	85.42
[17]	542721	100.00	64	85.42	100.00	64	89.58	100.00	64	95.83	100.00	64	81.25
[18]	576513	100.00	0	0.00	100.00	67	64.58	100.00	67	72.92	97.92	69	75.00
[19]	610305	100.00	72	81.25	100.00	71	64.58	100.00	71	56.25	97.92	72	75.00
[20]	644097	100.00	76	83.33	100.00	75	83.33	100.00	75	87.50	97.92	75	75.00
[21]	677889	100.00	79	77.08	100.00	79	77.08	100.00	79	66.67	100.00	82	85.42
[22]	711681	97.92	83	56.25	97.92	83	75.00	97.92	83	77.08	95.83	83	58.33
[23]	745473	97.92	87	81.25	97.92	87	75.00	97.92	86	60.42	97.92	89	81.25
[24]	779265	100.00	91	81.25	100.00	91	93.75	100.00	90	72.92	100.00	91	77.08
[25]	813057	100.00	95	95.83	100.00	94	77.08	100.00	94	83.33	100.00	95	83.33
[26]	846849	100.00	98	77.08	100.00	99	93.75	100.00	98	89.58	100.00	99	83.33
[27]	880641	100.00	102	83.33	100.00	102	89.58	100.00	102	91.67	100.00	102	66.67
[28]	914433	100.00	107	100.00	100.00	107	100.00	100.00	107	97.92	100.00	107	91.67
[29]	948225	100.00	110	79.17	100.00	109	64.58	100.00	109	60.42	97.92	110	62.50
[30]	982017	100.00	114	83.33	100.00	114	93.75	100.00	113	62.50	100.00	114	79.17
[31]	1015809	100.00	117	70.83	100.00	117	81.25	100.00	117	79.17	100.00	117	58.33
[32]	1049601	100.00	121	72.92	100.00	121	77.08	100.00	121	75.00	100.00	122	79.17
[33]	1083393	100.00	125	85.42	100.00	125	79.17	100.00	124	68.75	97.92	125	70.83
[34]	1117185	100.00	129	81.25	97.92	129	79.17	97.92	129	81.25	97.92	130	75.00
[35]	1150977	100.00	0	0.00	97.92	132	52.08	97.92	0	0.00	97.92	133	54.17
Watermark detected:				32			34			33			33
average % bits recovered				77.67			78.68			74.49			73.55

Table 12. FM MONO weak signal system performance

8.1.8 AM RADIO

	Original	AM MONO -2 dB			AM MONO -4 dB			AM MONO -6 dB			AM MONO -8dB		
	Position	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits	%orig.	Off.	%bits
[1]	2049	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00
[2]	35841	100.00	9	77.08	100.00	0	0.00	100.00	9	70.83	100.00	0	0.00
[3]	69633	100.00	12	66.67	97.92	14	81.25	97.92	12	52.08	100.00	11	43.75
[4]	103425	100.00	17	91.67	100.00	15	22.92	100.00	17	87.50	100.00	16	47.92
[5]	137217	100.00	21	95.83	100.00	21	87.50	100.00	22	89.58	100.00	20	50.00
[6]	171009	97.92	24	64.58	100.00	26	87.50	97.92	25	83.33	97.92	25	75.00
[7]	204801	97.92	29	77.08	97.92	28	62.50	97.92	30	81.25	93.75	29	60.42
[8]	238593	100.00	35	97.92	100.00	33	93.75	100.00	32	85.42	100.00	32	66.67
[9]	272385	100.00	33	6.25	100.00	37	97.92	100.00	0	0.00	100.00	41	91.67
[10]	306177	100.00	38	39.58	100.00	41	95.83	100.00	41	93.75	97.92	39	45.83
[11]	339969	95.83	43	66.67	95.83	44	75.00	93.75	45	66.67	89.58	45	75.00
[12]	373761	100.00	47	72.92	100.00	47	54.17	100.00	46	52.08	95.83	50	75.00
[13]	407553	100.00	52	89.58	100.00	47	14.58	100.00	53	95.83	97.92	54	87.50
[14]	441345	100.00	54	56.25	100.00	55	75.00	100.00	55	72.92	97.92	55	58.33
[15]	475137	100.00	58	77.08	100.00	58	58.33	100.00	58	52.08	97.92	60	81.25
[16]	508929	100.00	62	75.00	100.00	63	83.33	95.83	62	77.08	95.83	61	52.08
[17]	542721	100.00	0	0.00	100.00	36	62.50	100.00	0	0.00	100.00	28	52.08
[18]	576513	100.00	69	68.75	100.00	71	79.17	100.00	70	81.25	97.92	71	72.92
[19]	610305	100.00	74	89.58	100.00	74	81.25	100.00	77	93.75	97.92	73	56.25
[20]	644097	100.00	77	77.08	100.00	82	100.00	100.00	81	91.67	97.92	78	77.08
[21]	677889	100.00	80	39.58	100.00	82	72.92	100.00	80	41.67	100.00	82	77.08
[22]	711681	97.92	86	87.50	97.92	86	75.00	97.92	86	83.33	95.83	87	79.17
[23]	745473	97.92	90	83.33	97.92	91	89.58	97.92	88	41.67	97.92	92	81.25
[24]	779265	100.00	91	27.08	100.00	93	64.58	100.00	94	95.83	100.00	95	91.67
[25]	813057	100.00	96	68.75	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00
[26]	846849	100.00	104	97.92	100.00	101	81.25	100.00	103	97.92	100.00	103	95.83
[27]	880641	100.00	103	45.83	100.00	106	91.67	100.00	104	77.08	100.00	104	52.08
[28]	914433	100.00	108	83.33	100.00	109	93.75	100.00	108	77.08	100.00	109	81.25
[29]	948225	100.00	112	83.33	100.00	111	41.67	100.00	113	93.75	97.92	113	75.00
[30]	982017	100.00	117	100.00	100.00	118	100.00	100.00	118	100.00	100.00	117	91.67
[31]	1015809	100.00	119	70.83	100.00	121	93.75	100.00	118	37.50	100.00	121	87.50
[32]	1049601	100.00	123	75.00	100.00	124	79.17	100.00	123	64.58	100.00	125	91.67
[33]	1083393	100.00	125	39.58	100.00	126	43.75	100.00	124	33.33	97.92	0	0.00
[34]	1117185	100.00	130	60.42	97.92	132	72.92	97.92	130	56.25	97.92	133	77.08
[35]	1150977	100.00	137	45.83	97.92	0	0.00	97.92	0	0.00	97.92	0	0.00
Watermark detected:				32			31			30			30
average % bits recovered				71.81			74.60			74.24			71.67

Table 13. AM radio system performance