May 30, 2000
RT0367
Multimedia; Rights Management; Security 16 pages

Research Report

Statistical Model and Experiment of Reliability in Detecting Multi-bit Watermark

Taiga Nakamura, Ryuki Tachibana, and Seiji Kobayashi

IBM Research, Tokyo Research Laboratory
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Statistical Model and Experiment of Reliability in Detecting Multi-bit Watermark

Taiga Nakamura * Ryuki Tachibana * Seiji Kobayashi *

taiga@jp.ibm.com ryuki@jp.ibm.com kobayas@jp.ibm.com

Abstract

Digital watermarking is widely used as an identification and control mechanism for rights management of digital multimedia content. In this paper, on the basis of a statistical model, we offer a comprehensive scheme for computing bit error ratios, false negative error ratios, and false positive error ratios in detecting multiple bits of information from watermarked and unmarked content. We apply this scheme to the analysis of an audio watermarking system we developed and compare predicted and observed error ratios to show its validity and effectiveness experimentally. In our proposed scheme, the detection of multiple bits is modeled as a statistical decision with threshold parameters in the multi-dimensional watermark space, in which each component of the location corresponds to the normalized sum of the detected watermark pattern for each bit. Error ratios are computed as a function of the threshold parameters from the probability distribution in the watermark space. Further, from experiments on our audio watermarking system, we discuss how the deterioration of watermarked content changes the distribution in the multi-dimensional watermark space.

1 Introduction

Digital watermarking is a technique to conceal additional information within digital content such as images, motion pictures, and audio data. The information embedded by robust watermarking has the advantage of being extractable even after the content is converted to a new form or suffers from deterioration.

Most applications of watermarking such as copyright management and identification of content require the detected information to be reliable. For an example, it was proposed in SDMI and in the DVD CPTWG that an audio player and a video player would automatically stop playback of copied content as an unauthorized reproduction based on the detection of watermarks. However, if the player detects a false watermark from unmarked content and stops playback, the user suffers from a malfunction of the player.

Any reliable watermarking method should satisfy the followings requirements.

- It detects correct information from content which has been watermarked.
- It does not detect false watermark from content which has not been watermarked.

The purposes of this paper are to offer a general framework for computing the reliability of the detected multi-bit watermark, to apply it to the analysis of an audio watermarking system we developed, and to discuss how the deterioration of the content affects bit error. We consider the following three types of detection errors:

- False Positive Error (FPE)
  A watermark is extracted from unmarked content. FPE is sometimes called "false alarm."
- False Negative Error (FNE)
  No watermark is extracted from watermarked content. FNE is sometimes called "missed detection."
- Bit Error (BE)
  Wrong information, which differs from the embedded information, is extracted from watermarked content.

The so-called survivability is the reliability of watermark when the content has suffered deterioration and is given by the following equation:

$$\text{Survivability} = 1 - \text{FNE} - \text{BE} \tag{1}$$

The previous work [5, 6, 8, 4, 10] on the reliability of watermarks was concerned mainly with single-bit watermarking and FPE. However, BE is more important than FPE in the detection of multi-bit watermarks because the probability that the extracted information contains at least one bit error increases according to the number of bits in the information. Other research [13, 14] gave the survivability data for a specific robustness

*IBM Tokyo Research Laboratory, 1623-14 Shimotsuruma, Yamato-shi, Kanagawa-ken, Japan, 242-8502

test but did not discuss how BE increases or decreases according to the deterioration suffered by watermarked content.

In this paper we discuss BE in conjunction with the deterioration of watermarked content. We also show that correlations among multiple extracted values of bits are experimentally observed and affect BE computations.

In Section 2, we describe a framework for computing the reliability of a detected multi-bit watermark. First, we model the watermark detection process as a computation of the watermark strength and a statistical decision regarding it. The computed watermark strength forms a watermark vector in the multi-dimensional watermark space which corresponds to multi-bit information. The deterioration of watermarked content causes a change in watermark vector. We model such a change as a transition of moments of the probability distributions according to which the watermark strengths are distributed. We call such a transition of the moment a "deterioration trajectory." Using the deterioration trajectory, we can estimate error ratios for any given threshold and for any degree of deterioration by interpolating experimental data. We show by the simulation that:

- The bit error ratio becomes smaller when there is correlation in the components of the watermark vector.
- The bit error ratio is not necessarily a monotonically increasing function with respect to the degree of deterioration and in some cases, it has a peak at the midpoint of deterioration (halfway between a perfect watermark and complete loss of the watermark.)

In Section 3, we describe briefly the audio watermarking system we developed and experimentally show by use of various types of content that there is a correlation between the components of the watermark vector and that the deterioration of the content yields a relatively simple trajectory of the moments of the distributions. We observed that the variances of watermark vector distribution first increased and then gradually converged on the value one according to the degree of deterioration of the watermarked content. The value one corresponds to the distribution of watermark vector of unmarked content.

2 Model of watermark detection

In this section we model the detection of multi-bit watermarks within content as a statistical decision in the multi-dimensional watermark space and discuss the deterioration of the watermark as the distortion of the distribution of embedded watermarks in the watermark space.

2.1 Watermark detection

Watermark detection is a process to extract bit information from content data. Whether the content is watermarked or not is judged by comparing the value of the detected watermark's strength with a threshold. If the watermark value does not exceed the threshold, the content is regarded as unmarked. Otherwise, its bit values are determined.

The detection procedure of multi-bit watermarks is done in four steps:

1. Take a portion of the content as a detection unit. This is called a content chunk, denoted by $\boldsymbol{c}$ in this paper.
2. Calculate the value of a watermark vector $\boldsymbol{w} = [w_1, w_2, \ldots, w_n]^T$, from the content chunk $\boldsymbol{c}$ and its correlation with a detection pattern. Here T denotes the transpose of a vector. Each element of the vector, w_i, corresponds to the watermark strength of the ith bit b_i. We write the detection function $\mathcal{D}$ as:

   $$\boldsymbol{w} = \mathcal{D}(\boldsymbol{c}) \tag{2}$$

   The watermark space Ω is an n-dimensional space that contains all possible values of $\boldsymbol{w}$.
3. Judge whether the content is watermarked or not, by checking whether the watermark vector $\boldsymbol{w}$ is contained in the region assigned to "no bits," Ω_{unmarked}:

   $$\Omega_{\text{unmarked}} = \{\boldsymbol{w} \in \Omega \mid (-\xi_1 \le w_1 \le \xi_1) \vee \ldots \vee (-\xi_n \le w_n \le \xi_n)\} \tag{3}$$

   If $\boldsymbol{w}$ is within Ω_{unmarked}, the content is judged as not being watermarked. Otherwise, its bit values are decided in the next step.
4. Determine the bit values $\boldsymbol{b} = [b_1 b_2 \ldots b_n]$, according to the partition of the watermark space Ω. A simple partition is defined using plane thresholds: The region $\Omega_{\boldsymbol{b}}$ that corresponds to the bit $\boldsymbol{b} = [b_1 b_2 \ldots b_n]$, is defined as:

   $$\Omega_{\boldsymbol{b}} = \left\{ \boldsymbol{w} \in \Omega \,\middle|\, \begin{pmatrix} w_1 > +\xi_1 \text{ if } b_1 = 1 \\ w_1 < -\xi_1 \text{ if } b_1 = 0 \end{pmatrix} \wedge \ldots \wedge \begin{pmatrix} w_n > +\xi_n \text{ if } b_n = 1 \\ w_n < -\xi_n \text{ if } b_n = 0 \end{pmatrix} \right\} \tag{4}$$

   where x_i $(i = 1, 2, \ldots, n)$ denotes the position of the ith threshold.

2.2 Model of watermark distribution

Here we assume that the watermark vector $\boldsymbol{w}$ follows some distribution function $\phi(\boldsymbol{w})$ according to whether the content has been watermarked or not, in order to compute the detection error ratio. In principle, this distribution function can be obtained by collecting the frequency of $\boldsymbol{w}_\lambda = \mathcal{D}(\boldsymbol{c}_\lambda)$ in the watermark space Ω, with respect to a large number of chunk samples $\boldsymbol{c}_\lambda$.

Watermark vectors from any unmarked content are assumed to follow a single distribution. The distribution of the watermark vector from unmarked content, denoted by $\phi_{\text{unmarked}}(\boldsymbol{w})$, can be approximated by a standard normal distribution:

$$\phi_{\text{unmarked}}(\boldsymbol{w}) = \phi_{\text{normal}}(\boldsymbol{w}, 0, I) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}\right] \tag{5}$$

using an appropriate normalization of $\boldsymbol{w}$[17].

On the other hand, the distribution of the watermark vector from watermarked content, denoted by $\phi_{\text{watermarked}}(\boldsymbol{w})$, shows more complicated shape, especially when the content is post-processed. Here we approximate $\phi_{\text{watermarked}}(\boldsymbol{w})$ by an n-dimensional normal distribution with correlation, by counting up to the second order of moment of the distribution:

$$\phi_{\text{watermarked}}(\boldsymbol{w}) = \phi_{\text{normal}}(\boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})\right] \tag{6}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the variance-covariance matrix of the watermark vector $\boldsymbol{w}$ respectively. This approximation means that the moments higher than the second order are ignored. More accurate models can be obtained by taking moments of higher order into account.

2.3 Formulation of errors

The probability of any particular bit value being extracted can be calculated from the distribution of the watermark vector, $\phi(\boldsymbol{w})$, and the definition of the partition of the watermark space.

The three types of errors mentioned in Section 1 are defined using the definitions above.

- The False Positive Error (FPE) ratio, P_{FPE}, is the possibility that a sample watermark vector obtained from unmarked content takes a value that is outside of the bit region assigned to "no bits":

$$P_{\text{FPE}} = 1 - \int_{\Omega_{\text{unmarked}}} \phi_{\text{unmarked}}(\boldsymbol{w}) d\boldsymbol{w} \tag{7}$$

- The False Negative Error (FNE) ratio, P_{FNE}, is the possibility that a sample watermark vector obtained from content in which a certain bit has been embedded takes a value that is contained in the bit region assigned to "no bits":

$$P_{\text{FNE}} = \int_{\Omega_{\text{unmarked}}} \phi_{\text{watermarked}}(\boldsymbol{w}) d\boldsymbol{w} \tag{8}$$

- The Bit Error (BE) ratio, P_{BE}, is the possibility that a sample watermark vector obtained from content in which a certain bit has been embedded takes a value that is contained in the bit regions assigned to any other bits:

$$P_{\text{BE}} = 1 - P_{\text{correct}} - P_{\text{FNE}} \tag{9}$$

  where P_{correct}, the correct detection ratio, is defined as:

$$P_{\text{correct}} = \int_{\Omega_{\text{correct}}} \phi_{\text{watermarked}}(\boldsymbol{w}) d\boldsymbol{w} \tag{10}$$

  Ω_{correct} denotes the region that corresponds to the bit data that is to be extracted in successful detection.

The good threshold is the parameter that satisfies

$$P_{\text{FPE}} < \varepsilon_{\text{FPE}}, \quad P_{\text{BE}} < \varepsilon_{\text{BE}} \tag{11}$$

where ε_{FPE} and ε_{BE} denote the limits of the acceptable false positive error ratio and the bit error ratio, respectively. However, the distribution $\phi_{\text{watermarked}}(\boldsymbol{w})$ strongly depends on the actual implementation of the embedding process and the post-processing. Therefore we carefully examine the deterioration of the watermark in the next subsection.

Table 1: Models of deterioration trajectory

Model	$\mu(t)$	$\sigma^2(t)$	$\tau(t)$
1	$(1-t)\mu(0)+t\mu(1)$	$(1-t)\sigma^2(0)+t\tau^2(1)$	$(1-t)\tau(0)+t\tau(1)$
2	$(1-t)\mu(0)+t\mu(1)$	$(1-t)\sigma^2(0)+t\sigma^2(1)$	$(1-t)\tau(0)+t\tau(1)+\lambda_1 t(1-t)$
3	$(1-t)\mu(0)+t\mu(1)$	$(1-t)\sigma^2(0)+t\sigma^2(1)+\lambda_2 t(1-t)$	$(1-t)\tau(0)+t\tau(1)$
4	$(1-t)\mu(0)+t\mu(1)$	$(\sigma_r^2(t)+(n-1)\sigma_t^2(t))/n$	$(\sigma_r^2(t)-\sigma_t^2(t))/n$

*$\sigma_r^2(t)=(1-t)\sigma_r^2(0)+t\sigma_r^2(1)+\lambda_3 t(1-t)$, $\sigma_t^2(t)=(1-t)\sigma_t^2(0)+t\sigma_t^2(1)$

2.4 Deterioration of watermark

The content is exposed to various kinds of post-processing when it is used. The effects of post-processing on watermark extraction can be considered as a deformation of the distribution $\phi(\boldsymbol{w})$. We model the deterioration of watermark as the transition path from the initial state (the content just after the watermark was embedded) to the final state (after the content has been completely processed, severely damaging the watermark.) Using this model, we can calculate the error ratios that occur in bit extraction from post-processed content in any intermediate state by interpolating from experimental data.

When the initial condition and the transition path are symmetric for all bits, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are written as:

$$\boldsymbol{\mu}=[\mu,\mu,\ldots,\mu]^T,\quad \boldsymbol{\Sigma}=\begin{bmatrix}\sigma^2 & \tau & \cdots & \tau\\ \tau & \sigma^2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \tau\\ \tau & \cdots & \tau & \sigma^2\end{bmatrix} \tag{12}$$

where μ, σ^2, and τ are the mean, variance, and covariance of the distribution $\phi(\boldsymbol{w})$, respectively[16]. The distribution of the watermark vector is determined by specifying these three parameters:

$$[\mu,\sigma^2,\tau]=[\mu(t),\sigma^2(t),\tau(t)] \tag{13}$$

where t denotes a parameter of the degradation; $t=0$ and $t=1$ represent the initial state (just after embedding the watermark) and the final state (obliterated watermark, returning to the unmarked state), respectively.

Using these parameters, the deterioration trajectory is defined as a transition between them. Error ratios defined by the equations (7)–(9) can be computed at any point on the trajectory. Moreover, the greatest error ratio on the trajectory can be computed.

As examples, the error ratios along the following four deterioration trajectories have been computed. In all cases, the number of bits is $n=8$. The initial and the final state are set to $[\mu(0)=5,\sigma^2(0)=0,\tau(0)=0]$ and $[\mu(1)=0,\sigma^2(1)=1,\tau(1)=0]$, respectively.

Model 1 Linear model (Figure 1): This is the simplest trajectory, in which the initial and the final state are connected with a straight line. Since $\tau(0)=\tau(1)=0$, the distribution has no correlation in this model. This means that all components of the watermark vectors degrade independently. The BE ratio along this trajectory is shown in Figure 5. Note that the relation $(1-(\text{correct detection rate}))=P_{\text{FNE}}+P_{\text{BE}}$ holds. A one-bit error is the error when only a single bit out of n bits is incorrect, while a two-bit error is the error when exactly two bits are incorrect. As is shown, the ratio of two-bit errors is larger than the one-bit error ratio when the mean is near zero. This is because it follows a binomial distribution related to how many bits give the correct answer by chance.

Model 2 Linear model with quadratic correlation (Figure 2): It is natural to assume correlations exist for some degradation trajectories, because there are degradations which degrade only limited chunks of content severely. For this trajectory, the BE ratios are less than these of Model 1. The correlation causes an increase in the probability that the entire watermark vector degrades simultaneously and, consequently, false negative errors increase.

Model 3 Quadratic variance model without correlation (Figure 3): The implication of this model is that post-processing may have a slightly good effect as well as a bad effect on watermark if the degree of alteration is small. In such a situation, the variance of the extracted watermark vectors would increase, while their mean stays constant. This model emphasizes such a phenomenon. As Figure 7 shows, bit error occurs more frequently than in the cases of Model 1 and Model 2. This might be caused by wide and circular distributions without correlation.

Model 4 Quadratic variance model with correlation (Figure 4): This model is based on a supposition that the transverse component would not have a greater value than one during the transition. The BE ratio for this model is smaller than the BE ratio of Model 3 because of the effect of the correlation (Figure 8).

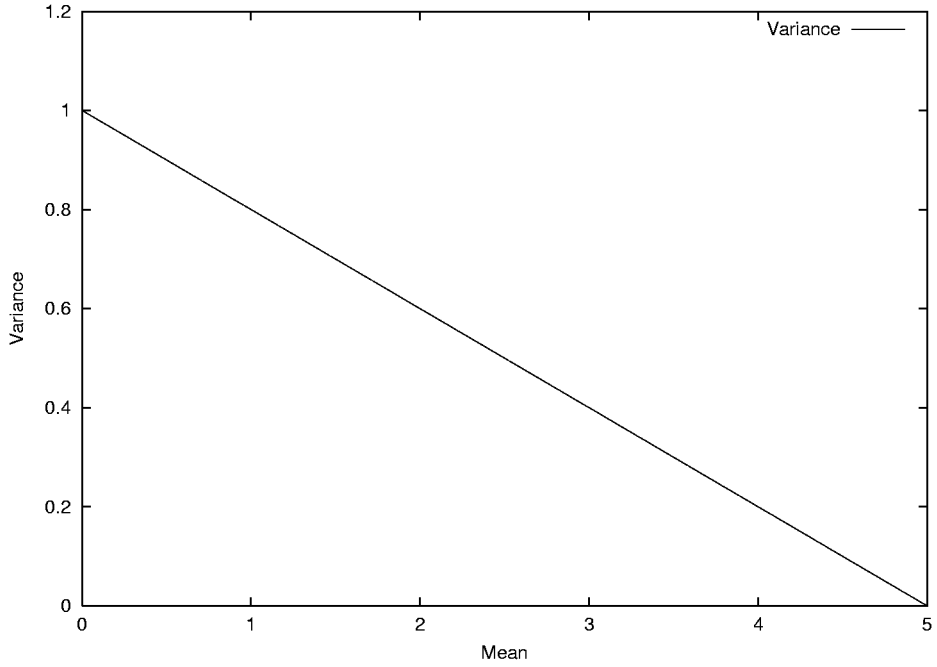

Figure 1: Deterioration trajectory model 1

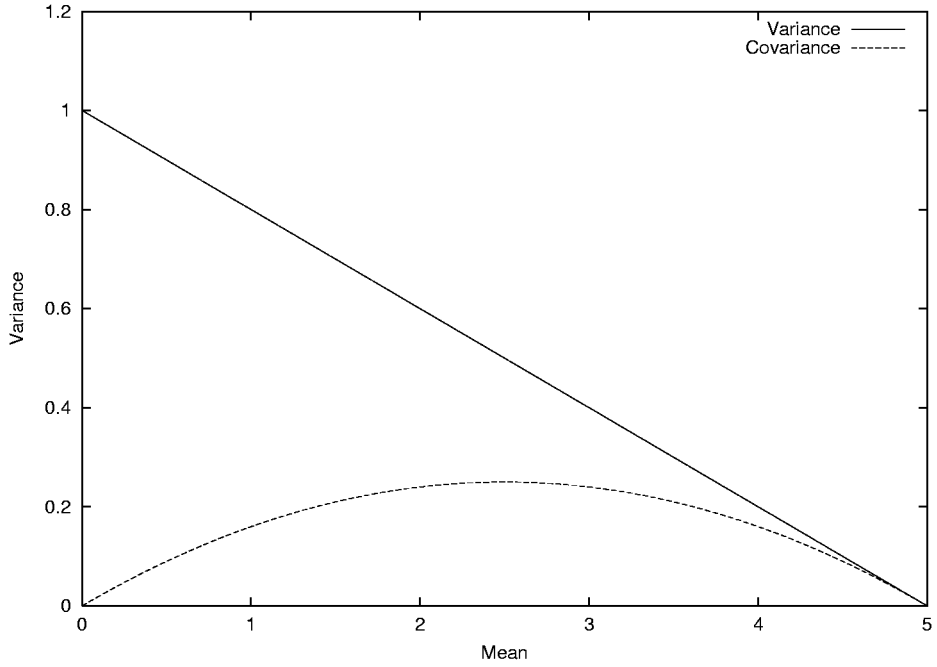

Figure 2: Deterioration trajectory model 2 with $\lambda_1 = 1$

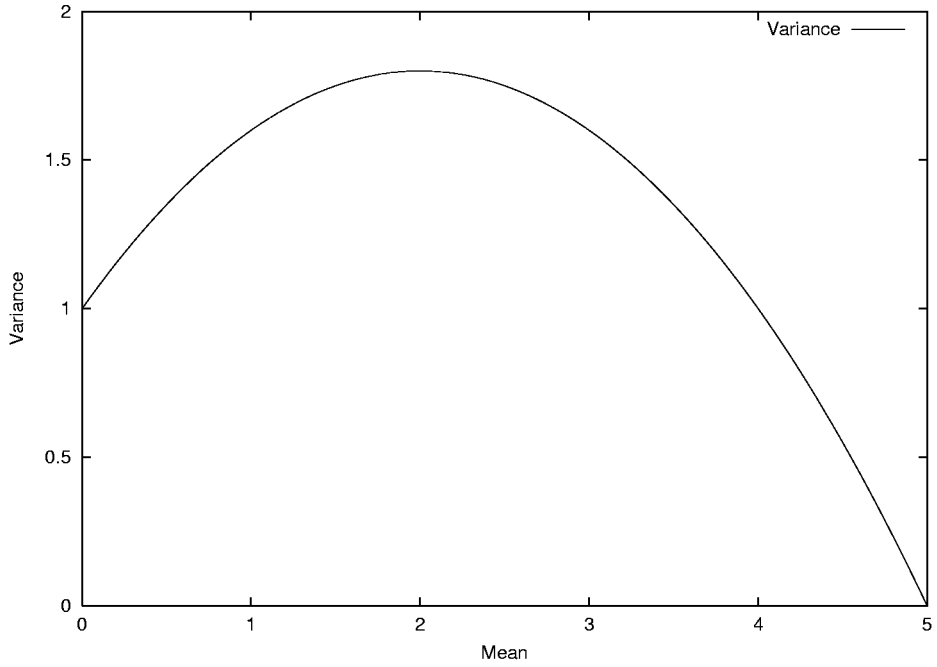

Figure 3: Deterioration trajectory model 3 with $\lambda_2 = 5$

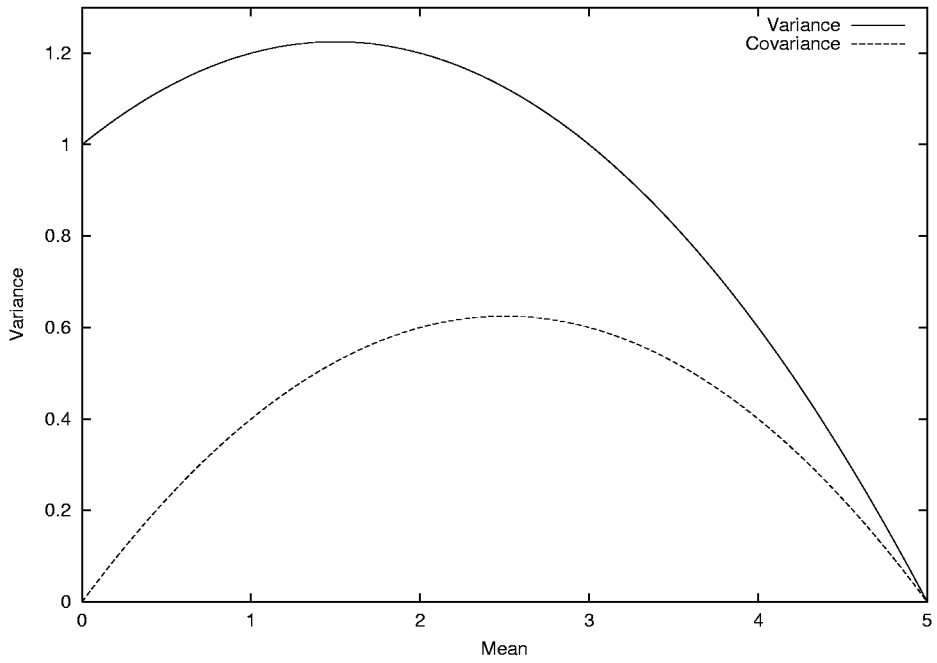

Figure 4: Deterioration trajectory model 4 with $\lambda_3 = 5$

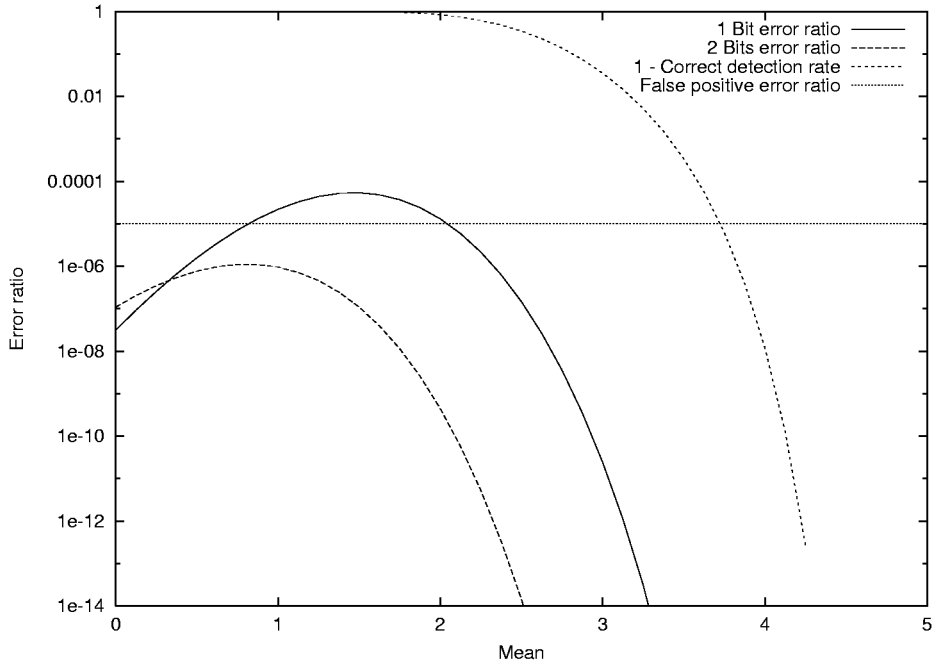

Figure 5: Bit error ratio for deterioration trajectory model 1

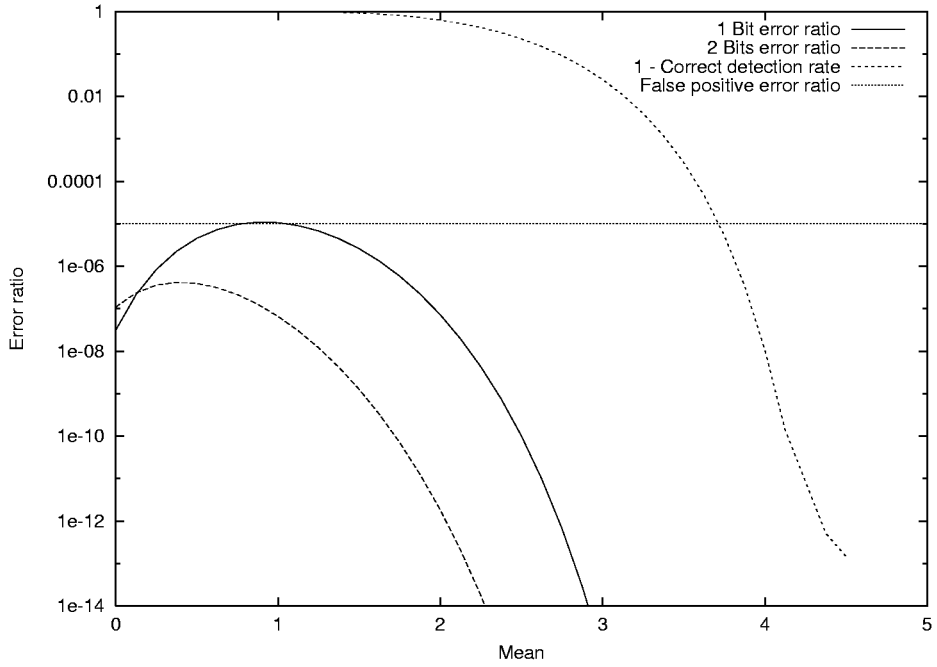

Figure 6: Bit error ratio for deterioration trajectory model 2

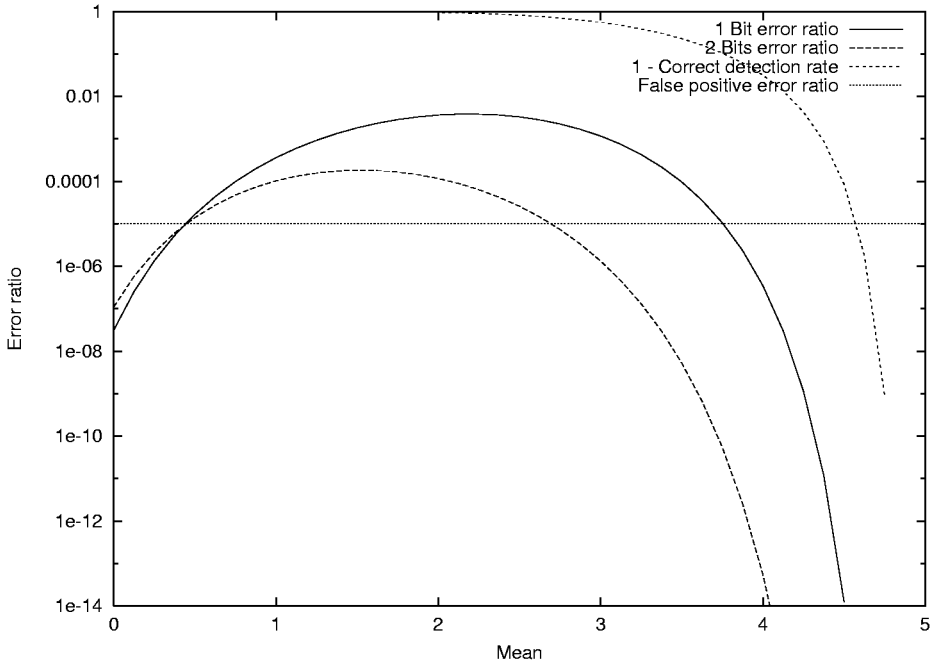

Figure 7: Bit error ratio for deterioration trajectory model 3

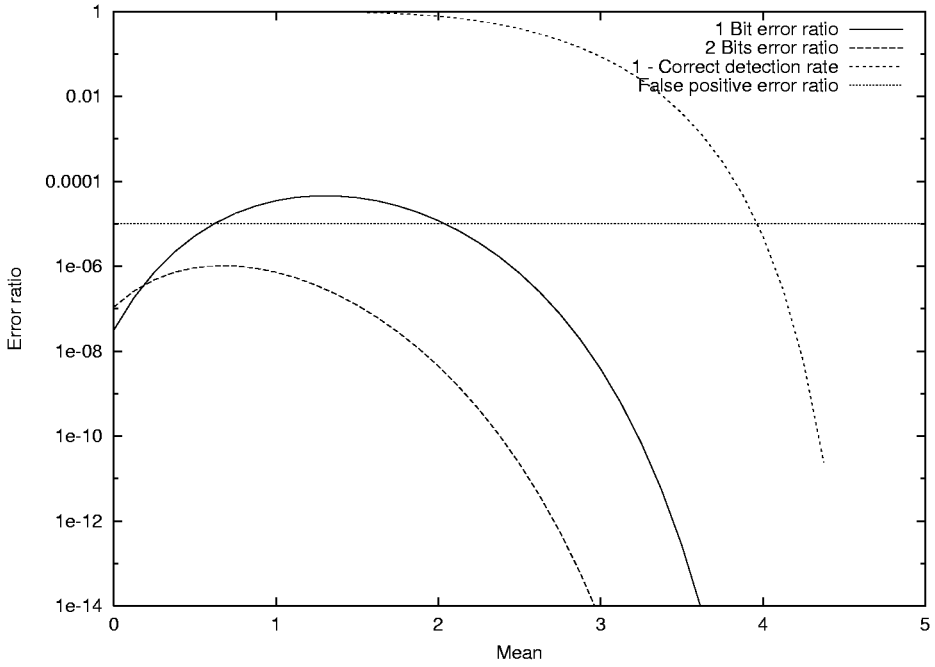

Figure 8: Bit error ratio for deterioration trajectory model 4

3 Experimental results

In this section, we verify the usefulness of our reliability model, by comparing the estimated error ratios with experimental results using the audio watermarking method we developed.

3.1 Audio watermarking method

Embedding

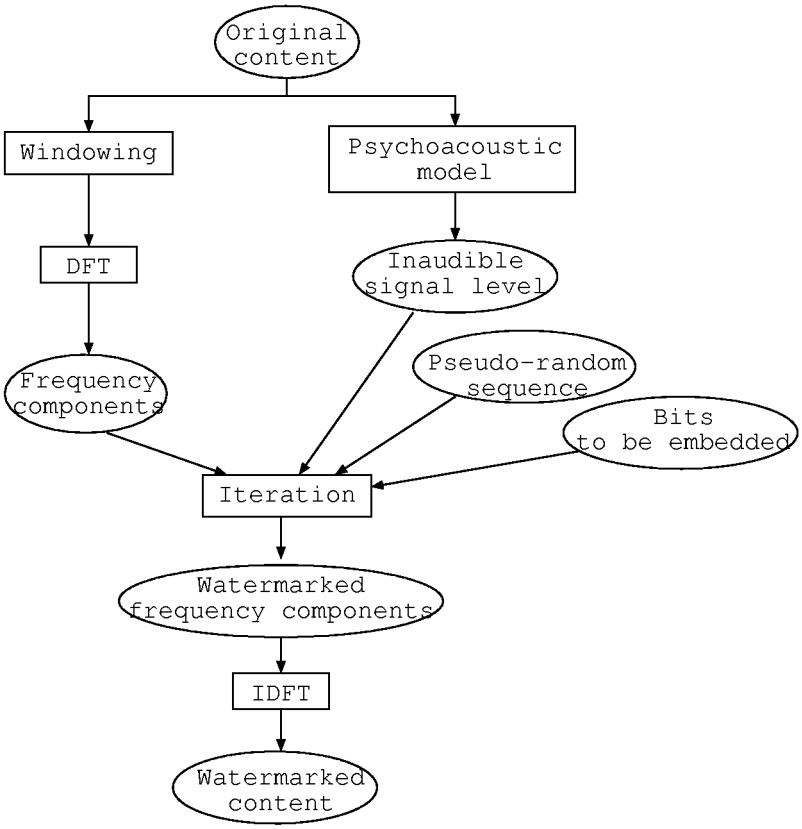

Figure 9: Flowchart of embedding

For embedding the watermark, the audio samples are divided into chunks, each of which receives embedded bit information without affecting the perceived sound quality. Figure 9 illustrates the embedding process. First, the inaudible signal level is computed in the frequency domain using the psychoacoustic model. The frequency components of the content chunk are also calculated by windowing and the Digital Fourier Transform (DFT). Next, watermarked frequency components are obtained within the inaudible signal level using a pseudo-random sequence and the bits to be embedded. Lastly, the watermarked content in the time domain is obtained by taking the inverse DFT of the watermarked frequency components. The iteration process is used to achieve the target watermark strength while remaining inaudible.

Detection

In detection, the same as for embedding, the frequency components of the content are calculated through windowing and DFT. Then the pseudo-random sequence used for the embedding is multiplied with the normalized frequency components, and the watermark vector of bits is detected. Bit decisions are made by comparing the vector with thresholds, which partitions the watermark space.

It should be noted that the position of each content chunk need to be determined. In the experiment, for simplicity we assumed that the positions of content chunks are already known. Knowledge of the original content is not required for this method to succeed in detecting the watermark.

In our experiment, two bits were embedded into and detected from each content chunk; the number of dimension of the watermark space is $n = 2$. We used three sources of audio data from different genres (jazz, pop music, and a soprano vocalist) as test samples. They are all approximately four minutes long, with CD quality (the sampling rate is 44.1kHz and the linearly quantization has 16-bit depth.)

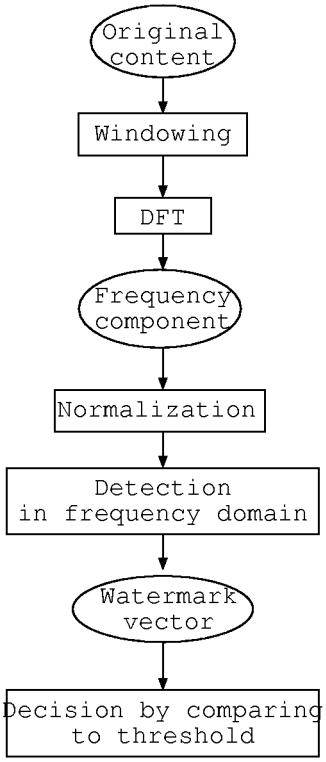

Figure 10: Flowchart of detection

At first, the distribution of watermark vectors obtained from unmarked test audio samples were observed to be compared with the assumption we made in Section 2, that the distribution should follow a normal distribution. Then for observing the degradation process of embedded watermarks caused by post processing watermark value and its reliability were observed while the level of processing was increased step by step.

3.2 Experimental results and discussion

Distribution of watermark vector of unmarked content

For the estimation of the FPE ratio, it is important that the distribution of the watermark vector obtained from unmarked content has predictable properties in the watermark space.

Figure 11 shows the histogram of detected watermark vectors for all three test audio samples. The theoretical curve, that is the probability density of normal distribution $N(0, 1^2)$ is also overlaid on the obtained histogram. Comparing with the theoretical curve, the distribution of watermark vectors detected from unmarked audio samples approximately follows the normal distribution as we had assumed. In this respect, once an appropriate threshold is set, the FPE ratio can be estimated based on the properties of this distribution model.

The distribution of watermark vector from unmarked content

Figure 12 plots the watermark vectors obtained from watermarked audio samples. In the experiment, the initial state of the watermark vector was set to be $(5.0, 5.0)$. For the comparison, the figure also plots the watermark vectors obtained from unmarked audio samples. The threshold is displayed in Figure 12, which divides watermark space into subregions. The embedded watermark vectors are distributed within a watermarked area. The distribution starts at $(5, 5)$, the initial value of the watermark, and spreads in the direction of the origin in the watermark space.

Deterioration trajectory

To observe the deterioration process of watermark vectors systematically, we added various level of white noise to the watermarked samples and detected the watermark vectors as a function of the noise level. Figure 13 displays the deterioration trajectories of the mean vector of the detected watermark at each additive noise level.

The different trajectories show the trajectories starts from different initial states for the embedded watermark vectors. Each dotted line drawn across the three degradation trajectories represents an equal additive noise level.

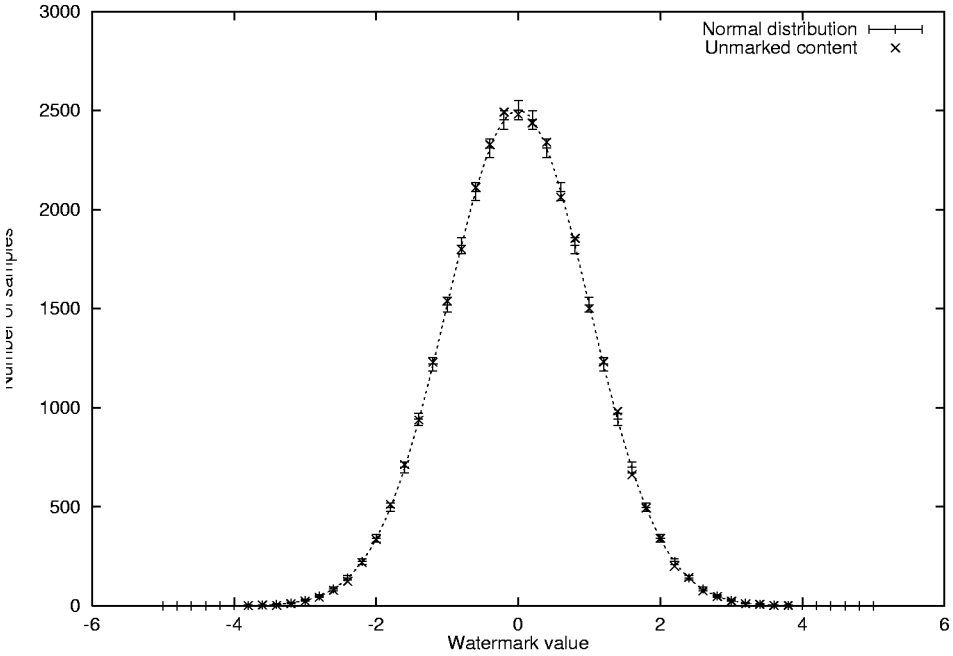

Figure 11: Histogram of watermark value detected from unmarked audio samples

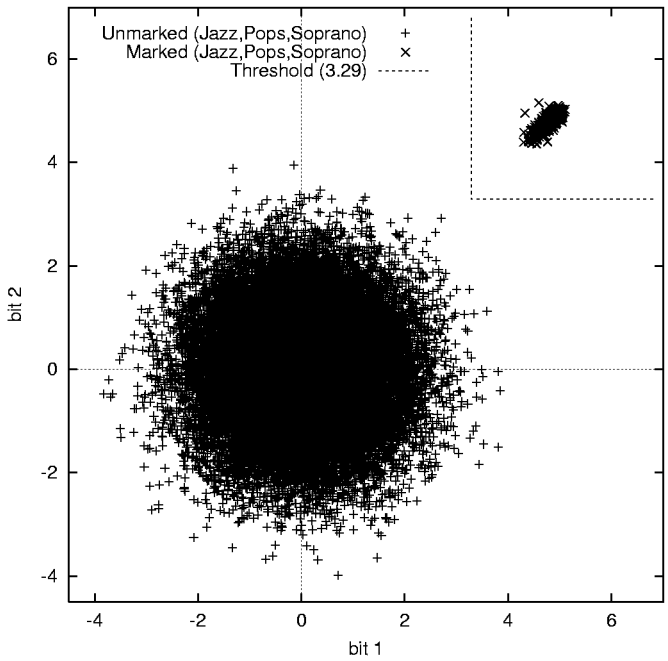

Figure 12: Watermark vector distribution of marked and unmarked content

The obtained result shows that in the watermark degradation process, the mean of watermark distribution changes linearly from the initial state towards the origin of watermark space independently of the initial state of the watermark vector. Table 2 summarizes other transition of each observed watermark moment along with each additive noise level.

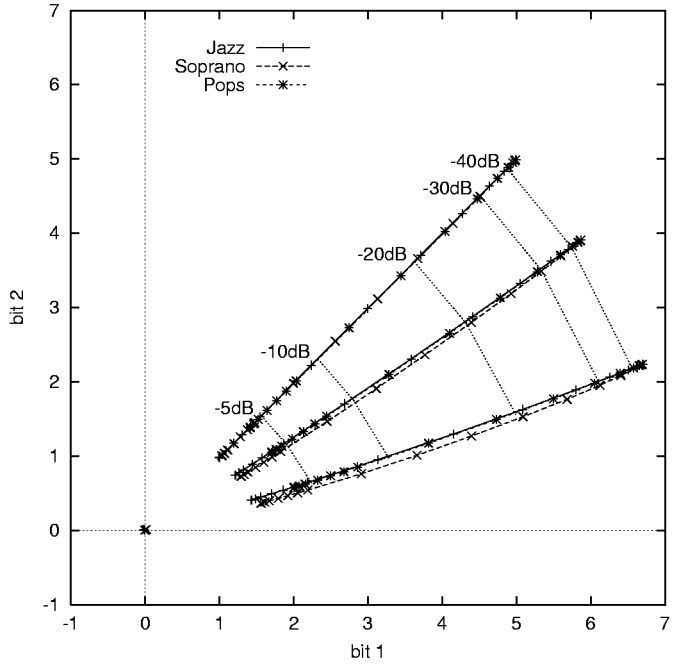

Figure 13: Watermark degradation path in two dimensional watermark space

Here, in these tables, μ, σ, τ, α, and β represent the mean, variance, covariance, skewness (the third-order moment), and kurtosis (the fourth-order moment) of the detected watermark vectors, respectively. The subscript (1 or 2) of each value denotes the bit index. As listed in the table, the covariance τ increases when the level of the additive noise was increased. That means σ_1^2 and σ_2^2 were correlatively degraded by noise addition. Then the variance-covariance matrix of the watermark vectors was diagonalized to find the axes where τ will be zero. σ_r^2 and σ_t^2 in these tables are the variances calculated during this process and θ shows the direction of the new axis (the direction of the eigenvector). According to the obtained θ, we can see that the direction of the new axis almost coincides with the direction of the mean vector.

The above observations of μ and σ^2 transitions suggests that during post processing, the mean vector μ is linearly degraded along the same direction, and the radial and transverse components of the variance, σ_r^2 and σ_t^2 are also spread along the direction of mean vector μ and its orthogonal direction, respectively, with no correlation. Figure 14 shows the deterioration trajectory where the horizontal and vertical axes represent the mean μ and the variance σ_r^2,σ_t^2 respectively. As simulated in the previous section, Figure 14 illustrates that the radial variance has a peak, while the transverse component changes linearly along the deterioration trajectory.

Estimation of bit error ratio

When a specific threshold is chosen to partition the watermark space based on the design of an acceptable FPE ratio, the transition of the FNE and BE ratio can be obtained for any deterioration trajectory. To show the validity of the proposed framework, we compared observed BE ratios and estimated BE ratios derived from approximated moments for a deterioration trajectory.

In this experiment we used the threshold value $\xi = 1.281551$, which yields a FPE ratio 10^{-2}.

Figure 15 shows the linear approximation of variances and covariances observed for the deterioration trajectory. Approximating the moments linearly, the BE ratios for the trajectory are estimated from Equation 6. Table 3 shows a comparison of the observed occurrence of BE and the predicted BE occurrence. As shown in Figure 16, the approximated value is quite close to the observed numbers. This result supports the assumption of the normal distribution approximation for practical use.

The consideration of higher moments in the distribution will further improve the agreement between observed and estimated values. As summarized in Table 2, the actual watermark distribution tends to be skewed towards

Table 2: Transition of the moments in degradation process

Jazz

Noise	μ_1	μ_2	σ_1^2	σ_2^2	τ_{12}	α_1	α_2	β_1	β_2	σ_r^2	σ_t^2	θ
none	4.9927	4.9929	0.0002	0.0002	0.0000	−3.4618	−1.7709	60.9372	17.5253	0.0003	0.0002	42.3900
−40dB	4.9314	4.9310	0.0329	0.0352	0.0171	−4.7340	−4.7629	42.5891	40.8446	0.0512	0.0169	46.9454
−30dB	4.6342	4.6343	0.2448	0.2474	0.1501	−2.5757	−2.6007	11.7620	12.1347	0.3963	0.0960	45.2425
−20dB	3.7107	3.7084	0.8411	0.8367	0.5054	−0.9302	−0.9018	1.3439	1.2835	1.3443	0.3334	44.8758
−10dB	2.2369	2.2253	1.3070	1.3033	0.6222	−0.1894	−0.1917	−0.1243	−0.1836	1.9274	0.6830	44.9153

Pops

Noise	μ_1	μ_2	σ_1^2	σ_2^2	τ_{12}	α_1	α_2	β_1	β_2	σ_r^2	σ_t^2	θ
none	4.9863	4.9869	0.0005	0.0005	0.0003	−4.7476	−5.0141	45.6701	52.9276	0.0008	0.0001	44.6047
−40dB	4.9511	4.9510	0.0147	0.0130	0.0042	−2.9598	−2.1304	22.3094	9.2196	0.0181	0.0095	39.3848
−30dB	4.7448	4.7420	0.1580	0.1508	0.0904	−2.2292	−1.9733	7.8698	5.2392	0.2449	0.0639	43.8574
−20dB	4.0363	4.0264	0.7896	0.7934	0.5424	−1.0598	−1.0367	1.0577	0.8867	1.3339	0.2491	45.0999
−10dB	2.7472	2.7270	1.4823	1.5044	0.9248	−0.2796	−0.2966	−0.3041	−0.2981	2.4182	0.5684	45.3420

Soprano

Noise	μ_1	μ_2	σ_1^2	σ_2^2	τ_{12}	α_1	α_2	β_1	β_2	σ_r^2	σ_t^2	θ
none	4.9756	4.9758	0.0034	0.0033	0.0030	−4.5337	−4.4364	27.8689	25.9972	0.0063	0.0004	44.5802
−40dB	4.7423	4.7357	0.1098	0.1119	0.0524	−1.7921	−1.6455	5.8274	4.6156	0.1633	0.0584	45.5590
−30dB	4.1447	4.1352	0.4816	0.4929	0.2848	−0.8892	−0.8264	0.9086	0.6610	0.7721	0.2025	45.5674
−20dB	3.1279	3.1181	1.0642	1.0942	0.6439	−0.2091	−0.1861	−0.2176	−0.3223	1.7233	0.4351	45.6676
−10dB	1.9946	1.9828	1.3420	1.3950	0.6847	0.1526	0.1432	−0.0240	−0.0838	2.0537	0.6833	46.1080

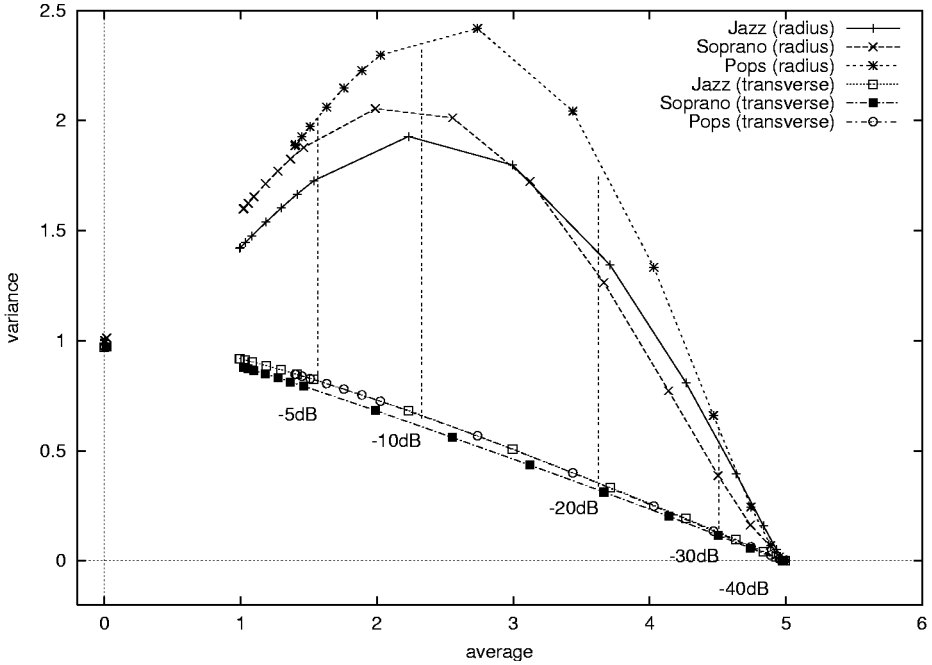

Figure 14: Deterioration trajectory

Table 3: Observed and estimated number of bit errors in 31410 samples

Mean	Observed BE	$\sqrt{(\text{Observed BE})}$	Estimated BE
2.995665	3	1.732051	4.488341
2.317840	18	4.242641	19.031391
1.672880	61	7.810250	65.421728
1.554465	79	8.888194	80.786762
1.440615	93	9.643651	98.452898
1.331630	109	10.440307	118.391285
1.227780	121	11.000000	140.471500
1.177810	129	11.357817	152.256173
1.138790	138	11.747340	162.010293
1.133965	140	11.832160	163.250824

the origin, rather than following a symmetric multi-dimensional normal distribution. The transitions of higher order moments also need to be modeled for more accurate approximation.

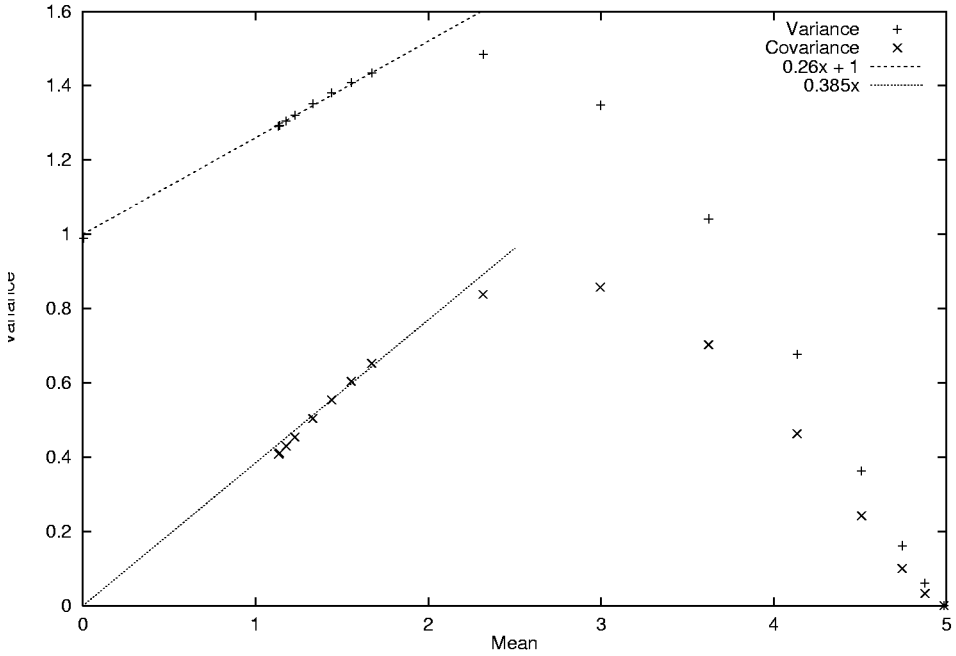

Figure 15: Deterioration trajectory by noise addition

4 Conclusion

The reliability of watermark detection is one of the most important points to be discussed in the commercial use of watermarking.

In this paper, we gave a comprehensive scheme for discussing the false positive error, the false negative error and the bit error during the detection of a multi-bit watermark. We showed how we could compute these error ratios from an assumed distribution function $\phi(\boldsymbol{w})$ of the watermark vector $\boldsymbol{w}$ with a specified threshold. Then we modeled the deterioration of the watermark by the distortion of the distribution function and assumed its moments would lie on a simple trajectory with respect to the degree of deterioration.

The usefulness of this scheme was supported by experimental result using our audio watermarking system. We observed:

- The distribution function of unmarked content can be well approximated by a normal distribution with a common variance equation (12).
- The distribution function of watermarked and damaged content has a correlation between the components of watermark vector which correspond to different bits.

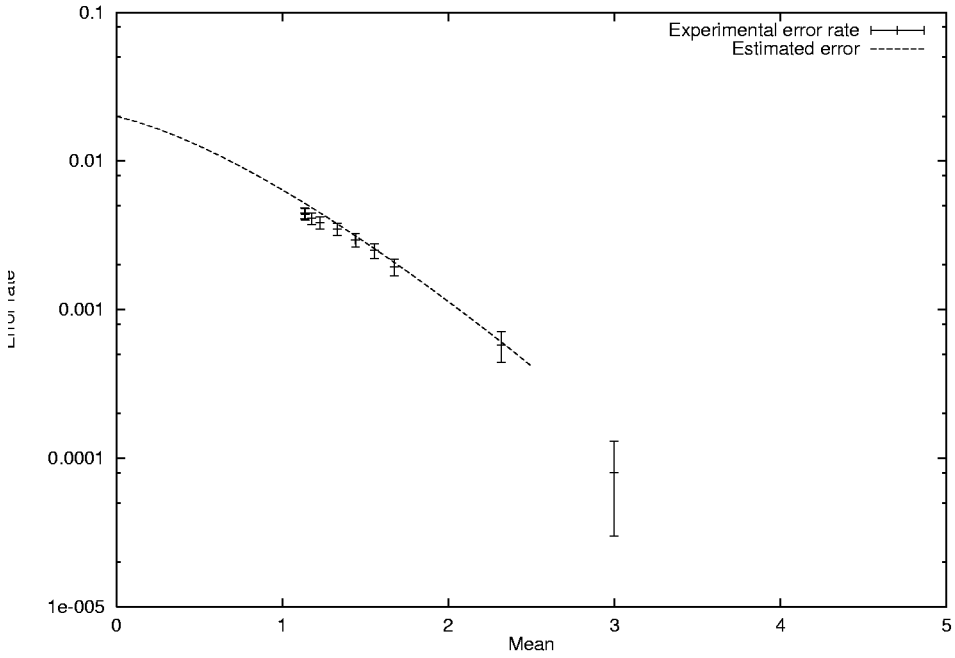

Figure 16: Bit error ratio

- The observed bit error ratio and false negative error ratio moderately coincide with the values which are computed from a multi-dimensional Gaussian distribution using the proposed scheme.
- The moments, averages, and covariance matrix are on a curved trajectory with respect to the degree of deterioration.

References

[1] J. R. Smith and B. O. Comiskey, "Modulation and Information Hiding in Images," *Proceedings of the First Workshop on Information Hiding*, pp. 207–226, 1996.

[2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for Data Hiding," *IBM Systems Journal*, Vol. 35, No. 3/4, 1996.

[3] Geert Depovere, Ton Kalker, and Jean-Paul Linnartz, "Improved Watermark Detection Reliability Using Filtering Before Correlation," *International Conference on Image Processing*, 1998.

[4] Matt L. Miller and Jeffrey A Bloom, "Computing the Probability of False Watermark," *Proceedings of the Workshop on Information Hiding*, 1999.

[5] Jean-Paul M. G. Linnartz, A. A. C. Kalker, G. F. G. Depovere and R. A. Beuker, "A reliability model for the detection of electronic watermarks in digital images," *Benelux Symposium on Communication Theory*, pp. 202–209, October 1997.

[6] Jean-Paul Linnartz, Ton Kalker, and Geert Depovere, "Modeling the False Alarm and Missed Detection Rate for Electronic Watermarks," *Proceedings of the Second Workshop on Information Hiding*, pp. 329–343, 1998.

[7] Frank Hartung and Martin Kutter, "Multimedia Watermarking Techniques", *Proceedings of the IEEE*, Vol. 87, No. 7, 1999.

[8] A. Piva, M. Barni, F. Bartolini and V. Cappellini, "Threshold Selection for Correlation-based Watermark Detection," *Proceedings of COST 254 Workshop on Intelligent Communications*, L'Aquila, Italy, pp. 67–72, June 4–6, 1998.

[9] M. Barni, F. Bartolini, A. De Rosa and A. Piva, "Capacity of the Watermark-Channel : How Many Bits Can Be Hidden Within a Digital Image ?," *Proceedings of SPIE*, Vol. 3657, Security and Watermarking of Multimedia Contents, Electronic Imaging '99, San Jose, CA, January 1999.

[10] Gerrit C. Langelaar, Reginald L. Lagendijk and Jan Biemond, "Watermarking by DCT Coefficient Removal: A Statistical Approach to Optimal Parameter Settings," *Proceedings of SPIE ELECTRONIC IMAGING '99, Security and Watermarking of Multimedia Contents*, 1999.

[11] Laurence Boney, Ahmed H. Tewfik and Khaled N. Harndy, "Digital Watermarks for Audio Signals," *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, pp. 473–480, June 1996.

[12] Mauro Barni, Franco Bartolini and Vito Cappellini, Alessandro Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, Vol. 66, pp. 357–372, 1998.

[13] Jiri Fridrich and Miroslav Goljan, "Comparing robustness of watermarking techniques," *Security and Watermarking of Multimedia Contents*, pp. 214–225, January 1999.

[14] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," *Security and Watermarking of Multimedia Contents*, pp. 226–239, January 1999.

[15] Laurent Piron, Michael Arnold, Martin Kutter and Wolfgang Func, Jean Marc Boucqueau, and Fiona Craven, "OCTALIS benchmarking : Comparison of four watermarking techniques," *Security and Watermarking of Multimedia Contents*, pp. 214–225, January 1999.

[16] Ryo Sugihara, "Bit Error Analysis of Video Watermarking," *Symposium on Cryptography and Information Security (SCIS)*, Okinawa, 1999, to be presented in a TRL Research Report, January 2000.

[17] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamoon, "A Secure, Robust Watermark for Multimedia," *Workshop on Information Hiding*, 1996.