Scale estimation in two-band filter attacks on QIM watermarks

Jinshen Wang[a,b], Ivo D. Shterev[a], and Reginald L. Lagendijk[a*]

[a] Delft University of Technology, 2628 CD Delft, Netherlands;
[b] Nanjing University of Science & Technolgy, China

ABSTRACT

This paper presents a scheme for estimating two-band amplitude scale attack within a quantization-based watermarking context. Quantization-based watermarking schemes comprise a class of watermarking schemes that achieves the channel capacity in terms of additive noise attacks[1]. Unfortunately, Quantization-based watermarking schemes are not robust against Linear Time Invariant (LTI) filtering attacks. We concentrate on a multi-band amplitude scaling attack that modifies the spectrum of the signal using an analysis/synthesis filter bank. First we derive the probability density function (PDF) of the attacked data. Second, using a simplified approximation of the PDF model, we derive a Maximum Likelihood (ML) procedure for estimating two-band amplitude scaling factor. Finally, experiments are performed with synthetic and real audio signals showing the good performance of the proposed estimation technique under realistic conditions.

Keywords: Watermarking, quantization, maximum likelihood estimation, multi-band

1. INTRODUCTION

Watermarking schemes based on quantization theory have recently emerged as a result of information theoretic analysis[1,2]. In terms of additive noise attacks, these schemes have proven to perform better than traditional spread spectrum watermarking because the used lattice codes achieve capacity for the AWGN channel. Another important feature of quantization-based watermarking schemes is that they can completely cancel the host signal interference, which makes them invariant to the host signal. A similar phenomenon exists in channel coding with side information at the encoder[3].

Unfortunately, quantization-based watermarking schemes such as Quantization Index Modulation watermarking with Distortion Compensation (QIM with DC) [2] are not robust against LTI filtering attacks. Considering the implementation of a quantization-based scheme in a LTI filtering setting, it is likely that the scheme will fail. Weakness against LTI filtering is a serious drawback, since many normal operations on images and audio are explicitly implemented with linear filters. The bass and treble adjustments in a stereo system apply simple filtering operations. In addition, many other operations, although not explicitly implemented with filters, can be modeled by them. For example, playback of audio over loudspeakers can also be approximated as a filtering operation.

In this paper, we focus on multi-band amplitude scaling problem in combination with additive noise attack. One of its applications of which is a multi-band equalizer that modifies the spectrum of the signal using the filter bank. The signal

Further author information: (Send correspondence to Jinshen Wang)

Jinshen Wang: E-mail: j.s.wang@ewi.tudelft.nl

Security, Steganography, and Watermarking of Multimedia Contents VIII, edited by Edward J. Delp III, Ping Wah Wong,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6072, 60720B, © 2006 SPIE-IS&T · 0277-786X/06/$15

frequency range is divided into a number of frequency bands and the signal may be amplified or attenuated in each of these bands independently. To see how serious the problem can be, figure 1 shows the behavior of QIM with DC for a variety of Document to Watermark ratio (DWR), when the watermarked signal is attacked by a two-band filter bank with a scaling β in the high frequency band depicted in Figure 2.

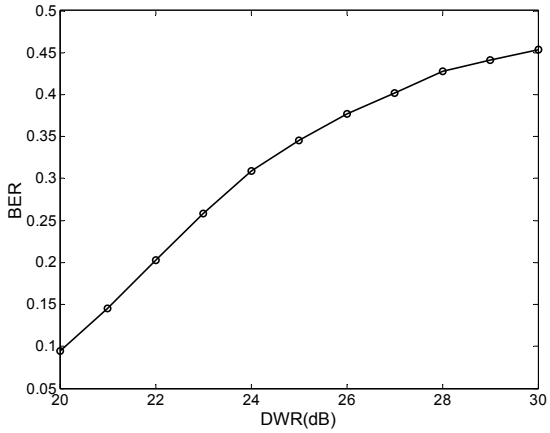

Figure 1. Probability of error for different values of DWR. β= 0.95, no noise.

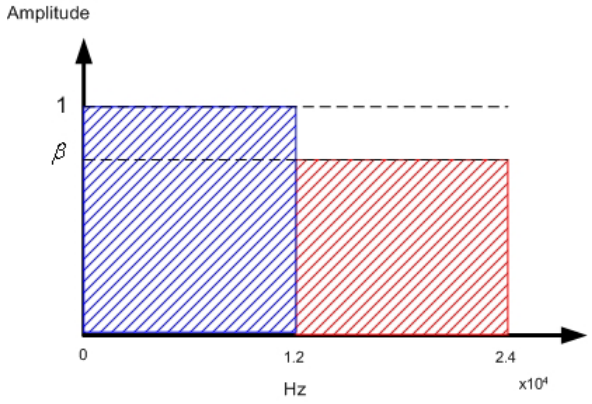

Figure 2. Amplitude response of the filter.

The solutions proposed so far to deal with one channel amplitude scaling attack, in the framework of QIM watermarking, can be grouped into two main categories: One of the approaches is based on designing watermarking codes that are resilient to amplitude scaling operation, such as trellis codes[4,8]. Another approach is based on estimation the amplitude scaling operation and inverting them prior to watermark decoding[5]. However, to the best of our knowledge, no earlier work with regard to multi-band amplitude scaling has been proposed before.

The paper is organized as follows: in Section 2 we formulate the multi-band amplitude scale attack and introduce some important notation. In section 3 we derive the PDF models for frequency band amplitude scaled signal and attacked signal respectively. A description of the estimation procedure is given in Section 4. Section 5 contains experimental results from synthetic and real audio host signals, and Section 6 concludes the paper.

2. MATHEMATICAL FORMULATION

In this section, we define some notational conventions. We assume that the host signal is arranged in an N-dimensional vector $\boldsymbol{x}$, i.e., $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where x_k ($k \in 1, \cdots, N$) refers to the k-th element. Throughout the paper, random variables are denoted by capital letters and their realizations by the respective small letters. The notation $X \sim f(x)$ indicates that the random variable X has a PDF $f(x)$. Vectors will be denoted by bold letters.

Figure 3. illustrates block-diagram of the system. It can be divided into: the basic quantization-based watermark embedding and decoding respectively, multi-band amplitude scaling attack, estimator and corrector. The basic embedding and decoding procedure are based on QIM with DC, proposed by Chen and Wornell[2]. In the watermark encoder, where $b_k \in \{0, 1\}$ denotes the message bits that are embedded in the host data, $\boldsymbol{x}$ is the host signal itself with a variance σ_x^2, $\boldsymbol{y}$ is the watermarked signal.

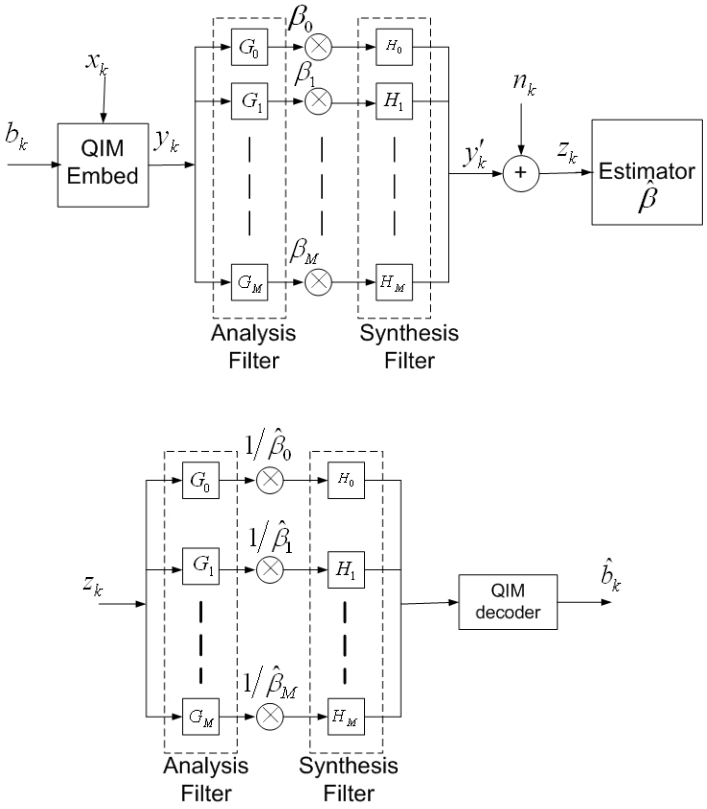

Figure 3. Block-diagram of the general system.

The multi-band amplitude scaling attack consists of an analysis/synthesis filter bank and a constant scaling of the amplitude of the watermarked signal in each band. Furthermore, we will assume that zero-mean additive white Gaussian noise $\boldsymbol{n}$ with variance σ_N^2 and independent of the output of the filter attack $\boldsymbol{y'}$ is also added by the attacker. Let $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_M]$, where $\beta_i > 0$, for all i, denotes the Multi-band amplitude scaling factor vector, and M is the number of the frequency channel. Following our model, the Fourier transform of the $\boldsymbol{y'}$ can be written as

$$
\begin{aligned}
Y'(e^{j\omega}) &= T(e^{j\omega})Y(e^{j\omega}) \\
&= [\beta_0 G_0(e^{j\omega})H_0(e^{j\omega}) + \beta_1 G_1(e^{j\omega})H_1(e^{j\omega}) + \cdots + \beta_M G_M(e^{j\omega})H_M(e^{j\omega})]\, Y(e^{j\omega}),
\end{aligned} \tag{1}
$$

where $G(e^{j\omega})$ and $H(e^{j\omega})$ are the transfer function of a lowpass filter and a highpass filter respectively.

Then, the attacked vector $\boldsymbol{z}$ is given as

$$
\boldsymbol{z} = \boldsymbol{y}' + \boldsymbol{n}\,. \tag{2}
$$

Finally, it is useful to define some quantities that relate the powers of the host, the watermark and noise. The *Document to Watermark Ratio* (DWR) is given by $10\log(\sigma_x^2/\sigma_w^2)$; the *Watermark to Noise Ratio* (WNR) is $10\log(\sigma_w^2/\sigma_N^2)$. These quantities are expressed in decibels.

3. PDF MODELS

In this section we derive the PDF models for frequency band amplitude scaled vector $\boldsymbol{y}$' and attacked vector $\boldsymbol{z}$ as a function of $\boldsymbol{\beta}$. These PDF models are the basis for the ML procedures for estimating β developed in section 4.

Referring to Figure 3, multi-band amplitude scaling attack in each frequency band consists of a twin LTI filters and a scaling factor β_k. Assume that the filter bank holds *Perfect Reconstruction* (PR) property and if the scaling vector $\boldsymbol{\beta} = \mathbf{1}$, we obtain:

$$
y_k = y_k'\,. \tag{3}
$$

For $\boldsymbol{\beta} \neq \mathbf{1}$, (3) does not any longer hold; hence it leads to watermark detection error because the watermarked signal is moved away from the correct centroids. From (1), we can see that transfer function $T(e^{j\omega})$ carries information of $\boldsymbol{\beta}$. Since our goal is to derive PDF of frequency band amplitude scaled vector $\boldsymbol{y}$', it would be reasonable to use time domain representation of (1). Then $\boldsymbol{y}$' can be written as:

$$
\begin{aligned}
y'(k) &= t(k) * y(k) \\
&= t(0)y(k) + t(1)y(k-1) + t(2)y(k-2) + \cdots + t(k)y(0)\,,
\end{aligned} \tag{4}
$$

where $t(k)$ denotes the impulse response of $T(e^{j\omega})$. Note that the impulse response $t(k)$ is known to the estimator.

We see that the overall filter operates by summing weighted delayed versions of the watermarked vector $\boldsymbol{y}$. In order to derive PDF of frequency band amplitude scaled vector $\boldsymbol{y}$', we assume that the host signal and the watermarked signal are independent identical distribution (i.i.d.) vector sources. We note that this assumption is only an approximation for the real world case. Thus, the frequency band amplitude scaled vector sample y_k' is a weighted sum of i.i.d. random variables y_k. In our previous publication[5], we have derived the PDF model for the watermarked data $\boldsymbol{y}$, i.e., $f_Y(y)$. Then, the PDF of the $\boldsymbol{y}$' is given as:

$$
f_{Y'}(y') = \frac{1}{|t(0)|} f_Y\!\left(\frac{y}{t(0)}\right) * \frac{1}{|t(1)|} f_Y\!\left(\frac{y}{t(1)}\right) * \cdots * \frac{1}{|t(k-1)|} f_Y\!\left(\frac{y}{t(k-1)}\right). \tag{5}
$$

To simplify the multi-band amplitude scaling problem, we confine ourselves to use a simplified model, namely, a two-band filter bank, and the scaling factor only exists in the high frequency band, in other words, the scaling factor vector is $\boldsymbol{\beta} = [\,1\ \beta]$.

Figure 4 illustrates the statistical distribution of the output of the filter attack $\boldsymbol{y'}$, showing the sufficient accuracy in the predicted PDF model. For $\beta = 1.0$ the analytical PDF is that of the typical QIM watermarked signal.

In addition, there are only several filter coefficients $t(k)$ which have relatively large magnitude. So it is reasonable to consider that these filter coefficients with larger magnitude play important role in (5). Therefore, $f_{Y'}(y)$ can be simplified by substituting only a few filter coefficients with larger magnitude into (5), instead of using all filter coefficients. Let L denote the necessary number of filter coefficients. Figure 5 illustrates $f_{Y'}(y)$ for different L.

From Figure 5, we can see that in this case, $L = 3$ is sufficient for (5). For large L, there is no evident improvement of accuracy of the analytical PDF model, which verifies that (5) can be simplified by substituting only a few filter coefficients with larger magnitude.

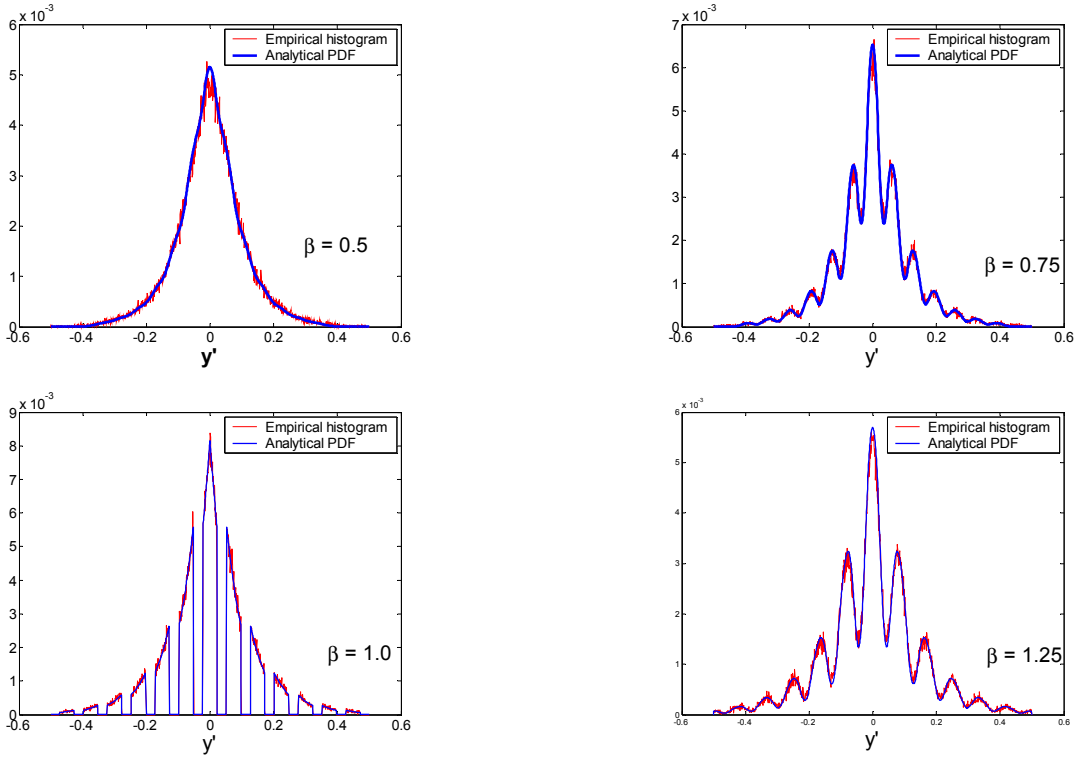

Figure 4. Analytical PDF for different β vs. empirical histogram for a Laplacian host, DWR = 15dB. The amplitude response of the filter is shown in figure 2.

Taking into account the additive noise $\boldsymbol{n}$, we obtain the PDF of the attacked vector $\boldsymbol{z}$:

$$f_Z(z) = f_N(n) * f_{Y'}(y'), \tag{6}$$

where the convolution * follows from the independence between additive noise $\boldsymbol{n}$ and $\boldsymbol{y'}$. $f_Z(z)$ is shown in Figure 6. We see that the PDF model of the attacked vector matches the histogram quite well for additive noise case.

4. MAXIMUM LIKELIHOOD ESTIMATION

The PDF model of attacked vector has been derived as a function of β in the previous section. We are now able to use the model to estimate β from the attacked vector $\boldsymbol{z}$.

Maximum Likelihood (ML) Estimation can be used to solve this problem. The ML estimation of β is done based on (6).

By definition[7], the ML estimation $\hat{\beta}$ of the scaling factor β is given as:

$$\hat{\beta} = \arg\max_{\beta} \boldsymbol{f}_{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_N}(z_1, z_2, \ldots, z_N \,|\beta) \,. \tag{7}$$

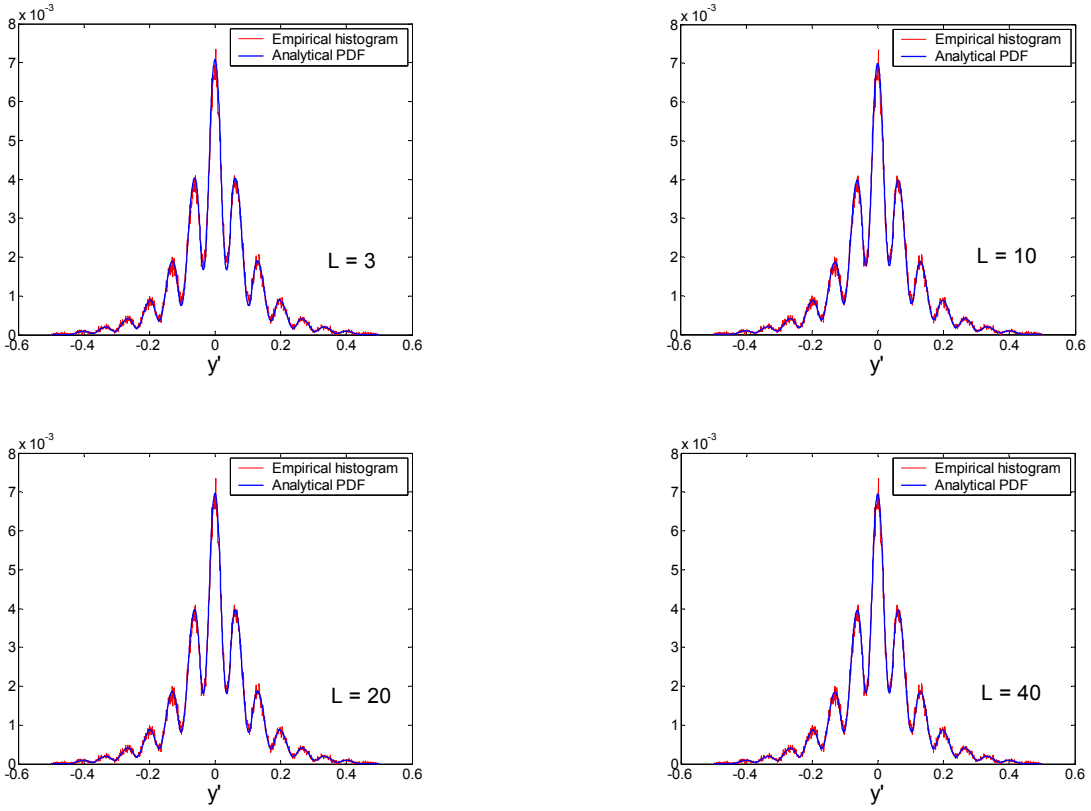

Figure 5. Analytical PDF for different L vs. empirical histogram for a Laplacian host, β = 0.8, DWR = 15dB. The amplitude response of the filter is shown in figure 2.

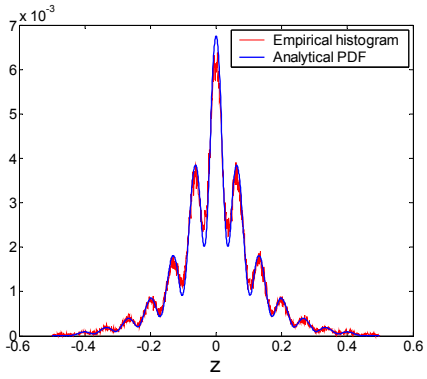

Figure 6. PDF of attacked vector z vs. empirical histogram for Laplacian host, β = 0.8, WNR = 3 dB, DWR = 15dB. The amplitude response of the filter is shown in figure 2.

However, it is difficult to derive the joint PDF from the PDF of z_k. Recall that for deriving (5), we have made an assumption that the frequency band amplitude scaled vector ***y'*** has i.i.d. components, so it is reasonable to consider that the vector ***z*** will also have approximately i.i.d. components.

Therefore, the joint PDF can be approximately written as a product of the marginal PDFs, that is,

$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} \prod_{i=1}^{N} f_{Z_i}(z_i \,|\beta) \\ &= \arg\max_{\beta} \sum_{i=1}^{N} \log f_{Z_i}(z_i \,|\beta)\,.\end{aligned} \tag{8}$$

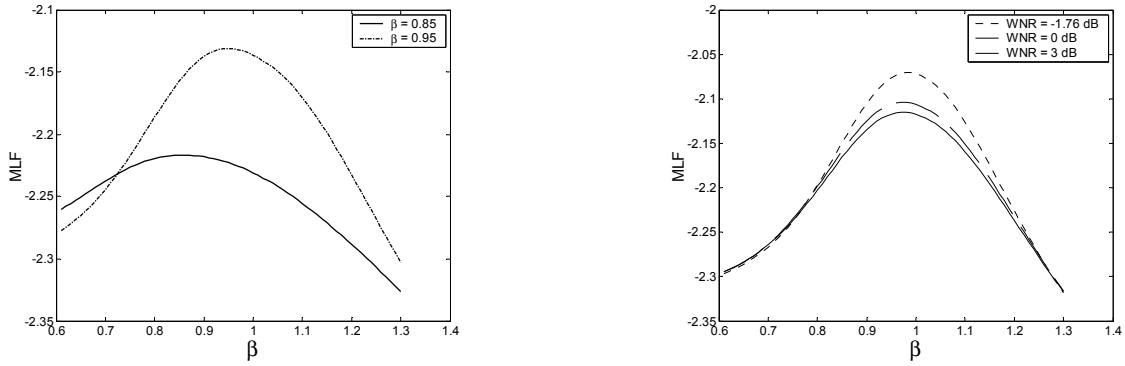

Figure 7. Graph of MLF for different values of $\hat{\beta}$ (a) and different values of WNR (b). Chosen settings are $X \sim L(0, 0.02)$, $N \sim N(0, 0.01)$, and $\sigma_W^2 = 0.01$. The amplitude response of the filter is shown in figure 2.

The Maximum Likelihood Functional (MLF) is the expression $\sum_i \log f_{Z_i}(z_i|\beta)$. Experimental curves of the MLF for different values of β and WNR are shown in figure 7. Since it is difficult to find an analytical expression of $\hat{\beta}$, we do a brute force search for the optimal value of β based on (8).

5. EXPERIMENTS

In this section we describe experiments with synthetic and real audio signals (with sampling frequency 48kH$_z$) carried out to test the estimation accuracy of the proposed techniques in terms of WNR, the parameter β, and the number of available signal samples N. Furthermore, we experimentally show how inverting the effect of the attack can significantly help to reduce the bit error rate.

5.1. Estimation Performance

Experimental results in terms of WNR and N are shown in figure 8. The assumed PDF model of the host signal at the estimator side is a zero-mean Laplacian PDF with variance equal to the variance of the sum of the variances of the host signal, watermark, and the noise in the attack channel, i.e., $L(0, \sigma_X^2 + \sigma_W^2 + \sigma_N^2)$. This is a realistic assumption, because the decoder has access to the received data and can estimate its variance. Furthermore, in practice most audio signals have a PDF that resembles the Laplacian PDF. The loss in performance of the ML approach is due to the approximation in $f_Z(z)$ and the fact that generally, ML estimation requires a large sample size[7]. In Figure 9, we plot experimental results of $\beta - \hat{\beta}$ as a function of β for different audio signals.

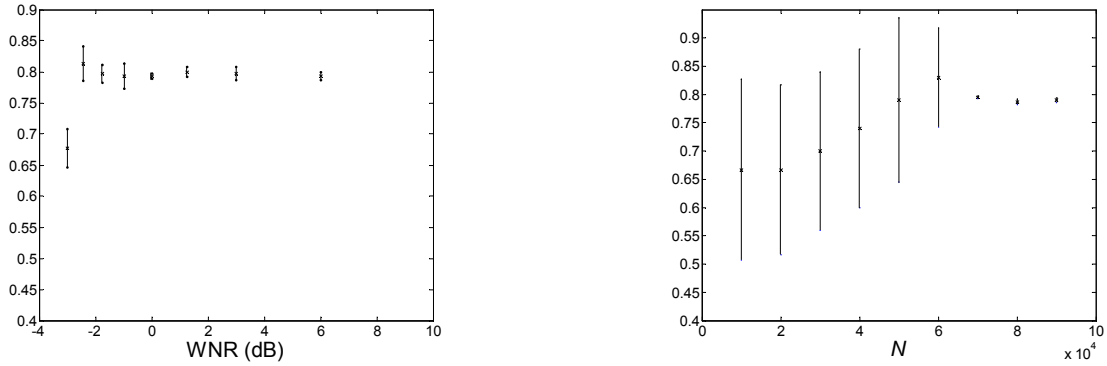

Figure 8. Graphs of $\hat{\beta}$ for real audio signals as a function of WNR (a) and as a function of available signal samples N (b). The crosses represent the estimation mean, and the lines the estimation standard deviation in both directions. DWR = 15dB. The assumption for the estimator is $\boldsymbol{X} \sim \boldsymbol{L}(0, \sigma_X^2 + \sigma_W^2 + \sigma_N^2)$. The amplitude response of the filter is shown in figure 2.

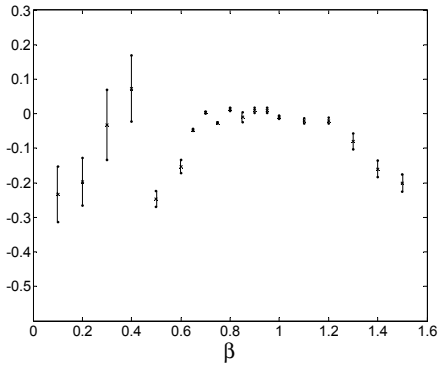

Figure 9. Graphs of $\beta - \hat{\beta}$ for real audio signals as a function of β. The crosses represent the mean, and the lines the standard deviation in both directions. DWR = 15dB, and WNR = 0dB. The assumption for the estimator is $X \sim L(0, \sigma_X^2 + \sigma_W^2 + \sigma_N^2)$. The amplitude response of the filter is shown in figure 2.

5.2. Inversion the Effect of Two-Band Amplitude Attack

Figure 10 shows the behavior of watermark decoder when the attacked signal is passed through the corrector depicted in Figure 3. The host signal is white noise, the DWR is 15dB, the number of signal samples is 80000 and the scaling factor is 0.8. The BER for reception of attacked signal and the BER for reception of corrected signal using the corresponding estimates are compared. Figure 10 illustrates how inversion of the effect of two-band amplitude attack leads to significant performance improvements. The BER increases as WNR decreases, since the estimation accuracy decreases due to the strong noise.

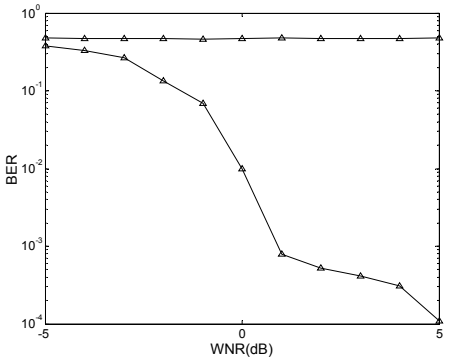

Figure 10. Watermark decoder performance. DWR=15dB, β=0.8. The amplitude response of the filter is shown in figure 2.

6. CONCLUSIONS

In this paper, we have presented a Maximum Likelihood estimation procedure for estimating a two-band amplitude scaling factor. The estimation approach performs well in terms of additive noise and for a relatively wide range of values for the parameter β, under realistic assumptions. The disadvantage is the need for a relatively large amount of signal

samples for estimating reliably β. Another disadvantage is that the method is computationally expensive and currently not suitable for real-time applications.

REFERENCES

1. P. Moulin and A. O'Sullivan. Information-Theoretic Analysis of Information Hiding. IEEE Transactions on Information Theory, 49(3):563–593, March 2003.
2. B. Chen and G. Wornell. Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. IEEE Transactions on Information Theory, 47:1423–1443, May 2001.
3. M. H. Costa. Writing on Dirty Paper. IEEE Transactions on Information Theory, 29(3):439–441, May 1983.
4. M. L. Miller, G. J. Doerr, and J. Cox. Dirty-Paper Trellis Codes For Watermarking. IEEE International Conference On Image Processing, 2:129–132, September 2002. Rochester, NY.
5. I. D. Shterev and R. L. Lagendijk, "Maximum Likelihood Amplitude Scale Estimation for Quantization-Based Watermarking in the Presence of Dither", SPIE Security, Steganography, and Watermarking of Multimedia Contents VII , San Jose, CA, January 2005.
6. J. J. Eggers, R. Bauml, and B. Girod, "Estimation of Amplitude Modifications before SCS Watermark Detection," SPIE Security and Watermarking of Multimedia Contents IV, vol. 4675, pp. 387-398, January 2002, San Jose, CA, USA.
7. H. V. Poor. An Introduction to Signal Detection and Estimation. Springer-Verlag, second edition, 1994.
8. F. Pérez-González, M. Barni, A. Abrardo, and C. Mosquera. Rational Dither Modulation: a novel data-hiding method robust to value-metric attacks. IEEE International Workshop on Multimedia Signal Processing, Sept 29 - Oct 1 2004, Siena, Italy.