OPTIMAL DISTORTION COMPENSATION FOR QUANTIZATION WATERMARKING

Marius Staring, Job Oostveen, Ton Kalker

Philips Research Eindhoven, The Netherlands

ABSTRACT

In this paper we study the problem of optimizing the distortion compensation parameter for the Scalar Costa Scheme, which is a practical version of the class of Distortion Compensated Dither Modulation schemes. In the literature, a number of results are known for finding the value of the distortion compensation parameter that maximizes the capacity of the watermarking channel. Instead, in this paper, we look at minimization of the bit error probability as the criterion for determining the optimal value of the distortion compensation parameter. To this end, we derive a model for the bit error probability, which is subsequently approximated and minimized. This is done both for the cases of Gaussian noise and uniform noise. The results match very well with earlier results by Eggers.

1. INTRODUCTION

In this paper we study optimization of the distortion compensation parameter α of a class of watermarking schemes introduced by Chen and Wornell (see, for instance, [1]) and known as *"Distortion Compensated Dither Modulation"*. It builds directly on the scheme introduced by Costa in [2]. The scheme is based on large, random code-books. Watermark embedding boils down to replacing a vector of samples by a nearby codeword that corresponds to the to-be-embedded symbol. In order to optimize the rate of the watermark channel an additional parameter $\alpha (0 \leq \alpha \leq 1)$ is introduced, referred to by Chen and Wornell as the distortion compensation parameter. The interpretation of this parameter is that, rather than replacing a sample x by the nearby codeword c, an intermediate point is chosen. That is, x is replaced by $x + \alpha(c - x) = (1 - \alpha)x + \alpha c$. Note that setting $\alpha = 1$ equals embedding without distortion compensation. At the other extreme, setting $\alpha = 0$ is equivalent to complete distortion compensation, i.e., no embedding at all.

Both papers [1, 2] derive an expression for the rate-maximizing value α^* of the distortion compensation parameter α for the *"Ideal Costa Scheme"* (ICS):

$$\alpha^*_{\text{ICS}} = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_n^2}, \tag{1}$$

Submitted to ICIP 2003
job.oostveen@philips.com

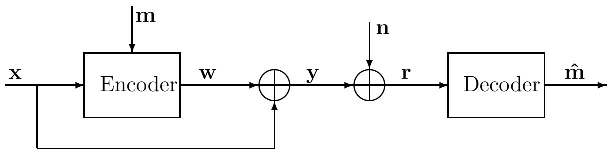

Fig. 1. The set-up for the quantization watermarking schemes

where σ_w^2 and σ_n^2 are the watermark variance and the noise variance respectively. The naming "Ideal Costa Scheme was introduced by Eggers in [3].

In practice, the schemes of Costa and Chen and Wornell cannot be used easily. Both are based on very large, random code-books, which lead to impractical and complex watermark embedders and detectors. A very practical approximation of these schemes was published by Eggers et al. [3]. Their *Scalar Costa Scheme* (SCS) replaces the large, random code-books by very simple, structured code-books consisting of sample-wise uniform quantizers. For this simpler scheme (which from the perspective of capacity is sub-optimal) a different value of α will maximize the rate. In [3] a numerical approach was taken to derive an expression for the rate-maximizing value α^*_{SCS} of α for the Scalar Costa Scheme.

In this article we take a different approach for optimizing the distortion compensation parameter α for the Scalar Costa Scheme. In Section 3, we derive an analytical model for the bit error probability for the cases of Gaussian and uniform noise sources. In Section 4 we proceed by finding the value α^*_{BE} that minimizes this bit error probability. We compare the resulting α^*_{BE} with the expression for α^*_{SCS}, derived by Eggers. We start by recalling the Scalar Costa Scheme in some detail in Section 2.

2. THE SCALAR COSTA SCHEME

In this section we introduce the Scalar Costa Scheme. See Figure 1 for a schematic drawing of the setting. Assume that the host signal $\mathbf{x}$ is an i.i.d. random vector of length N. For the moment, we will not make any assumptions on the distribution of the samples of $\mathbf{x}$. However, it is assumed that the power of $\mathbf{x}$ equals σ_x^2. A message vector $\mathbf{m}$ is embedded into this host signal. The resulting, watermarked signal is denoted $\mathbf{y}$ and the watermark $\mathbf{w}$ is defined as the difference between the host signal and the watermark signal: $\mathbf{w} = \mathbf{y} - \mathbf{x}$. The decoder acts on the received signal $\mathbf{r}$, which is a noise-corrupted version of the watermarked signal $\mathbf{y}$: $\mathbf{r} = \mathbf{y} + \mathbf{n}$, where the noise signal $\mathbf{n}$ has power σ_n^2. We assume that the message $\mathbf{m}$ is encoded such that it is a vector of length N (the same length as the host signal) over the alphabet $\mathcal{D} = \{0, 1, \ldots, D-1\}$.

The Scalar Costa Scheme works on a sample-by-sample basis. The value of the to-be-embedded symbol m determines the selection of a quantizer $Q_\Delta^m(\cdot)$, which quantizes according to

$$Q_\Delta^m(x) \quad = \quad \text{round}\left(\frac{x}{\Delta} - \frac{m}{D}\right)\Delta + \frac{m\Delta}{D}. \qquad (2)$$

That is, $Q_\Delta^m(x)$ denotes rounding x to the nearest value of the form $\left(k + \frac{m}{D}\right)\Delta$ for some integer k. For the case of binary messages ($D = 2$) this boils down to rounding x to the nearest odd or even multiple of $\frac{\Delta}{2}$, depending on the value of m. A message symbol m is embedded in the corresponding host signal sample x according to the following formula:

$$\begin{aligned} y &= (1-\alpha)x + \alpha Q_\Delta^m(x) \qquad (3) \\ &= x + \alpha\left(Q_\Delta^m(x) - x\right). \end{aligned}$$

In this formula, α is the distortion compensation parameter. It determines what fraction of the quantization error $Q_\Delta^m(x) - x$ is embedded as watermark; $w = \alpha(Q_\Delta^m(x) - x)$.

Detection is done by rounding the received value r to the nearest value $(k + \frac{\hat{m}}{D})\Delta$ and computing the corresponding value $\hat{m}$:

$$\hat{m} \quad = \quad \text{round}\left(\frac{rD}{\Delta}\right) \mod D. \qquad (4)$$

A useful interpretation of the distortion compensation parameter is the following. Assume that there is a power-constraint σ_w^2 for the watermark. The question then becomes how to use this power-budget, in such a way that the robustness of the watermark is maximized. The power of the watermark $\mathbf{w}$ is determined by the product $\alpha\Delta$. Similarly, a measure for robustness (like the expected bit error probability at a given noise-intensity) will depend on both Δ and α, as well. This leads to the question of optimizing the robustness measure for Δ and α, subject to the power-constraint. This is the approach we will follow in the remainder of this paper.

Any practical quantizer would include the use of a dither. The application of dither has two important effects. First it reduces the effect of various quantization-related artifacts in the quantized signal (like, for instance, false contours in images). Secondly, it allows to shape the probability density function of the to-be-quantized signal. Typically the dither is chosen such that a distribution which is uniform within each quantization interval is obtained (see [4]). For simplicity, we do not explicitly incorporate dither in the formulas. However, we do assume that the host signal $\mathbf{x}$ has a uniform distribution within each quantization interval $[-\Delta/2, \Delta/2]$. Explicit incorporation of dither in our analysis would not change any of the conclusions, but only complicate the formulas.

3. A MODEL FOR THE BIT ERROR PROBABILITY

In this section, we derive a model for the bit error probability of the SCS watermarking channel under an additive noise attack. So, we assume that the received sample $r = y + n$, where n is drawn from a zero-mean Gaussian or uniform distribution with variance σ_n^2 and where y is given by Equation (3). Recall that x denotes the original, unwatermarked, host signal sample. The results are presented for the case of Gaussian noise in Section 3.1 and for uniform noise in Section 3.2.

From the detection formula (4), it can be seen that the message symbol m is associated with the sequence of intervals $I_{m,k}$ ($k \in \mathbb{Z}$) given by

$$I_{m,k} \quad = \quad \left[(k + \frac{2m-1}{2D})\Delta, (k + \frac{2m+1}{2D})\Delta\right).$$

The message symbol m is correctly decoded if and only if $r \in I_{m,k}$ for some $k \in \mathbb{Z}$. Assuming (without loss of generality) $m = 0$, the bit error probability P_{BE} can be written as

$$\begin{aligned} P_{\text{BE}} &= 1 - \sum_{k\in\mathbb{Z}} P(r \in I_{0,k}) \\ &= 1 - \sum_{k\in\mathbb{Z}} P\big(r - Q_\Delta^0(x) \in I_{0,k}\big) \\ &= 1 - \sum_{k\in\mathbb{Z}} P\left((k - \frac{1}{2D})\Delta \le v + n < (k + \frac{1}{2D})\Delta\right), \end{aligned}$$

where we defined $v = (1-\alpha)(x - Q_\Delta^0(x))$ as the difference between x and the center of the nearest quantization bin corresponding to $m = 0$. By assumption, x is distributed uniformly over the quantization bin, and consequently v is distributed uniformly on $[-\frac{(1-\alpha)\Delta}{2}, \frac{(1-\alpha)\Delta}{2}]$.

As this infinite sum is hard to evaluate, we approximate it by the K^{th} order truncation

$$P_{\text{BE}}^K = 1 - \sum_{k=-K}^{K} P\left((k - \frac{1}{2D})\Delta \le v + n < (k + \frac{1}{2D})\Delta\right).$$

Note that P_{BE}^K is a monotonically decreasing sequence of nonnegative numbers, and hence it converges to P_{BE}.

In the following sections, we give expressions for the 0-th order truncation of the above sum, for the cases that the noise signal n is distributed according to a Gaussian or a uniform distribution.

3.1. The Gaussian noise case

In this section we derive an explicit expression for the 0^{th}-order approximation P_{BE}^0 assuming a Gaussian noise signal $\mathbf{n}$. For reasons of space we will only state the result here. The derivation can be done by first deriving an expression for the conditional bit error probability, given the value v, and subsequently integrating the result over the distribution of v. As a result, we have

$$\begin{aligned} & P_{\mathrm{BE}}^0 \qquad (5) \\ = \quad & 1 + \frac{1-(1-\alpha)D}{2(1-\alpha)D}\,\mathrm{erf}\left(\frac{(1-(1-\alpha)D)\Delta}{2\sqrt{2}D\sigma_n}\right) \\ & - \frac{1+(1-\alpha)D}{2(1-\alpha)D}\,\mathrm{erf}\left(\frac{(1+(1-\alpha)D)\Delta}{2\sqrt{2}D\sigma_n}\right) \\ & + \frac{\sigma_n\sqrt{2}}{(1-\alpha)\Delta\sqrt{\pi}} \cdot \\ & \left(e^{-\frac{\Delta^2(1+(1-\alpha)D)^2}{8D^2\sigma_n^2}} - e^{-\frac{\Delta^2(1-(1-\alpha)D)^2}{8D^2\sigma_n^2}}\right). \end{aligned}$$

Similar, but increasingly complicated formulae can be derived for P_{BE}^K, with $K > 0$.

The above formula is to complicated to interpret. Still it is useful for determining the optimal value of the distortion compensation parameter.

3.2. The uniform noise case

In this section we derive an explicit expression for the 0^{th}-order approximation P_{BE}^0 assuming a uniform noise signal $\mathbf{n}$. For reasons of space we will only state the result here. The derivation can be done by first deriving an expression for the conditional bit error probability, given the value v, and subsequently integrating the result over the distribution of v. Let γ denote the ratio σ_w/σ_n. As a result, we have

$$P_{\mathrm{BE}}^0 = \begin{cases} 1 - \frac{1}{2(1-\alpha)} & \text{if } 0 \le \alpha \le \frac{\gamma}{2+2\gamma}, \\ 0 & \text{if } \gamma \ge 2 \text{ and } \alpha \ge \frac{\gamma}{2\gamma-2}, \\ \frac{(\gamma-2\alpha)^2}{16\gamma\alpha(1-\alpha)} + \frac{2-\gamma}{4} & \text{otherwise.} \end{cases} \qquad (6)$$

Similar, but increasingly complicated formulae can be derived for P_{BE}^K, with $K > 0$.

The above formula shows that in the case of uniform noise, the Scalar Costa Scheme can provide an error-free channel, provided that the watermark-to-noise ratio is sufficiently large.

4. OPTIMAL DISTORTION COMPENSATION

In this section we will use the expression for the bit error probability of the previous section to find an optimal value $\alpha^*_{\mathrm{BE,K}}$ of the distortion compensation parameter α. Again, we will split the problem in two cases: the case of Gaussian noise and the case of uniform noise. For the latter it will be possible to derive an analytical optimum. The Gaussian case is solved numerically.

4.1. The Gaussian noise case

As analytical minimization of the expression for P_{BE}^K is a difficult task, we will instead use numerical optimization routines. For simplicity we will do this for the binary case ($D = 2$). This makes it possible to compare to the result by Eggers et al. [3]. In that paper a numerical model is used for the mutual information, which is optimized numerically to obtain the rate-maximizing value α^*_{SCS} of the distortion compensation parameter. By curve-fitting through these numerical values, the following expression was obtained:

$$\alpha^*_{\mathrm{SCS}} = \sqrt{\frac{\sigma_w^2}{\sigma_w^2 + 2.71\sigma_n^2}}. \qquad (7)$$

In Figure 2, we plot the optimal distortion compensation parameter as a function of the watermark-to-noise ratio. The dotted line is $\alpha^*_{\mathrm{BE,0}}$ computed from Equation (5), the solid line is α^*_{SCS} from Equation (7) and the dash-dotted line is α^*_{ICS} from Equation (1). We have also computed $\alpha^*_{\mathrm{BE,K}}$ for larger values of K (which lead to a better approximation of the bit-error-probability minimizing value of α). Because these were hardly distinguishable from $\alpha^*_{\mathrm{BE,0}}$, we have only plotted the latter. We conclude that the bit error probability minimizing value $\alpha^*_{\mathrm{BE,0}}$ is very close to the capacity maximizing value α^*_{SCS} as derived by Eggers.

4.2. The uniform noise case

For the case of uniform noise, we can explicitly minimize the 0^{th} order approximation of the bit error probability P_{BE}^0, as given by Equation (6). We distinguish the two cases $\gamma \ge 2$ and $\gamma \le 2$. For reasons of space we will only state the result and omit the derivation.

In case $\gamma \ge 2$, $P_{\mathrm{BE}}^0 = 0$ for all $\alpha \ge \frac{\gamma}{2\gamma-2}$. In case $\gamma \le 2$, we have to deal with the different cases in Equation (6) separately. From the analysis it follows that the minimal bit error probability in this case equals $P_{\mathrm{BE}}^0 = (2-\gamma)/4$, which is obtained for $\alpha^*_{\mathrm{BE}} = \gamma/2$. Summarizing:

$$\alpha^*_{\mathrm{BE}} = \frac{\gamma}{2}, \text{ if } \gamma \le 2, \qquad (8)$$

$$\alpha^*_{\mathrm{BE}} \ge \frac{\gamma}{2\gamma-2}, \text{ if } \gamma \ge 2. \qquad (9)$$

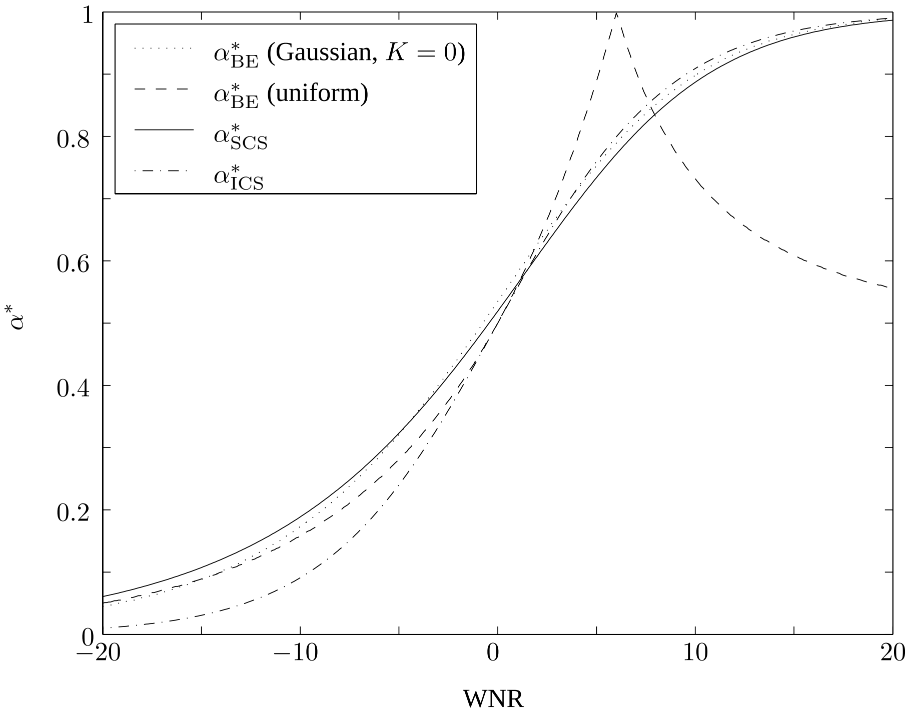

Fig. 2. The optimal values of the distortion compensation parameter

This optimal distortion compensation parameter for the uniform case has been plotted in Figure 2, as well (dashed line).

Note that for low watermark-to-noise ratios the optimal value of α depends linearly on the quotient $\gamma = \sigma_w/\sigma_n$. In the Gaussian noise case, above, there is approximately a linear dependence, as well, as for low watermark-to-noise ratios (7) reduces to $\alpha^*_{\mathrm{BE}} \approx \gamma/\sqrt{2.71}$.

For large watermark-to-noise ratios a continuum of values of alpha optimizes the bit error probability. This region corresponds to the situation where the bit error probability equals 0. In practice, one would probably choose α close to 1 in this case, as that would be lead to a better performance in case non-uniformly distributed noise would be present (as indicated by the curve for the Gaussian case).

5. CONCLUSIONS

In this paper we have derived a model for the bit error probability for the watermarking system known as the Scalar Costa Scheme. Expressions were derived both for the case of additive Gaussian noise and of additive uniform noise. The model was subsequently used to compute the optimal value of the distortion compensation parameter α, where the bit error probability was used as optimality criterion. The result was compared with approaches that select the distortion compensation parameter such that the capacity of the watermarking system is maximized. Our bit error probability minimizing values appear to be very similar to these capacity maximizing values.

6. REFERENCES

[1] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[2] M.H.M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[3] J.J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE transactions on Signal Processing*, vol. 51, no. 3, pp. 1003–1019, 2003,

[4] S.P. Lipschitz, R.A. Wannamaker, and J. Vanderkooy, "Quantization and dither: a theoretical survey," *Journal of the audio engineering society*, vol. 40, no. 5, pp. 355–375, 1992.