MODEL-BASED DEADZONE OPTIMIZATION FOR STACK-RUN AUDIO CODING WITH UNIFORM SCALAR QUANTIZATION

Marie Oger[1]*, Stéphane Ragot*[1]*, and Marc Antonini*[2]

[1]France Télécom R&D/TECH/SSTP, Av. Pierre Marzin, 22307 Lannion Cedex
[2]Lab. I3S-UMR 6070 CNRS and Univ. of Nice Sophia Antipolis, rte des Lucioles, 06903 Sophia Antipolis
E-mail: {marie.oger, stephane.ragot}@orange-ftgroup.com, am@i3s.unice.fr

ABSTRACT

In this paper we present a model-based method to optimize the deadzone of uniform scalar quantization in transform audio coding. The input signal is coded in modified discrete cosine transform (MDCT) domain by uniform scalar quantization followed by context-based arithmetic coding. The optimal deadzone is derived using a non-asymptotic method, assuming that the distribution of MDCT coefficients is approximated by a generalized Gaussian model. We show that deadzone optimization improves slightly quality, especially at low bitrates.

Index Terms— Transform coding, audio coding.

1. INTRODUCTION

Using a deadzone for scalar quantization is well-known to improve the performance of audio or image coding. For example a deadzone is used in ITU-T G.722.1 Recommendation [1], MPEG audio standards (where it is related to the so-called "magic number") or in JPEG2000 [2]. The main contribution of this work lies in the application of generalized Gaussian model to optimize the deadzone for scalar quantization. The inclusion of a deadzone for quantization was studied in [3] for Laplacian distribution. It has shown that under high rate assumption the optimal deadzone z is close to the stepsize q. In the case of low bitrate for a Laplacian distribution the optimal deadzone z is two times the stepsize q [4].

This paper is organized as follows. We present the principle of deadzone optimization based on generalized Gaussian model in Section 2. Then the proposed coder is presented in Section 3. Objective and subjective quality results are presented in Section 4 before concluding in Section 5.

2. DEAD ZONE OPTIMIZATION BASED ON GENERALIZED GAUSSIAN MODEL

2.1. Preliminary: generalized Gaussian model

The probability density function (pdf) of a zero-mean generalized Gaussian random variable x of standard deviation σ is given by [5]:

$$g_{\sigma,\alpha}(x) = \frac{A(\alpha)}{\sigma}\, e^{-|B(\alpha)x/\sigma|^{\alpha}}, \qquad (1)$$

This work was supported in part by the European Union under Grant FP6-2002-IST-C 020023-2 FlexCode.

where α is a shape parameter describing the exponential rate of decay and the tail of the density function,

$$A(\alpha) = \frac{\alpha B(\alpha)}{2\Gamma(1/\alpha)} \quad \text{and} \quad B(\alpha) = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}, \qquad (2)$$

with

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha+1}\, dt. \qquad (3)$$

The special cases $\alpha = 1$ and 2 correspond to the Laplacian and Gaussian distributions respectively. In order to estimate the shape parameter α we use a method proposed by Mallat [6].

2.2. Deadzone optimization based on a generalized Gaussian model

We consider the encoding of N zero-mean random variables x_i of variances σ^2 with respect to the mean square error criterion. We assume that the variables x_i have a generalized Gaussian pdf $g_{\sigma,\alpha}(x_i)$ of shape parameter α. The variables x_i are coded by scalar quantization with the *same* step size q. For a given bit allocation R in bits per sample, the bit allocation problem is to minimize the distortion D under the constraint that $\sum_{i=1}^{N} b_i \leq R$. Solving this problem is the minimization of a function with Lagrangian techniques. The criterion $J(z, q, \lambda)$ is defined as:

$$J(z, q, \lambda) = D\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right) - \lambda\left(b\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right) - R\right) \qquad (4)$$

where λ is the Lagrange multiplier.

The quantization mean square error D_Q resulting for the encoding of N random variables x_i is given by [7]:

$$D_Q = \int_{-z/2}^{z/2} x^2 g_{\sigma,\alpha}(x)dx + 2\sum_{m=1}^{+\infty} \int_{z/2+(m-1)q}^{z/2+mq} (x - \hat{x}_m)^2\, g_{\sigma,\alpha}(x) \qquad (5)$$

where $\hat{x}_m$ is the reconstruction level of each quantization level m. Here, we consider the special case of a reconstruction level set to mid-value so:

$$\hat{x}_m = \frac{z}{2\sigma} + (m - \frac{1}{2})\frac{q}{\sigma} \qquad (6)$$

After simplifying we have the following relationship:

$$\begin{aligned} D_Q &= \sigma^2 + 2\sum_{m=1}^{+\infty} \hat{x}_m^2 \int_{-z/2+(m-1)q}^{z/2+mq} g_{\sigma,\alpha}(x)dx \\ &\quad -4\sum_{m=1}^{+\infty} \hat{x}_m \int_{-z/2+(m-1)q}^{z/2+mq} x g_{\sigma,\alpha}(x)dx \end{aligned} \qquad (7)$$

By using Eq. 6 we can write that:

$$D_Q = 2\sigma^2 \sum_{m=1}^{+\infty} \left(\frac{z}{2\sigma} + \left(m - \frac{1}{2} \right) \frac{q}{\sigma} \right)^2 f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) - 4\sigma^2 \sum_{m=1}^{+\infty} \left(\frac{z}{2\sigma} + \left(m - \frac{1}{2} \right) \frac{q}{\sigma} \right) f_{1,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) + \sigma^2 \quad (8)$$

where $f_{n,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right)$ is a function defined as:

$$f_{n,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) = \int_{z/2\sigma+(m-1)q/2\sigma}^{z/2\sigma+mq/2\sigma} x^n g_{1,\alpha}(x) dx \quad (9)$$

So the mean square error D_Q is a function of the stepsize q, the deadzone z, the shape parameter α and the variance σ^2. The distortion is defined as [5]:

$$D\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) = \frac{D_Q}{\sigma^2} \quad (10)$$

The bit rate is defined as:

$$b = -\sum_{m=-\infty}^{+\infty} p(m) \log_2 p(m) \quad (11)$$

where $p(m)$ is the probability of having the quantization level m. The generalized Gaussian distribution is symmetrical so we have the relationship $p(m) = p(-m)$ and finally:

$$b = p(0) \log_2 p(0) - 2 \sum_{m=1}^{+\infty} p(m) \log_2 p(m) \quad (12)$$

where $p(m)$ is defined as:

$$p(m) = \int_{z/2+(m-1)q}^{z/2+mq/2} g_{1,\alpha}(x) dx = f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) \quad (13)$$

The bit rate is also a function of stepsize q, deadzone z, shape parameter α and variance σ^2. Finally, we have the relationship [5]:

$$b\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) = -f_{0,0}\left(\alpha, \frac{z}{\sigma} \right) \log_2 f_{0,0}\left(\alpha, \frac{z}{\sigma} \right) - 2 \sum_{m=1}^{+\infty} f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) \log_2 f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right) \quad (14)$$

It can be shown that the optimal dead zone z is given by the solution to the equation [5]:

$$\frac{\frac{\partial D}{\partial z}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right)}{\frac{\partial b}{\partial z}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right)} = \frac{\frac{\partial D}{\partial q}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right)}{\frac{\partial b}{\partial q}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma} \right)} \quad (15)$$

From Eq. 15 we have a derivative systems from which it is possible to extract a relationship between $\ln(q/\sigma)$ and z/q. So for practical implementation, we stores tables of this relationship for different values of the shape parameters α.

Fig. 1 presents charts in order to have the optimal deadzone z depending of the shape parameter α, the stepsize q and the variance σ. As we can see, as the stepsize is getting smaller, which means for high bitrate, the deadzone is equal to the stepsize. For lower bitrate, the size of the deadzone increases. For typical audio and speech bitrate coding, $ln(q/\sigma)$ is between 1 and -2 which give a ratio z/q between 1 and 2.

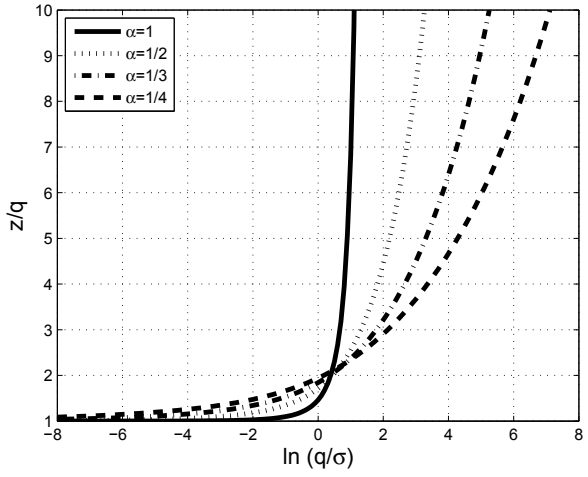

Fig. 1. Optimal deadzone for a uniform scalar quantizer with centroid set to mid-value (assuming a generalized Gaussian model).

3. PROPOSED CODING STRUCTURE

3.1. Encoder

The proposed encoder is illustrated in Fig. 2. The input sampling frequency is 16000 Hz, while the frame length is 20 ms with a lookahead of 25 ms. The effective bandwidth of the input signal $x(n)$ is 50-7000 Hz. Weighting and transform on $x(n)$ are the same as described in [6]. They consist of linear-predictive weighting followed by modified discrete cosine transform (MDCT) and low-frequency pre-shaping. The distribution of the spectrum $X_{pre}(k)$ is approximated by a generalized Gaussian model and Mallat's method [8] is used to estimate the shape parameter α. The deadzone optimization is based on the pdf of $X_{pre}(k)$ as described in Section 2. The pre-shaped spectrum $X_{pre}(k)$ is divided by stepsize q and the resulting coefficients $Y(k)$ are encoded by scalar quantization with deadzone:

$$\tilde{Y}(k) = \begin{cases} X_{pre}(k)/q - (z-q)/2q & \text{if } X_{pre}(k) > z/2 \\ X_{pre}(k)/q + (z-q)/2q & \text{if } X_{pre}(k) < -z/2 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Only the first 280 coefficients of the $Y(k)$ spectrum corresponding to the 0-7000 Hz band are coded; the last 40 coefficients are discarded. The rate control consist in finding the appropriate step size q so that the number of bits, $nbit$, used for stack-run coding matches the allocated bit budget as described in [6]. Finally, a noise estimation is performed on the spectrum $Y(k)$ after stack-run coding. The noise floor σ is estimated as in [6]:

$$\sigma = r.m.s.\left\{ X_{pre}(k) \,|\, Y(k) = 0 \right\} \quad (17)$$

The step size q is scalar quantized in log domain with 7 bits. The noise floor σ is quantized by coding the ratio $\sigma/\hat{q}$ in linear domain with 3 bits. In the case of stack-run coding with $z = z_{opt}$, the ratio z/q is quantized in linear domain with 2 bits. Otherwise for stack-run coding with $z = q$ or $z = 2q$ we don't need to transmit the ratio z/q.

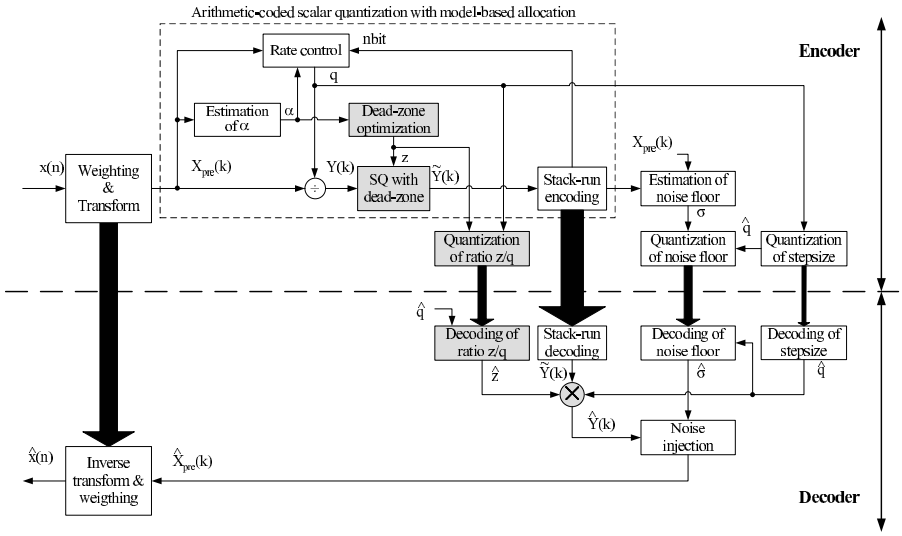

Fig. 2. Block diagram of the proposed predictive transform coder with deadzone.

3.2. Decoder

The decoder in error-free conditions is illustrated in Figure 2. The reconstructed spectrum $\hat{Y}(k)$ is given by:

$$\hat{Y}(k) = \begin{cases} \hat{q}\tilde{Y}(k) + (\hat{z} - \hat{q})/2 & \text{if } \tilde{Y}(k) > 0 \\ \hat{q}\tilde{Y}(k) - (\hat{z} - \hat{q})/2 & \text{if } \tilde{Y}(k) < 0 \\ 0 & \text{if } \tilde{Y}(k) = 0 \end{cases} \quad (18)$$

where $\tilde{Y}(k)$ is found by stack-run decoding, $\hat{z}$ and $\hat{q}$ are respectively the decoded deadzone and the decoded stepsize. The spectrum $\hat{X}(k)$ is de-shaped by using an inverse weighting and transform presented in [6].

3.3. Bit allocation

The parameters of the proposed coder are line spectrum frequency (LSF) parameters, step size q, and noise floor level σ. The ratio z/q is transmitted to the decoder only for scalar quantization with optimal deadzone. The bit allocation to the parameters is detailed in Table 1, where B_{tot} is the total number of bits per frame. For instance at 24 kbit/s, $B_{tot} = 480$ bits. The allocation (in bits per sample) to stack-run coding with deadzone scalar quantization is $B = (B_{tot} - 52)/280$.

Table 1. Bit allocation for the coding scheme.

Parameter	Number of bits
LSF	40
Step size q	7
Noise floor σ	3
Stack-run coding with $z = q$	B_{tot}-50
Stack-run coding with $z = 2q$	B_{tot}-50
Ratio z/q	2
Stack-run coding with $z = z_{opt}$	B_{tot}-52
Total	B_{tot}

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this work we used the same experimental setup as in [6]. A database of 24 clean speech samples in French language (6 male and female speakers×4 sentence-pairs) and 16 clean music samples (4 types×4 samples) of 8 seconds is used. These samples are sampled at 16 kHz, preprocessed by the P.341 filter of ITU-T G.191A and normalized to -26 dB$_{ov}$ using the P.56 speech voltmeter.

4.1. Optimization of the dead zone

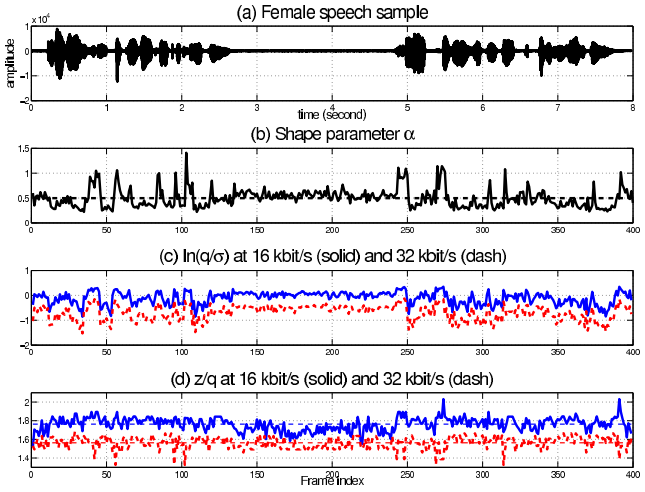

Fig. 3. Example of dead-zone optimization.

We presented in Fig. 3 an example of the dead-zone optimization with a centroid at middle-value for a French female speaker sample of 8 seconds at two bitrates 16 and 32 kbit/s. As we can see in Fig. 3 (c) $ln(q/\sigma)$ at 32 kbit/s is smaller than the one at 16

kbit/s because the stepsize at high bitrate is smaller than the one at low bitrate. Also in Fig. 3 (d) the mean value of $z/q \approx 1.6$ at 32 kbit/s and the mean value of $z/q \approx 1.8$ at 24 kbit/s which confirm the theory that at high bitrate z/q is getting closer to 1.

4.2. Objective quality results

WB-PESQ [9] is used to evaluate the quality of the proposed coder and compare it with ITU-T G.722.1. Only clean speech samples are used to compute the average WB-PESQ scores at various bitrates. The bit rate varies from 16 to 40 kbit/s with a step of 4 kbit/s for our coder. ITU-T G722.1 is tested at 24 and 32 kbit/s.

Fig.4 shows the WB-PESQ scores obtained for the two coders As we can see, using a scalar quantizer with an optimal dead-zone $z = z_{opt}$ or a dead-zone equal to two times the stepsize $z = 2q$ improves the performance at low bitrate. It seems that having $z = z_{opt}$ or $z = 2q$ is equivalent, it could be explain by the fact that we need two bits to transmit the optimal deadzone z_{opt} which is not the case if $z = 2q$ or $z = q$.

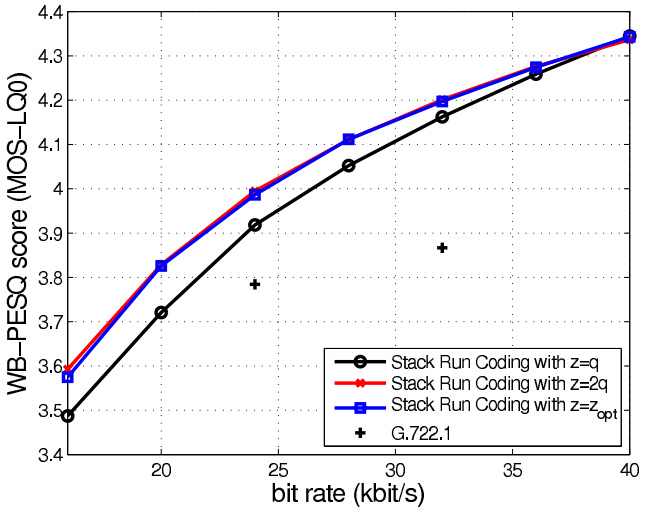

Fig. 4. Average WB-PESQ score (centroid set to mid-value).

4.3. Subjective quality results

An informal AB tests at 24 kbit/s has been conducted for speech in order to compare the stack-run coding with or without deadzone . In total 9 experts participated in the test. Fig. 5 shows the results. Stack-run coding with $z = z_{opt}$ was preferred for speech in 50% of cases. The results confirmed the objective quality results at 24 kbit/s. Subjective tests have also been conducted at 32 kbit/s for speech and music and the two coders are equivalent. Consequently, the use of an optimized deadzone z_{opt} does improve slightly quality, especially at low bitrates.

4.4. Complexity

The algorithmic complexity of stack-run coding is 45 ms (20 ms for the frame, 20ms for the MDCT and 5 ms for the lookahead), while that of G.722.1 is 40 ms. The computational complexity of G.722.1 is low which is not the case with the stack-run coding. Indeed in the latter case, a rate control loop is needed [6] and in practice stack-run coding is performed several times per frame. The memory requirements (in terms of data ROM) for the stack-run coding consists mainly of the storage of GMM parameters for LPC quantization and MDCT computation tables. To compute the optimal deadzone z_{opt} for each MDCT frame we have to store the tables which gives the relationship between $\ln(q/\sigma)$ and z/q.

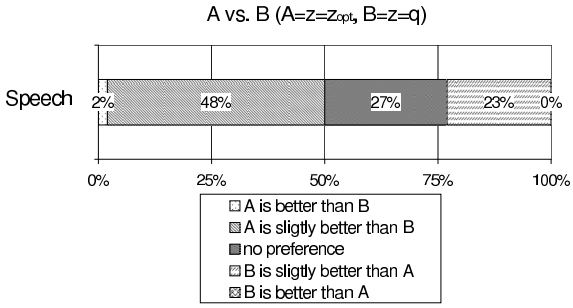

Fig. 5. AB test results for speech at 24 kbit/s.

5. CONCLUSION

In this paper we proposed a non-asymptotic method to have optimal deadzone for scalar quantization, assuming that the distribution of MDCT coefficients is approximated by a generalized Gaussian model. In fact, stack-run coding with $z = 2q$ is near-equivalent to $z = z_{opt}$. This result relies on the assumption of generalized Gaussian modeling and the use of ideal entropy coding. Still, we can consider that $z = 2q$ is a general solution for scalar deadzone optimization and it is not specific to stack-run coding. Finally this result confirms [3]. However we don't have a Laplacian distribution and we don't assume high bitrate.

REFERENCES

[1] ITU-T G.722.1, *Coding at 24 kbit/s and 32 kbit/s for Hand-free Operations in Systems with Low Frame Loss*, 1999.

[2] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Springer, 2001.

[3] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Trans. on Information Theory*, vol. 42, no. 5, pp. 1365–1374, sept 1996.

[4] S. G. Mallat, "Analysis of low bit rate image transform coding," *IEEE Trans. on Signal Proc.*, vol. 46, pp. 1027–1042, April 1998.

[5] C. Parisot, M. Antonini, and M.Barlaud, "3d scan based wavelet transform and quality control for video coding," *EURASIP*, vol. 1, pp. 521–528, Jan 2003.

[6] M. Oger, S. Ragot, and M. Antonini, "Transform audio coding with arithmetic-coded scalar quantization and model-based bit allocation," *Proc. ICASSP*, May 2007.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1993.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, pp. 674–693, Jul 1989.

[9] ITU-T Rec P.862.2, *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, Nov 2005.