FACTA UNIVERSITATIS (NIŠ)
SER.: ELEC. ENERG. vol. 17, August 2004, 165-184

Spherical Logarithmic Quantization and its Application for DPCM

Dedicated to Prof. Dr. Karlheinz Tröndle on the occasion of his retirement and 65th birthday

Johannes B. Huber and Bernd Matschkal

Abstract: A new method for efficient digitizing analog signals while preserving the original waveform as close as possible with respect to the relative quantization error is presented. Logarithmic quantization is applied to short vectors of samples represented in sphere coordinates. The resulting advantages, i.e. a constant Signal-to-Noise Ratio over a very high dynamic range at a small loss with respect to rate-distortion theory are discussed. In order to increase the Signal-to-Noise Ratio (SNR) by exploitation of correlations within the source signal, a method of combining differential pulse-codemodulation (DPCM) with spherical logarithmic quantization is presented. The resulting technique achieves an efficient digital representation of waveforms with a high longterm as well as segmental SNR at an extrem low delay of the signal.

Keywords: DPCM, spherical logaritmic quantization, quantization noise, signal-to-noise ratio, gradient descent algorithm.

1 Introduction

For efficient and robust transmission, processing or storage of analog waveforms, digital representation offers lots of advantages. For this purpose the analog signal usually is quantized and digitized. The development towards digital representation of analog signals had been pioneered in Germany by the textbook on pulsecode-modulation (PCM) of K. Tröndle and Weiß [2] in early days. After PCM-encoding usually data compression methods are applied exploiting redundancy within the

Manuscript received March 17, 2004. This paper is an extended reprint of a paper presented at the 5th International ITG Conference on Source and Channel Coding, Jan. 14-16, 2004, see [1].

The authors are with the Institute for Information Transmission, University of Erlangen–Nuremberg, Germany, www.LNT.de, (e-mail: [huber,matschkal]@LNT.de).

source signal and irrelevance due to special properties of the consumer of the signal. By this, efficient digital transmission or storage at a tolerable data rate is combined with a sufficient quality of the reconstructed analog waveform.

The numerous different methods for digitizing analog signals may be subdivided into two basic categories:

i) methods where the reconstructed waveform approximates the original one closely, i.e. no exploitation of irrelevance. Here we use the expression "waveform conserving" instead of the often applied term "lossless" waveform coding, because digitizing analog signals at a finite data rate is not "lossless" in principle (i.e. infinite entropy of a continuous random variable).

ii) non-waveform conserving methods. These approaches are very important in audio signal and video compression, as long as only the subjective quality of the signal at the receiver output is relevant (e.g. exploitation of psychoacoustic masking). Usually signal processing for exploitation of irrelevance produces a reconstructed waveform quite far away from the original one and additionally introduces a high signal delay (e.g. due to spectral transforms or equivalent block based operations). The quality of these waveform coding methods cannot be expressed by a Signal-to-Noise Ratio (SNR) in the classical sense; it rather has to be determined in complex performance tests by well-trained persons. But in many applications, such as processing of measurement data, recording of waveforms for further signal processing, real time signal transmission using digital modulation schemes without tolerance of a noticeable signal delay as e.g. for cordless digital stage microphones, non-waveform conserving signal coding schemes are completely useless.

In this paper a new waveform conserving method, spherical logarithmic quantization, is proposed in order to meet the following requirements: a) low data rate by exploitation of favorite packing properties of multidimensional lattices (vector quantization), b) extreme high dynamic range, i.e. preservation of a constant high SNR for variations of 60 dB and more of the short time variance of the analog source signal, c) high segmental SNR for short segments of waveform samples, d) insensitivity to special signal parameters such as segmental probability density function etc., and most important e) introduction of an extreme low signal delay in the order of a few (up to 10) samples.

2 Logarithmic Quantization

The intention of logarithmic quantization is a high dynamic range. This implies a wide range of the signal level where the SNR and thus the maximum relative

quantization error $(\Delta q_i/2)/|r_i|$ is constant. Here, let Δq_i denote the width of the i-th quantization interval and r_i the corresponding reconstruction value. This leads us to the well-known logarithmic quantization e.g. according to the $\mathscr{A}$-law, which is in use for many years in telecommunications [2]. Regarding a medium signal level and $R \gg 1$ the SNR can be expressed by:

$$10\log_{10}(\mathrm{SNR}) = R \cdot 6.02\ \mathrm{dB} + 10\log_{10}\frac{3}{\left(1+\ln(\mathscr{A})\right)^2} \tag{1}$$

where R denotes the average bits per sample (rate) and $\mathscr{A}$ the usual parameter of this logarithmic quantization. This quantization provides a constant relative maximum quantization error for samples with an absolute value $\geq 1/\mathscr{A}$ (referred to a signal range of the quantization from -1 to +1). The parameter $\mathscr{A}$ specifies the quotient of the maximum to the minimum width of quantization intervals. Therefore, signals with power levels greater than

$$-B_1 := 20\log_{10}\frac{1}{\mathscr{A}}\ \mathrm{dB} \tag{2}$$

are quantized and reconstructed at a SNR given by (1). B_1 characterizes the dynamic range of scalar logarithmic quantization. Within this dynamic range, the SNR is independent of the probability density function (pdf) of the signal which provides an universal applicability of this method. Comparing (1) to the result for an uniformly distributed source signal and using an uniform quantizer or to the rate distortion function for independent, identically distributed (i.i.d.) Gaussian random variables, the term $10\log_{10}\left(3/(1+\ln(\mathscr{A}))^2\right)$ represents a "SNR-loss" by companding, which has obviously to be paid for a wide dynamic range.

3 Quantization in Sphere Coordinates

3.1 Sphere coordinates in *D* dimensions

Spherical logarithmic quantization belongs to the family of vector quantization methods (of e.g. [3, 4]). It is well known that the property of dense sphere packing in multidimensional lattices offers some gains for quantization, even if there are no statistical interdependencies within the samples combined to a vector. A vector $\boldsymbol{x} := (x_1, \ldots, x_D)$ of D samples in Cartesian coordinates at first is expressed by polar coordinates $\boldsymbol{u} := (\varphi_1, \ldots, \varphi_{D-1}, r)$. The $D-1$ angles φ_i as well as the radius r are given by the following equations (j: imaginary unit, $\arg(\cdot)$: argument-function delivering the angle of a complex number in rad):

$$\varphi_1 = \arg(x_1 + jx_2) \in [-\pi, +\pi) \tag{3}$$

$$\varphi_i = \arg\left(\sqrt{\sum_{l=1}^{i} x_l^2} + jx_{i+1}\right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right), \quad i \in \{2, \ldots, D-1\}, \tag{4}$$

$$r^2 = \sum_{l=1}^{D} x_l^2 \tag{5}$$

The reconstruction of the Cartesian components from vector $\boldsymbol{u}$ is obtained from

$$x_i = r \cdot b_{i-1} \cdot \sin(\varphi_{i-1}), \quad i \in \{D, ..., 2\} \tag{6}$$

$$x_1 = r \cdot b_1 \cdot \cos(\varphi_1) = r \cdot b_0 \tag{7}$$

with the radii b_i of the "circles of latitude" of a unit sphere (radius 1):

$$b_{D-1} = 1 \tag{8}$$

$$b_i = \prod_{l=i+1}^{D-1} \cos(\varphi_l), \quad i \in \{D-2, ..., 0\}\ . \tag{9}$$

The recursions for this transformations forward and backward may be efficiently implemented using the well-known CORDIC-algorithm [5] with low complexity.

3.2 Spherical logarithmic quantization

In contrast to usual vector quantization, individual, i.e. mutually independent, quantization of the polar coordinates is applied for saving complexity. This is because, under the requirements stated in section 1, a "true vector quantization" would offer only marginal further gains which we waive in favour of reduced complexity. In order to preserve the properties of a logarithmic quantization, i.e. independence of the SNR of the signal variance and its special (short time) pdf, we propose to apply usual logarithmic quantization according to the $\mathscr{A}$-law for the radius (magnitude). For the angle variables φ_i we simply apply individual uniform quantization but with quantization intervals being functions of the quantized angle variables $\widehat{\varphi}_l$, $l \in \{i+1, i+2, ..., D-1\}$ of higher orders, used for reconstruction.

Thus, a very simple implementation of the quantization and signal reconstruction is achieved, where the iterative procedure according eqs. (3) to (9) allows the processing of the coordinates step by step in the same way as in the case of a scalar quantization.

Because of the proportionality of a circular arc segment to the radius, the feature of logarithmic quantization, i.e. the proportionality of the width of a quantization

interval to the signal value, is already present for uniform quantization of the angles, see Fig. 1 for a simple example of two dimensions ($D = 2$). Thus, the second term in eq. (1) for logarithmic quantization has to be paid only for one out of D dimensions. This is the essential source of the gains due to spherical logarithmic quantization.

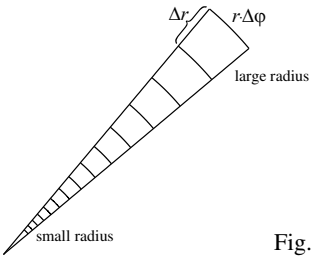

Fig. 1. Proportionality of a circular arc segment to the radius

The uniform quantization of the angles corresponds to a quantization of the surface of a D-dimensional sphere roughly into $(D-1)$-dimensional (hyper-) cubes (e.g. ordinary squares for $D = 3$) as long as the number of quantization cells is very large.

We will show in section 3.3 that this suboptimal quantization of the surface of the sphere results in a loss of only $10\log_{10}(\pi e/6) = 1.53$ dB for $D \to \infty$, which corresponds to a rate loss of about 1/4 bit/sample, which we accept for the benefit of low complexity.

In the following, we will discuss how to split the M^D (i.e. $M := 2^R$) quantization levels (which are available per quantization step) into quantization intervals for the radius r and the surface of a D-dimensional unit (hyper-)sphere in an optimum way.

In the logarithmic area $r_0/\mathscr{A} \le r \le r_0$ of $\mathscr{A}$-Law companding the following compression function is applied for the radius as usual:

$$k(r) = r_0 \left(c \cdot \ln \frac{r}{r_0} + 1 \right) \quad \text{with} \quad c := \frac{1}{1 + \ln \mathscr{A}} . \tag{10}$$

Its derivative is:

$$k'(r) := \frac{\mathrm{d}k(r)}{\mathrm{d}r} = \frac{c \cdot r_0}{r} . \tag{11}$$

As mentioned above, $\mathscr{A}$ is the parameter of the $\mathscr{A}$-Law and r_0 describes a scale factor which will be fixed in the next paragraph. If we assume M_D quantization intervals for the radius (D^{th} component of the vector $\boldsymbol{u}$), the width of a quantization cell in radial direction reads

$$\Delta r(r) \approx \frac{r_0}{M_D \cdot k'(r)} = \frac{1}{M_D \cdot c} \cdot r . \tag{12}$$

Notice, that $\Delta r(r)$ is independent of r_0 in the considered logarithmic area.

In order to achieve a similar performance of the quantization in polar coordinates as in cartesian coordinates with respect to overloading the quantizer, we normalize the maximum value r_0 of the radius in that manner that the volume of the corresponding D-dimensional sphere equals the volume of a D-dimensional cube with edge length 2 (quantization range $x_i \in [-1;1]$ in every dimension), i.e. $V_{\text{sphere}} \stackrel{!}{=} V_{\text{cube}}$:

$$\alpha_D \cdot r_0^D \stackrel{!}{=} 2^D$$

with the volume α_D of the unit sphere in D dimensions [6]:

$$\alpha_D = \frac{\pi^{\frac{D}{2}}}{\Gamma(\frac{D}{2}+1)}. \tag{13}$$

Here $\Gamma(\cdot)$ denotes the usual Gamma-Function $(\Gamma(x+1) = x!,\ x \in \mathbb{R}^+)$. This leads to:

$$r_0 = \frac{2}{\alpha_D^{\frac{1}{D}}} \quad \text{with} \quad r_0 \geq 1 \quad \forall D \in \mathbb{N}. \tag{14}$$

Simplifying the calculation of the SNR, we chose $r = 1$ in the following, i.e. we consider the quantization cells covering the surface of the unit sphere. This is possible without loss of generality because the SNR does not depend on the radius within $r_0/\mathscr{A} \leq r \leq r_0$ due to its logarithmic quantization and the natural scaling of arc segments with the radius. The width of the quantization cells with respect to the quantization of the radius here reads:

$$\Delta := \frac{1}{M_D \cdot c} \tag{15}$$

By uniformly quantizing the angles, the surface of the D-dimensional unit sphere is divided into M_φ cells resembling $(D-1)$-dimensional cubes as long as the number of cells is high. The surface of a sphere in D dimensions is given by [6]:

$$S = \beta_D \cdot r^{D-1} \quad \text{with} \quad \beta_D = D \cdot \alpha_D\,. \tag{16}$$

For a fair partitioning of the number of M^D quantization cells for radius and surface of the unit sphere following considerations hold: regarding our approximation of cubic quantization cells, i.e. enforcing an approximately equal width of quantization cells in all dimensions (radius and circular arc segments), the surface of the unit sphere has to be subdivided into M_φ equal cubes in $(D-1)$ dimensions with the same edge length Δ as the radius for $r = 1$:

$$S = \beta_D \stackrel{!}{=} M_\varphi \cdot \Delta^{D-1}. \tag{17}$$

This construction establishes a fair split of quantization intervals between radius and angles. As we have $M = 2^R$ intervals per sample, additionally

$$M_\varphi \cdot M_D \stackrel{!}{=} M^D \tag{18}$$

has to be claimed.

With (15), (17) and (18) the number of intervals that are available for the quantization of the radius, can now be represented by

$$M_D := M \cdot \frac{1}{\beta_D^{\frac{1}{D}}} \frac{1}{c^{\frac{D-1}{D}}}. \tag{19}$$

In Fig. 2 quotients $M_D/(M/2)$ of the number of quantization cells devoted to the radius normalized to the number of cells per absolute value of a sample in scalar quantization $(D = 1)$ are given versus number of dimensions for several examples of the parameter $\mathscr{A}$ of the $\mathscr{A}$-law companding. The values of $\mathscr{A}$ are chosen in that way, that applying $\mathscr{A}$-law companding results in smallest quantization intervals being by factor $2^{\Delta n}$, $\Delta n \in \mathbb{N}$ smaller than for uniform quantization at the same entire number of intervals:

$$\Delta r\,(r \leq 1/\mathscr{A}) = 2^{-\Delta n} \frac{1}{M_D}, \tag{20}$$

i.e.

$$2^{\Delta n} = \frac{\mathscr{A}}{1 + \ln \mathscr{A}}. \tag{21}$$

Thus, resolution for small values of r is increased by Δn bit.

Fig. 2 shows that spherical logarithmic quantization devotes up to more than factor 8 more intervals to the radius than a scalar quantization with equivalent resolution of very small signal values! Thus, we find a gain of 3 bit/sample or 18 dB for spherical logarithmic quantization in this area!

Using (15) we get for the width of the quantization cells on the surface of the unit sphere:

$$\Delta = \frac{1}{M} \cdot \left(\frac{\beta_D}{c} \right)^{\frac{1}{D}}. \tag{22}$$

The constant c is determined by the selected dynamic range, see eqs. (2) and (10).

The actual uniform quantizations of the angle variables for equal circular arc segments on the unit sphere follow from the recursion eqs. (8) and (9). We have M_i

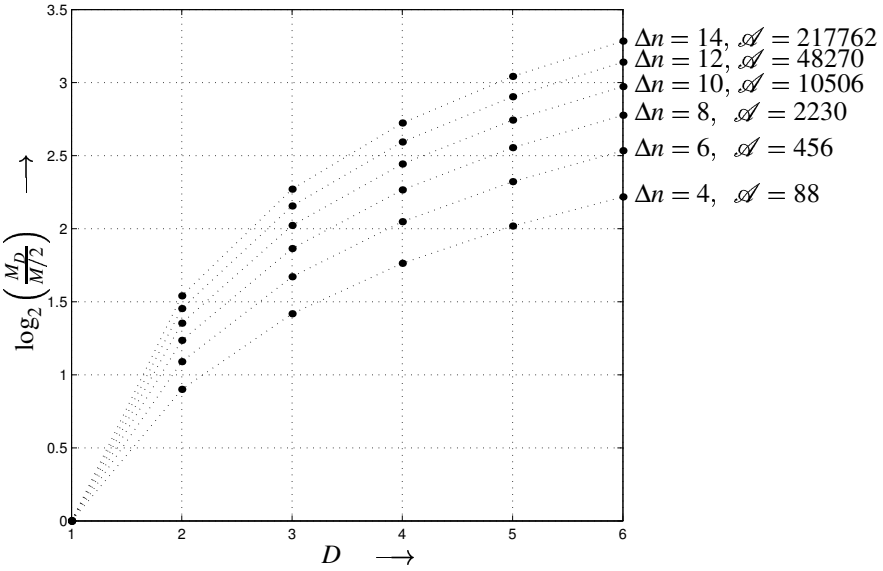

Fig. 2. Number of quantization cells devoted to the radius normalized to the number of calls per absolute value of a sample in scalar quantization

intervals for the coordinate φ_i:

$$M_{D-1} = \left\lfloor \frac{\pi}{\Delta} \right\rfloor \tag{23}$$

$$M_i(\widehat{\varphi}_{i+1}, \ldots, \widehat{\varphi}_{D-1}) = \left\lfloor \frac{\pi \cdot \widehat{b}_i}{\Delta} \right\rfloor, \quad \text{for} \quad i = D-2, D-3, \ldots, 2 \tag{24}$$

$$M_1(\widehat{\varphi}_2, \ldots, \widehat{\varphi}_{D-1}) = \left\lfloor \frac{2\pi \widehat{b}_1}{\Delta} \right\rfloor \tag{25}$$

with $\lfloor x \rfloor \in \mathbb{N}$: largest integer $\leq x$ with $x \in \mathbb{R}^+$ and $\widehat{b}_i$ corresponding to eqs. (8) and (9) for quantized angles $\widehat{\varphi}_i$.

Please notice, that the numbers M_i of quantization intervals in dimension i is a function of the selected quantization cell in dimension $i+1, ..., D-1$ and therefore can be calculated iteratively starting with M_{D-1}. Calculation in advance is not possible.

An assignment of an index $N \in \{0, 1, ..., M^D - 1\}$ to the actual quantization cell and reconstruction may be performed by application of nested look-up tables as known from Shell Mapping, e.g., cf. [7]:

The assignment of an index to the radius is independent of the index of the cell on the unit sphere. Therefore we can concentrate on the problem to assign indices

to the cells on the unit sphere.

The starting number of the index is 0, the maximum ending index number is $M_\varphi - 1$ which equals the number of quantization cells on the surface of the unit sphere. M_{D-1} is the number of subspheres of dimensionality $D-1$ and can be obtained from eq. (23). For the benefit of a fast implementation, a first-level lookup-table with M_{D-1} entries $N_0, ..., N_{M_{D-1}-1}$ is used. Here the number N_i denotes the smallest index of all cells corresponding to the i^{th} quantization interval for angle $\widehat{\varphi}_{D-1}$:

$$\begin{aligned} N_0 &= 0 \\ N_{i+1} &= \sum_{\nu=0}^{i} M_{D-2}\left(\widehat{\varphi}_{D-1} = \nu\frac{\pi}{\Delta}\right), \quad i \in \{0,\ldots,(M_{D-1}-2)\} \end{aligned} \tag{26}$$

Thus, there exist $N_{i+1} - N_i$ cells for fixed i^{th} value of $\widehat{\varphi}_{D-1}$. For each N_i, a second-level lookup-table is calculated containing the index numbers $O_{i,j}$ denoting the smallest index of all cells corresponding to the j^{th} quantization interval for angle $\widehat{\varphi}_{D-2}$ within the $(D-1)$-dimensional subsphere which N_i is referred to.

$$\begin{aligned} O_{i,0} &= 0, \\ O_{i,j+1} &= \sum_{\nu=0}^{j} M_{D-3}\left(\widehat{\varphi}_{D-2} = \nu\frac{\pi}{\Delta}, \widehat{\varphi}_{D-1} = i\frac{\pi}{\Delta}\right), \\ & \quad i \in \{0,\ldots,(M_{D-1}-1)\}, \\ & \quad j \in \left\{0,\ldots,\left(M_{D-2}\left(\widehat{\varphi}_{D-1} = i\cdot\frac{\pi}{\Delta}\right) - 2\right)\right\} \end{aligned} \tag{27}$$

Proceeding in this manner leads to $D-2$ levels of nested lookup-tables for indexing the $D-1$ angles, because the lookup-table of level $D-1$ (angle $\widehat{\varphi}_1$) holds subsequent cell indices and therefore does not need to be tabularized.

3.3 Quantization noise and SNR

As long as there is a sufficiently high number of quantization cells in D dimensions, the usual approximation of an uniformly distributed quantization error within these cubic cells, each of it represented by its center point, may be applied. Thus, the noise variance for a signal vector with radius r is:

$$\frac{\Delta^2(r)}{12}D. \tag{28}$$

Spherical logarithmic quantization enforces $\Delta(r) = \Delta \cdot r$ in all dimensions for

$r_0/\mathscr{A} \le r \le r_0$, thus we have

$$\mathrm{SNR} = \frac{r^2}{\Delta^2 \cdot r^2 \frac{D}{12}} = F(D)M^2 \tag{29}$$

with

$$F(D) := \frac{12}{D}\left(\frac{c}{\beta_D}\right)^{\frac{2}{D}}, \tag{30}$$

see eq. (22). As designed, the SNR is independent of the variance of the signal in this area.

Inserting (13) and (16) yields:

$$F(D) = \frac{12}{\pi}\frac{1}{D^{\frac{D+2}{D}}}\left((D/2)!\right)^{\frac{2}{D}} c^{\frac{2}{D}}. \tag{31}$$

Looking at (29), $F(D)$ may be considered as the loss with regard to the rate-distortion-bound for i.i.d. Gaussian random variables (6 dB-per-bit-rule). It is presented in Fig. 3 for different values of $\mathscr{A}$.

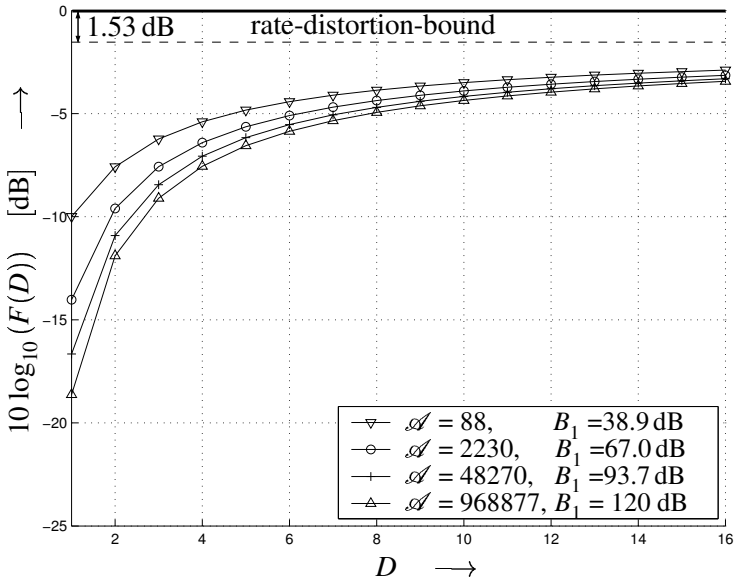

Fig. 3. SNR-loss with respect to $R \cdot 6.02$ dB versus number of dimensions for different values of $\mathscr{A}$

Using Stirling's approximation $x! \approx \sqrt{2\pi x}(x/\mathrm{e})^x$ the limit

$$\lim_{D\to\infty} F(D) = \frac{6}{\pi \mathrm{e}} \quad \widehat{=} \quad -1.53\,\mathrm{dB} \tag{32}$$

follows, which represents the loss with respect to the rate-distortion-bound caused by the suboptimal cubical quantization cells (instead of $(D-1)$-dimensional hyperspheres for quantization of the surface of a D-dimensional unit sphere and $D \to \infty$). In other words: By using spherical logarithmic quantization, it is possible to compensate the loss due to companding described in (1) up to a residual margin of 1.53 dB. Therefore, it is possible in principle to choose the parameters $\mathscr{A}$ and D in that way that an infinite dynamic range is achieved without having to accept a noticeable loss in the maximum attainable SNR.

The asymptotic SNR-loss of 1.53 dB corresponds to a rate loss of 1/4 bit/sample which we accept for the advantage of an extreme reduction in complexity. Our requirement does not allow to have detailed knowledge on the pdf of the source signal, as such a knowledge would not be compatible with the claims for a high dynamic range of the quantizer and high segmental SNRs for nonstationary source signals. On the other hand the rate-distortion-function for Gaussian i.i.d. random variables is an upper bound on the minimum achievable rate (distortion, resp.) at a given distortion (rate, resp.) (Berger's upper bound of the rate-distortion-function [8]). Therefore, further improvements beyond 1.53 dB or 1/4 bit/sample, resp., are not possible at all under the given constraints and requirements. Furthermore, the approach towards the rate-distortion-function for Gaussian i.i.d. random variables within 1.53 dB is achieved without any knowledge on the source signal. No transmission of side information for amplitude scaling etc. is necessary as long as overload is avoided, which is no problem at all because of the extreme high dynamic range of spherical logarithmic quantization.

Fig. 4 shows the distance between the SNR and the rate-distortion-bound $R \cdot 6\text{dB}$ depending on the average signal level ($10\log_{10}(\text{variance})$) of i.i.d. Gaussian random variables for the example choosing $\mathscr{A} = 48270$ versus dimensionality of the spherical logarithmic quantization. These simulation results meet exactly the theoretical analysis according to Fig. 3. At a first glance, the extreme high values for $\mathscr{A}$, chosen for the examples in Fig. 3 and 4 and additionally in the examples of section 5 may look quite unrealistic and not implementable. But please notice that usually much more than M intervals are devoted by eq. (19) to the quantization of the radius ($M_D > M/2$), see Fig. 2. Even for extreme low rates, e.g. $R = 4$ bit/sample very fine quantization intervals still exist for the radius. Additionally, eq. (10) represents an invertible function for any value of $\mathscr{A} \geq 1$ and therefore is well suited for specification of a certain nonuniform quantizer in any case. On the other hand, this approach for true waveform coding at an extreme low delay of the signal should not be applied for rates below 3 bit/sample.

Furthermore it becomes apparent from Fig. 4 that the dynamic range expands by increasing dimensionality at a constant $\mathscr{A}$ due to two effects. First, the scal-

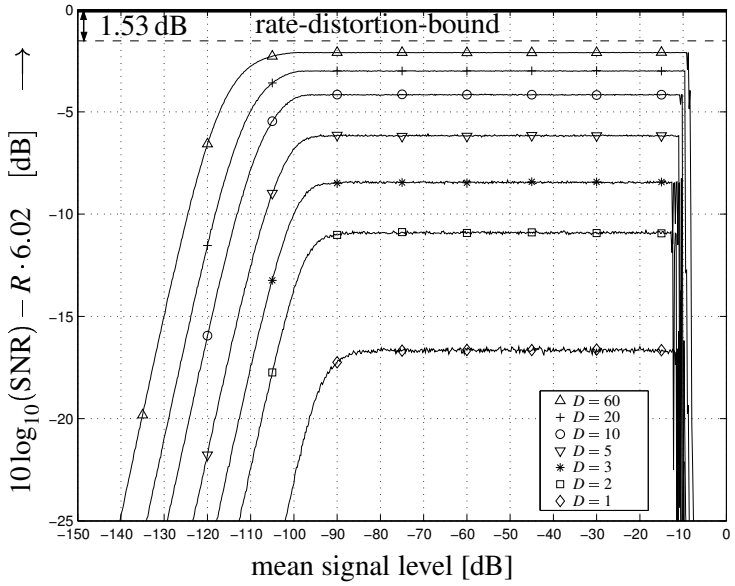

Fig. 4. SNR of i.i.d. Gaussian distributed signal values using spherical logarithmic quantization ($\mathscr{A} = 48270$, $B_1 = 93.7$ dB, $R \gg 1$ bit/sample)

ing factor r_0 increases and second, the limitations of the logarithmic compression only apply to a single dimension (the radius) whereas for the other $D-1$ dimensions the proportionality of the width of the quantization cells to the signal value is maintained even for smallest signal values. Thus the dynamic range expands in D dimensions approximately to

$$B_D \approx B_1 + 20\log_{10}(r_0) + 10\log_{10}(D)\,. \tag{33}$$

Moreover, because of the averaging effect of combining D samples, the robustness to overload increases which results in a further expansion of the dynamic range. In the limit $D \to \infty$, an infinite dynamic range is offered for any value of $\mathscr{A}$.

The structural delay of spherical logarithmic quantization is, as for any vector quantization method, exactly D samples. As to be seen from Fig. 3 and 4, the major portion of the possible gain is already achieved at very small values of D (up to 5).

4 Combination of Spherical Logarithmic Quantization and DPCM

Correlations of waveform samples $q[k]$ are efficiently exploited by differential PCM (DPCM), see Fig. 5. A prediction error signal $x[k]$ is formed by the subtraction of

predicted samples which are generated out of reconstructed waveform samples $\hat{q}[k]$ via a linear prediction filter $H_{\mathrm{P}}(z) \cdot z^{-1}$. For an ideal predictor, the prediction error sequence $x[k]$ shows a white power spectral density (psd) and minimum variance. Usually this prediction filter is adaptive in order to track a nonstationary source. Here, examples are presented only for a fixed prediction filter because we want to describe its interaction with spherical logarithmic quantization as simple as possible for high readability. Regarding audio signals for example, even very short fixed prediction error filters (designed w.r.t. a compromise criterion) usually offer average gains of more than 18 dB at a sampling frequency of 44.1 kHz. Even segmental gains, averaged over 6000 samples (0.136 s), lower than 15 dB can be observed only very rarely, see Fig. 7 in comparison with Fig. 3. Additionally, when signal delay is rather restricted to a few samples the gain due to adaptive prediction is limited, if noticeable effects generated by coefficient updates have strictly to be avoided. Of course, the subsequent material can immediately be generalized to adaptive prediction. Please notice that a SNR gain, expressed by a decrease of the mean squared

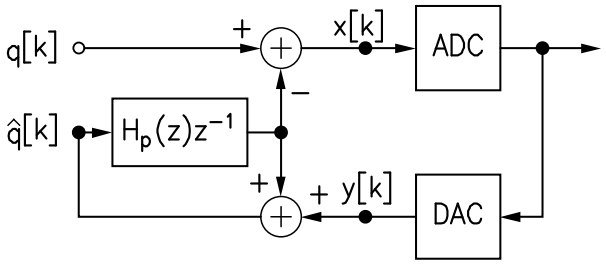

Fig. 5. Block diagram: DPCM-Encoder using backward prediction

error between original and reconstructed samples, is only possible for application of the so-called backward prediction as shown in Fig. 5. It is well known that for logarithmic quantization the SNR for the prediction error signal $x[k]$ is independent of its variance or equivalently the power of the quantization noise is proportional to the signal power, the prediction gain, i.e. the quotient of the variances of $q[k]$ and $x[k]$ is directly transformed into a SNR gain. Thus, logarithmic quantization is a favorite choice for DPCM. Additionally no further signal delay is introduced by DPCM w.r.t. PCM, because an optimum prediction error filter for maximum prediction gain is causal and strictly minimum phase in principle [9] and therefore invertible without any structural delay.

4.1 Gradient descent algorithm

For the application of spherical log. quantization to DPCM with backward prediction the same problem arises as of any vector quantization method: for calculation

of the actual prediction error sample $x[k]$ all previous reconstructed samples $\widehat{q}[k-i]$, $i = 1,2,\ldots$ have to be available and for a high prediction gain the most recent values ($i = 1,\ i = 2$) are undeniable. But this claim is inconsistent with quantizing blocks of length D of samples.

For solutions of this problem any method mentioned in literature for combining vector quantization with DPCM can be applied. Here, we address a method very closely related to the principle "analysis by synthesis" well known from CELP waveform coding methods [10] and we combine this approach with a very simple discrete step gradient descent algorithm.

In order to tackle the contradiction between DPCM and vector quantization the squared Euclidean distance between the vector of samples

$$\boldsymbol{q}[l] = (q[l \cdot D], q[l \cdot D + 1], \ldots, q[l \cdot D + D - 1]) \tag{34}$$

and a corresponding reconstruction vector $\widehat{\boldsymbol{q}}[l]$ is minimized, $l = \lfloor k/D \rfloor$. Please notice that besides spherical log. quantization, the calculation of the corresponding prediction error signal and the inversion of the prediction error filter have to be appropriately included in the computation of a pair $\boldsymbol{q}, \widehat{\boldsymbol{q}}$ whereas previously selected reconstruction vectors $\widehat{\boldsymbol{q}}[l-m]$, $m = 1,2,\ldots$ are addressed but remain fixed. The aim of the algorithm is to find that quantization cell for $\boldsymbol{x}[l]$ for which the metric

$$d^2(\boldsymbol{q}, \widehat{\boldsymbol{q}}) = \sum_{i=0}^{D-1} (q[D \cdot l + i] - \hat{q}[D \cdot l + i])^2. \tag{35}$$

is minimized. For finding a well suited starting point for the algorithm we propose to begin with a forward prediction for the actual D samples; i.e. to disable the chain ADC and DAC in Fig. 5 or equivalently, to feed $q[k]$, $k = l \cdot D$, $l \cdot D + 1, \ldots, l \cdot D + D - 1$ into the prediction filter (of Fig. 5) instead of $\widehat{q}[k]$. Subsequently, spherical log. quantization is applied to the resulting vector $\boldsymbol{x}$ generating an initial vector $\boldsymbol{y}$ from which a vector $\widehat{\boldsymbol{q}}$ follows by usual inversion of the prediction error filter (DPCM-receiver structure). Thus, metric calculation for a given quantized vector $\boldsymbol{y}_i$ is straight forward. Given a certain actual reconstruction vector $\boldsymbol{y}_i$, all $2D$ nearest neighbor reconstruction vectors $\boldsymbol{y}_{j[i]}$ in D dimensions are addressed and the corresponding metrics (eq. (35)) are compared. The vector with the smallest metric is selected for the next iteration, i.e. $\arg\min_j d^2(\boldsymbol{q}, \widehat{\boldsymbol{q}}_{j[i]})$ provides an update for i with $\widehat{\boldsymbol{q}}_i$ corresponding to the quantized vector $\boldsymbol{y}_i$ via inversion of the prediction error filter. If there is none, the algorithm stops and delivers the index of $\boldsymbol{y}_i$ to be transmitted.

Notice that the neighboring cells usually have quite different indices and it is not trivial to identify these cells. For example, regarding $D = 3$ the angles of the

azimuth (φ_1) at large elevations (φ_2 near $\pm\pi/2$) will be quantized more coarsely than at small elevations (φ_2 near 0), because the latter have a larger circle of latitude on the surface of the sphere. Therefore, the indices of the neighboring cells may be stored in a ROM for all cells for a fast implementation as long as M_φ is not too high. Because linear prediction is incorporated into the optimization process the resulting SNR sometimes is higher than predicted by adding the gains (in dB) due to spherical logarithmic quantization and prediction (DPCM). Especially for low rates (e.g. $R < 5$ bit/sample) remarkable additional SNR gains are observable, see section 5 and Fig. 6.

An analytic result for the SNR achievable by this algorithm is yet not known to the authors.

Simulations have shown that the average number of iterations is about 0.25 (per D samples) and that there is no significant SNR-loss in comparison to unlimited search even if the maximum number of iterations is restricted to three. Therefore, a realtime implementation seems to be feasible. Additionally, small values of D already offer a high gain, cf. Figs. 3 and 6.

Regarding the receiver side, there are no differences to usual DPCM. Notice that the overall delay of the transmission system is only D samples and thus this method is excellent for situations that have to cope with an extreme low signal delay.

4.2 Variants

Performing the search for the quantization cell which leads to the minimum distortion according to eq. (35) all algorithms for so-called lattice-decoding, resp. the search for a maximum-likelihood-codeword (channel decoding) are applicable in order to speed up the search in principle, see e.g. [11] and the references therein. Of course, the lattice decoding methods have to be transformed to polar coordinates for this purpose.

A variant without any iterative search of the quantization cell arises by non-uniform quantization of the angle coordinates instead of exploiting correlations within the actual D samples by means of DPCM whereas correlations of these D samples to previous samples may be exploited by DPCM as before. This leads to an update of the prediction filter every D steps.

We propose to maintain the logarithmic quantization of the radius and to normalize the signal vector to radius 1. This enables an analytical calculation of the probability density function of the signal points on the surface of the unit sphere, e.g. by the assumption of a Gaussian process and exploiting the autocorrelation function of the source signal or by means of a direct experimental determination of

the relative frequency of the signal points on the surface of the sphere, resp.

For that purpose it is possible to develop a non-uniform cubic quantization using the optimization strategy proposed in [12] as long as we assume $M \gg 1$. This approach enforces that the average contribution of every quantization cell to the quantization noise should be identical. With the Δ_i representing the width of a $(D-1)$-dimensional cubic quantization cell and $\boldsymbol{z}_i$ the corresponding center, as well als $f(\boldsymbol{v})$; $\boldsymbol{v} := (\varphi_1, \ldots, \varphi_{D-1})$ representing the joint probability density function or relative frequency of angles, resp., the following condition has to be satisfied:

$$(D-1)\frac{\Delta_i^2}{12} f(\boldsymbol{z}_i)\Delta_i^{D-1} = \text{const.} \tag{36}$$

with the constraint

$$\sum_{i=1}^{M_\varphi} \Delta_i^{D-1} = \beta_D. \tag{37}$$

This leads immediately to a $(D-1)$-dimensional compressor function $\boldsymbol{k}(\boldsymbol{v})$ which determines a non-uniform quantization of the surface of the sphere. The non-uniform quantization can be performed e.g. like in the one-dimensional case by a nonlinear mapping of the vector $\boldsymbol{v}$ into a vector $\boldsymbol{z} := \boldsymbol{k}(\boldsymbol{v})$ and a subsequent uniform quantization which leads to the vector $\boldsymbol{z}_i$ as we showed in section 3.2. The application of the inverse function yields the reconstruction vector $\widehat{\boldsymbol{v}} := \boldsymbol{k}^{-1}(\boldsymbol{z}_i)$.

The resulting $(D-1)$-dimensional compressor function can be either approximated by an analytic function or by $(D-2)$-dimensional partial plains. This resembles the well-known approximation by straight lines (e.g. the 13-segment-characteristic) as it is used for one-dimensional compression. By using an orthogonal raster for this approximation concerning the $D-1$ angles in the space $(-\pi/2, \pi/2]^{D-2} \times (-\pi, \pi]$, an easy implementation of the compressor function is achievable.

5 Simulation Results

The simulation of spherical locarithmic quantization in combination with DPCM according to chapter 4 was performed for the ouverture and the aria "Der Vogelfänger bin ich ja" of the opera "Zauberflöte" of Wolfgang Amadeus Mozart [13]. Fig. 6 shows simulation results for $R = 3$, $R = 4$ and $R = 7$ bit/sample at $\mathscr{A} = 102726$, $\mathscr{A} = 48270$, $\mathscr{A} = 4858$ and various numbers of dimensions D. First the signal was coded, then decoded and the SNR was calculated by comparison with the original CD-signal. Here the Signal-to-Noise Ratio is averaged on the whole pieces of music. At the low rates the SNR increases even steeper with D

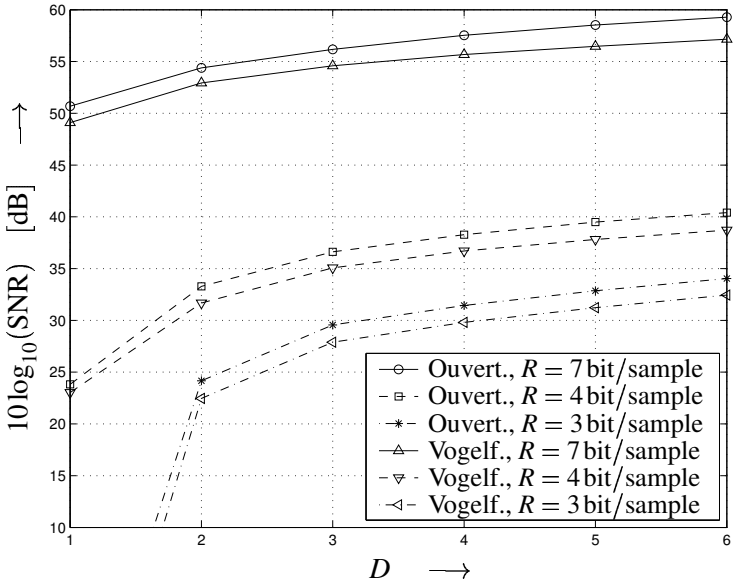

Fig. 6. SNR of spherical logarithmic quantizationin combination with DPCM.
——"Ouvertuere", (W. A. Mozart), [13] Track 1,
Time = 7:11 min, −27.20 dB, P=2.
- - -"Der Vogelfänger bin ich ja", (W. A. Mozart), [13] Track 3,
Time = 3:02 min, −32.15 dB, P=2.
$R = \{7, 4, 3\}$ bit/sample, $\mathscr{A} = \{4858, 48270, 102726\}$.

when compared to Fig. 3. Notice the different values for $\mathscr{A}$ while comparing the results. The aria "Vogelfänger" is stored at a mean signal level of −32.15 dB and offers a challenging example for audio coding due to signal dynamic and timbres (prelude, singing, reed pipe). Using an universal prediction filter of low prediction order ($P = 2$), these simulation results can be considered as representative for a multiplicity of audio signals.

From a comparison of Fig. 3 (or eq. (1)) and Fig. 6 at $D = 1$, $R = 7$ it is obvious that the average prediction gain is about 20 dB to 23 dB for this simple predictor (sampling frequency 44.1 kHz!). For $R = 7$ the overall gain is well approximated by the sum of both individual gains by spherical logarithmic quantization and prediction at all values of D, whereas for $R = 4$ this is only true for $D \geq 5$. (Of course, eq. (1) is not applicable for $R = 3$ or $R = 4$ and $D = 1$, $\mathscr{A} = 48270$.)

The mean signal level of the ouvertuere is −27.20 dB; it is characterized by a wide dynamic range from below -70 dB up to -17 dB, see also Fig. 7.

In Fig. 7 segmental signal level and segmental SNRs are plotted for the example ouvertuere; each segment consists of 6000 samples, i.e. 0.136 s. Spherical

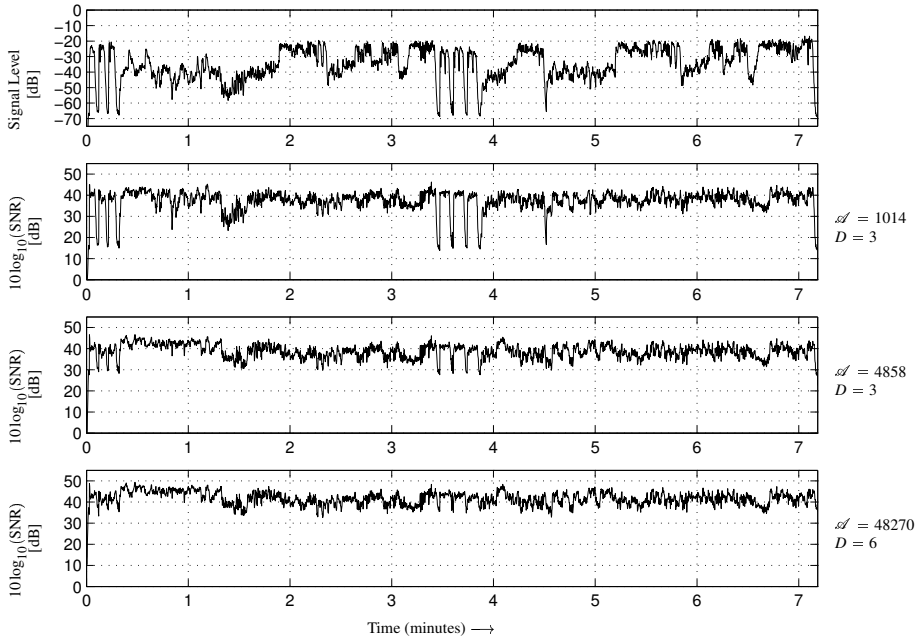

Fig. 7. Segmental signal level and segmental SNR of spherical logarithmic quantization in combination with DPCM "Ouvertuere" (W. A. Mozart), [13] Track 1, Time = 7:11 min, −27.20 dB, P=2, R=4 bit/sample Segments of 6000 samples or 0.136s, resp.

log. quantization in $D = 3$ dimensions with $\mathscr{A} = 1014$ and $\mathscr{A} = 4858$ as well as in $D = 6$ dimensions with $\mathscr{A} = 48270$ and the same compromise predictor with degree $P = 2$ as in the first example are applied. The upper curve shows the high dynamics which are caused by well known long heavy tones separated by general pauses over entire bars at the beginning and in the middle of this popular piece of music. The curves for $\mathscr{A} = 1014$ and $\mathscr{A} = 4858$ show the benefit of an increasing $\mathscr{A}$ regarding the segmental SNR especially for the general pauses. The lower curve in Fig. 7 demonstrates this aspect for a further increase of $\mathscr{A}$ in combination with another 3 dB gain in SNR due to the higher dimensionality of $D = 6$ (see also Fig. 3). Please notice, that despite of the low rate $R = 4$ bit/sample a $10\log_{10}(\text{SNR})$ >35 dB (w.r.t. the original CD) is even maintained during this general pauses.

The value $\mathscr{A} = 48270$ corresponds to a compression of 12 bits, e.g. for $D = 1$, $R = 12$ a resolution corresponding to 24 bits is achieved for very small samples. Of course, for $D = 1$, $R = 4$ a resolution according to 16 bit is never present because of the low number of intervals. But applying spherical logarithmic quantization in 6 dimensions the extreme high value $\mathscr{A} = 48270$ indeed is well designed for

maximum minima of the segmental SNR. Here the resolution is even as high as that of the original CD-data for the segments with a signal level of -70 dB and lower. No person – even if familiar with audio coding artifacts – have been found yet, who were able to hear any difference between the quantized signal at $R = 4$, $D = 6$, $\mathscr{A} = 48270$ and the original one with high reliability. The interested reader is invited to evaluate these audio examples [14] by himself.

6 Conclusion

We presented a waveform-conserving digitizing scheme for analog source signals, which on the one hand side combines gains from multidimensional and logarithmic quantization at a tolerable complexity for implementation and on the other hand side is capable of further SNR gains by means of linear prediction. Finally we want to emphasize that besides an advantageous tradeoff between rate and distortion this method is particularly characterized by its extreme high dynamic range and a very low structural signal delay of only a few sample periods.

References

[1] J. B. Huber and B. Matschkal, "Spherical logarithmic quantization and its application for DPCM," in *Proc. 5th International ITG Conference on Source and Channel Coding (SCC)*, Erlangen, Germany, Jan. 2004, pp. 349–356.

[2] K. Tröndle and R. Weiß, *Einführung in die Pulscodemodulation*. Munich: Oldenbourg Verlag, 1974.

[3] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Information Theory*, pp. 2325–2383, Oct. 1998.

[4] M. Herbert, "Lattice-Quantisierung von Sprach- und Sprachmodellsignalen," in *Ausgewählte Arbeiten über Nachrichtensysteme*, H.-W. Schüßler, Ed., Erlangen, 1991, no. 79.

[5] J. Volder, "The CORDIC trigonometric computing technique," *IRE Trans. Electronic Computing*, vol. EC-8, pp. 330–334, Sept. 1959.

[6] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*, 3rd ed. Springer-Verlag.

[7] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: John Wiley & Sons, Inc., 2002, iSBN 0471 22410 3.

[8] T. Berger and J. Gibson, "Lossy source coding," *IEEE Trans. Information Theory*, pp. 2693–2723, Oct. 1998.

[9] L. Pakula and S. Kay, "Simple proofs of the minimum phase property of the prediction error filter," *IEEE Trans. on ASSP*, vol. 31, 1983.

[10] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[11] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. on Information Theory*, pp. 2201–2214, Aug. 2002.

[12] P. F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, Jan. 1951.

[13] "Mozart: "Die Zauberflöte"," Philips Classics Productions (DDD), Polygram Records #442569-2, 1994, tracks 1 and 3.

[14] Audio examples. [Online]. Available: http://www.LNT.de/LIT/audio-examples