ENTROPY-CONSTRAINED QUANTIZATION OF EXPONENTIALLY DAMPED SINUSOIDS PARAMETERS

Olivier Derrien

Université de Toulon / CNRS LMA,
Marseille, France
derrien@lma.cnrs-mrs.fr

Roland Badeau, Gaël Richard

Télécom ParisTech / CNRS LTCI,
Paris, France
firstname.lastname@telecom-paristech.fr

ABSTRACT

Sinusoidal modeling is traditionally one of the most popular techniques for low bitrate audio coding. Usually, the sinusoidal parameters are kept constant within a time segment but the exponentially damped sinusoidal (EDS) model is also an efficient alternative. However, the inclusion of an additional damping parameter calls for a specific quantization scheme. In this paper, we propose an asymptotically optimal entropy-constrained quantization method for amplitude, phase and damping parameters. We show that this scheme is nearly optimal in terms of rate-distortion trade-off. We also show that damping consumes the smallest part of the total entropy of quantization indexes, which suggests that the EDS model is truly efficient for audio coding.

Index Terms— Parametric audio coding, Exponentially damped sinusoids, Quantization, Entropy.

1. INTRODUCTION

For low bitrate audio coding applications, parametric coders are an efficient alternative to transform coders. Many parametric models were proposed, but the sinusoidal model remains the most popular, because most real-world audio signals are dominated by tonal components. Traditionally, in sinusoidal models used for parametric coding, the amplitude of each component is kept constant within a time segment. Both parametric codecs included in the MPEG-4 Audio standard, HILN and SSC, use a sinusoidal model combined with a noise model (and an additional transient model in SSC). However, some studies have shown that an exponentially damped sinusoidal (EDS) model is an efficient alternative for audio modeling [1, 2]. In HILN and SSC, sinusoidal parameters are quantized independently: frequency is quantized at Just Noticeable Distortion, amplitude uses a log-uniform scalar quantizer, and phase a uniform scalar quantizer. Recently, more efficient joint-quantizers for amplitude-phase [3] and amplitude-phase-frequency [4] have been proposed, which take advantage of the statistical dependence between the parameters. With the EDS model, the inclusion of the additional damping parameter calls for a new quantization scheme. The purpose of this paper is to propose a solution for joint quantization of amplitude, damping and phase, the bitrate constraint being formulated in terms of entropy of quantization indexes. For the moment, we do not consider the joint quantization of frequency, neither the repartition of the bitrate between several sinusoids. This method can be seen as an extension of the work by Vafin et al. [3] for constant-amplitude sinusoids. First, we present the EDS model. Then, we describe our quantization scheme. Finally, we evaluate the performance of our method on both synthetic and real data. We compare our scheme with a trained vector-quantizer and with a polar quantizer associated with an independent damping quantizer. We also consider the distribution of entropy between parameters.

2. EDS SIGNAL MODELING

The modeling of a signal $x(t), t \in [0, T]$ can be written as

$$x(t) = \sum_{k=0}^{K-1} s_k(t) + \varepsilon(t) \qquad (1)$$

where K is the model order, T is the length of the analysis window and $\varepsilon(t)$ is a white noise. s_k is an exponentially damped sinusoid (EDS) defined as

$$\begin{cases} s_k(t) & = & a_k \, e^{\delta_k(t/T-1)} \, e^{i(\omega_k t + \phi_k)}, & \text{if } \delta_k \geq 0, \\ s_k(t) & = & a_k \, e^{\delta_k t/T} \, e^{i(\omega_k t + \phi_k)}, & \text{if } \delta_k < 0. \end{cases} \qquad (2)$$

Each EDS is characterized by a set of 4 parameters: amplitude a_k, damping δ_k, pulsation ω_k and phase ϕ_k. Note that damping can be positive (increasing envelope) or negative (decreasing envelope). Using different expressions for positive and negative dampings avoids numerical errors while estimating amplitudes for high dampings.

Actually, the most popular schemes for EDS parameter estimation are subspace methods. During the last years, many studies have been published about the optimization of these methods. In this paper, we selected the estimation scheme proposed in [5], which was developed for audio signals.

Considering the quantization of a single sinusoid, we omit index k. The mean square error (MSE) distortion measure between an EDS s and the reconstructed EDS $\hat{s}$ can be written using the continuous-time signal model:

$$d = \frac{1}{T} \int_0^T |s(t) - \hat{s}(t)|^2 \, dt. \qquad (3)$$

Assuming that pulsation is not quantized, we get:

$$d = a^2 h(2\delta) + \hat{a}^2 h(2\hat{\delta}) - 2a\hat{a}\cos(\phi - \hat{\phi})h(\delta + \hat{\delta}) \qquad (4)$$

where $\hat{a}$, $\hat{\phi}$ and $\hat{\delta}$ are respectively the reconstructed amplitude, phase and damping. h is the real-valued function defined as

$$h(x) = \frac{1 - e^{-|x|}}{|x|} \quad \forall x \in \mathbb{R}\backslash\{0\}, \quad h(0) = 1. \qquad (5)$$

We assumed that $\hat{\delta}$ and δ have the same sign. A sufficient condition is that the damping quantizer is symmetric around zero. In practice, $x(t)$ is a discrete-time signal, but (4) is still a good approximation as long as $1/T$ is small compared to the sampling frequency.

3. OPTIMAL QUANTIZER IN HIGH RESOLUTION

In this section, we propose an asymptotically optimal method for high-resolution quantization, (i.e. assuming a large number of quantization cells). Basically, we follow the same method as in [3] and [4], which is similar to the one introduced by A. Gersho [6]. The quantizers are defined by their quantization cell density (QCD), which can be seen as the inverse of the quantization step-size. In order to derive an expression for the optimal QCD, we make a simplifying assumption: amplitude, damping and phase are quantized with scalar quantizers, but depending on one another. This is called *joint-scalar quantization*. The QCD can be split in 3 scalar functions: g_A, g_Δ and g_Φ respectively for amplitude, damping and phase.

We define $p = \{a, \delta, \phi\}$ as the set of EDS parameters. We denote $\hat{p} = \{\hat{a}, \hat{\phi}, \hat{\delta}\}$ the set of reconstructed parameters and $i = \{i_a, i_\phi, i_\delta\}$ the set of quantization indexes associated with $\hat{p}$. We note P, $\hat{P}$ and I the random variables associated respectively with p, $\hat{p}$ and i. The optimal quantizer minimizes the mean distortion $D = \mathbb{E}[d(P, \hat{P})]$ under the constraint $H(I) \leq R$, where $H(I)$ denotes the entropy of quantization indexes and R the target entropy.

3.1. Mean distortion

We consider the quantization cell $\mathcal{C}$ associated to $\hat{p}$. In high resolution, quantization cells are centered on reconstruction values because the probability mass function of parameters is approximately constant over each cell. We note Δ_a, Δ_ϕ and

Δ_δ the length of quantization intervals. The mean distortion over cell $\mathcal{C}$ is defined as:

$$d_\mathcal{C} = \int_{\hat{a}-\Delta_a/2}^{\hat{a}+\Delta_a/2} \int_{\hat{\phi}-\Delta_\phi/2}^{\hat{\phi}+\Delta_\phi/2} \int_{\hat{\delta}-\Delta_\delta/2}^{\hat{\delta}-\Delta_\delta/2} d(p,\hat{p}) \, da \, d\phi \, d\delta \quad (6)$$

where $d(p,\hat{p})$ is given by (4). Assuming that Δ_a, Δ_ϕ and Δ_δ are small, expression (6) can be developed in Taylor series. Keeping only the most significant terms, we get:

$$d_\mathcal{C} \approx \frac{1}{12}\left[h(2\hat{\delta})\Delta_a^2 + \hat{a}^2 h''(2\hat{\delta})\Delta_\delta^2 + \hat{a}^2 h(2\hat{\delta})\Delta_\phi^2\right] \quad (7)$$

where $h''(x)$ denotes the second order derivative of $h(x)$. The length of quantization intervals are related to the QCDs:

$$\Delta_a = g_A(\hat{p})^{-1} \quad \Delta_\phi = g_\Phi(\hat{p})^{-1} \quad \Delta_\delta = g_\Delta(\hat{p})^{-1} \quad (8)$$

Thus, equation (7) can be re-written as:

$$d_\mathcal{C} \approx \frac{1}{12}\left[\frac{h(2\hat{\delta})}{g_A^2(\hat{p})} + \frac{\hat{a}^2 h''(2\hat{\delta})}{g_\Phi^2(\hat{p})} + \frac{\hat{a}^2 h(2\hat{\delta})}{g_\Delta^2(\hat{p})}\right]. \quad (9)$$

The mean distortion over all quantization cells is:

$$D = \sum_n \rho_n \, d_{\mathcal{C}_n} \qquad (10)$$

where $\rho_n = \text{proba}\{P \in \mathcal{C}_n\}$. Assuming that the probability mass function $\rho_P(p)$ is constant in each cell, we get $\rho_n \approx \rho_P(\hat{p}_n) V_n$, where V_n is the volume of $\mathcal{C}_n$. Combining previous equations leads to:

$$D \approx \frac{1}{12}\sum_n \rho_P(\hat{p}_n)\left[\frac{h(2\hat{\delta}_n)}{g_A^2(\hat{p}_n)} + \frac{\hat{a}^2 h''(2\hat{\delta}_n)}{g_\Delta^2(\hat{p}_n)} + \frac{\hat{a}^2 h(2\hat{\delta}_n)}{g_\Phi^2(\hat{p}_n)}\right] V_n. \quad (11)$$

The sum can be approximated by an integral:

$$D \approx \frac{1}{12}\int \rho_P(p)\left[\frac{h(2\delta)}{g_A^2(p)} + \frac{a^2 h''(2\delta)}{g_\Delta^2(p)} + \frac{a^2 h(2\delta)}{g_\Phi^2(p)}\right] dp. \quad (12)$$

3.2. Entropy constrained quantizers

The joint entropy of quantization indexes can be approximated by [6]:

$$H(I) \approx \mathcal{H}(P) + \int \rho_P(p) \log_2\left[g_A(p)g_\Delta(p)g_\Phi(p)\right] dp \quad (13)$$

where $\mathcal{H}(P)$ is the joint differential entropy of EDS parameters defined as

$$\mathcal{H}(P) = -\int \rho_P(p) \, \log_2(\rho_P(p)) \, dp. \qquad (14)$$

A Lagrange optimization technique finally leads to the QCDs which minimize D under the constraint $H(I) \leq R$:

$$\begin{cases} g_A(\delta) & \approx & h(2\delta)^{\frac{1}{2}} 2^{\frac{1}{3}(R-\sigma)} \\ g_\Delta(a,\delta) & \approx & a \, h''(2\delta)^{\frac{1}{2}} 2^{\frac{1}{3}(R-\sigma)} \\ g_\Phi(a,\delta) & \approx & a \, h(2\delta)^{\frac{1}{2}} 2^{\frac{1}{3}(R-\sigma)} \end{cases} \quad (15)$$

The constant σ is defined as

$$\sigma = \mathcal{H}(P) + \int \rho_\Delta(\delta) \log_2(h(2\delta)h"(2\delta)^{\frac{1}{2}})d\delta$$

$$+2 \int \rho_A(a) \log_2(a)da \qquad (16)$$

where $\rho_A(a)$ and $\rho_\Delta(\delta)$ are respectively the probability mass functions of amplitude and damping. One can observe that the amplitude and phase quantizers are uniform with respect to the quantized variable, but parametrized respectively by δ and (a,δ). The damping quantizer is parametrized by a, but not uniform: the QCD is higher when δ is small. This can be explained by the fact that an EDS with a high damping affects only a small part of the analysis segment, and thus does not contribute much to the MSE and can be quantized more roughly. Note that for $\delta = 0$, our solution reduces to the polar quantizer described in [3].

Combining equations (12) and (15), we get the entropy-distortion function:

$$D \approx \frac{1}{4} 2^{\frac{2}{3}(\sigma - R)}. \qquad (17)$$

4. PERFORMANCE EVALUATION

4.1. Implementation details

Quantizers defined by equations (15) can be implemented with compression/expansion functions and a scalar uniform quantizer, the QCD being the slope of the compression function [6]. For amplitude and phase, the compression functions are linear. For damping, computing the compression function is not straightforward. So we pre-computed numerically a sampled version of the compression function, and interpolated between the samples. For each scalar quantizer, we choose zero as the central reconstruction value, and for each value of the amplitude, the step-size of the phase quantizer is slightly modified in order to cover $[0, 2\pi]$ with an integer number of quantization cells.

4.2. Entropy-distorsion function

First, we evaluate the performance of our quantization scheme on synthetic data. Like in [3] and [4], we assume that amplitude, phase and damping are statistically independent. In the literature, the amplitude is usually Rayleigh distributed and the phase is uniformly distributed over $[0, 2\pi]$. With the EDS model, we found out that the amplitude is more likely Gamma distributed ($p = 1$ and $\theta = 0.21$). For damping, because of equation (5), only the distribution of $|\delta|$ is significant. Experiments showed that $\log(|\delta|)$ follows approximately a centered Gaussian distribution (of variance 1.2). Using equation (16), we computed $\sigma = -5.66$ bits.

We evaluated the entropy-distorsion curve on $N = 10^8$ sets of parameters where amplitude, phase and damping are

independently generated using the distributions described above. Results are plotted on figure 1. Compared to the theoretical relation given by equation (17), one can observe that the practical curve diverges in low resolution but converges in high resolution. We also compared our method with an entropy-constrained vector quantizer (VQ) as described by Chou et al. [7]. The VQ is slightly better at medium resolution, but has similar performance in very low and high resolution. However, in terms of complexity, the joint scalar quantizer clearly outperforms the VQ.
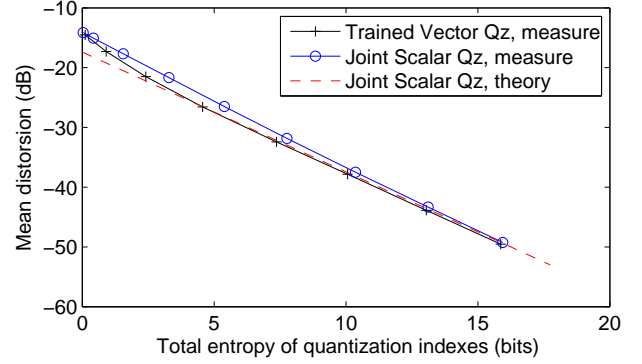

Fig. 1. Joint-scalar quantizer (vs) trained VQ.

4.3. Comparison with a simpler quantization scheme

A simpler alternative is to apply a polar quantizer (as described in [3]) to amplitude and phase, and an independent entropy-constrained scalar quantizer to damping. This method requires an a-priori distribution of the entropy between the two quantizers. We tested 5 values of the damping/polar entropy ratios r and plotted the entropy-distortion curve on figure 2. One can see that the joint scalar quantizer is always better. Furthermore, with the two-quantizers solution, the optimal entropy ratio depends on the target entropy, while our quantizer automatically adjusts the entropy balance.
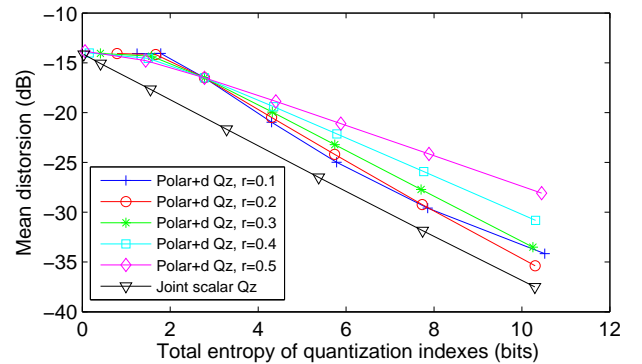

Fig. 2. Joint-scalar quantizer (vs) polar + damping quantizers.

4.4. Distribution of entropy between the parameters

We also considered the distribution of entropy between the quantization indexes associated to the three parameters. For different values of the target entropy, we computed the ratio $H(I_x|I_y, I_z)/H(I_x, I_y, I_z)$, x being amplitude, phase or damping, and $\{y, z\}$ the other two parameters. The results are plotted on figure 3. One can notice that phase always requires the greatest part of the entropy (which is consistent with the results reported in [4]), and the damping always requires the lowest part, especially in low resolution. Asymptotically, all three parameters seem to contribute equally.
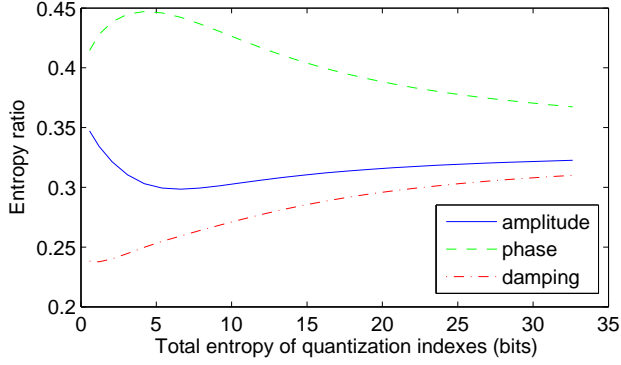
Fig. 4. Entropy-Distortion function on real audio data.

scheme for amplitude, phase and damping. We showed that our method performs better than a more simple scheme, and almost as well as a trained vector quantizer, which is theoretically the best solution, although practically not suitable in the context of audio coding. We also showed that the additional damping parameter requires fewer coding bits compared to amplitude and phase. This suggests that the EDS model is an efficient alternative to constant amplitude sinusoidal models for parametric audio coding. However, this is still a work in progress: we will extend our study to include the quantization of frequency, and consider the repartition of the coding bits between several sinusoids using a hearing model.

Fig. 3. Distribution of entropy between parameters.

4.5. Performance on real audio data

We applied our quantizer to a corpus of EDS parameters obtained by analyzing real audio signals: 8 audio excerpts from various musical styles, sampled at 44.1kHz. We use the same analysis method as in [5]. In the preprocessing stage, the signal is segmented in variable-length time-segments aligned with onset positions in order to get physically consistent EDS parameters. But this is not directly related to the design of the quantization stage. We set a constant model order in each time-segment ($K = 50$), so that approximately 95% of the energy is captured. The analysis gives a corpus of 64000 sets of parameters. We also use a second database, similar to the first one, to train a VQ. As one can see on figure 4, the measured entropy-distortion function for the joint-scalar quantizer is asymptotically similar to the one obtained with synthetic data. For the VQ, as explained in [7], the minimum achievable distortion is limited by the small size of the training database, which penalizes the VQ in high resolution.

5. CONCLUSION

We considered the quantization of exponentially damped sinusoids parameters. Given a constraint on the entropy of quantization indexes, i.e. the number of coding bits per sinusoid, we proposed a new asymptotically optimal quantization
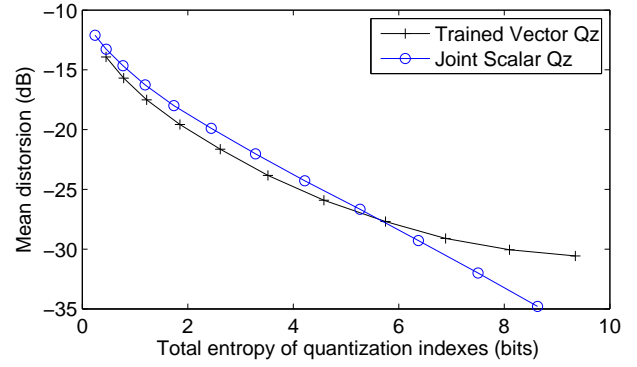
6. REFERENCES

[1] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere, "Robust exponential modeling of audio signals," in *Proc. ICASSP*, Seattle, WA, USA, May 1998.

[2] J. Jensen and R. Heusdens, "A comparison of sinusoidal model variants for speech and audio representation," in *Proc. EUSIPCO*, Toulouse, France, 2002.

[3] R. Vafin and W.B. Kleijn, "Entropy-constrained polar quantization and its application to audio coding," *IEEE Tr. Speech, Audio Process.*, vol. 13, no. 2, Mar. 2005.

[4] P. Korten, J. Jensen, and R. Heusdens, "High-resolution spherical quantization of sinusoidal parameters," *IEEE Tr. Audio, Speech, Lang. Process.*, vol. 15, no. 3, 2007.

[5] R. Badeau, R. Boyer, and B. David, "EDS parametric modeling and tracking of audio signals," in *Proc. DAFx*, Hamburg, Germany, Sept. 2002.

[6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Tr. Inf. Theory*, vol. 25, no. 4, July 1979.

[7] P.A. Chou, T. Lookbaugh, and R.M. Gray, "Entropy-constrained vector quantization," *IEEE Tr. Acoust., Speech, Signal Process.*, vol. 37, no. 1, Jan. 1989.