

# PERFORMANCE BOUNDS FOR LPC SPECTRUM QUANTIZATION

*Per Hedelin, Jan Skoglund, and Jonas Samuelsson*

Information Theory  
Department of Signals and Systems  
Chalmers University of Technology  
SE-412 96 Göteborg, Sweden

## ABSTRACT

This paper presents a method for obtaining numerical estimates of high rate vector quantization (VQ) performance suitable for sources for which the pdf is not analytically available. In the proposed method, the VQ point density is described from a Gaussian mixture model optimized for the data. Employing this method for LPC spectrum quantization, we obtain high rate expressions for both the average spectral distortion (SD) and the distribution function of the SD. We estimate the minimum bits required for a quantizer to obtain an average SD of 1 dB and the outlier statistics for that quantizer. We find that approximately 3 bits can be saved as compared to a 2-split LSF-based vector quantizer.

## 1. INTRODUCTION

The linear prediction coefficients (LPC) are often used to represent the short-time spectral envelope of the speech signal in source-filter based speech coders. The LPC parameters generally consumes a considerable fraction of the total bit rate for many low rate coders. Much effort has therefore been spent on efficient quantization of LPC parameters. The spectral distortion (SD) has become the standard measure for evaluating spectrum quantization. Performance is usually presented as the average SD and the percentage of outlier spectra having SD greater than predefined thresholds [1]. For sufficient quality, a spectrum quantizer must provide an average SD around 1 dB. For unconstrained VQ, technology limits codebook sizes to a maximum of, say 13-15 bits, for current hardware. This is not sufficient to achieve the 1 dB requirement. However, the minimum size for a quantizer achieving a 1 dB performance is not thoroughly established. Informal rough bounds around 20 bits have been given in e.g. [2, 3]. We will in this paper explore the performance bounds with the assistance of high-rate VQ analysis.

In VQ literature much theoretical work has been presented for high rate quantization, see e.g [4, 5]. The theory presents expressions for minimum distortion of quantizers operating at high rates. The formulas contain integrals over the probability density function (pdf) of the data, and closed form expressions can only be derived for simple pdfs. For such analytically tractable pdfs, there is a high accuracy in the performance predicted by high rate theory and the performance obtained for optimally designed, e.g. using the generalized Lloyd algorithm [6], high rate quantizers. An example of this for a Gaussian source can be found in [7]. For a general and unknown data pdf, the formulas can be calculated numerically using histogram pdf estimates [5]. However, histograms are only feasible for low dimensions and when the available data

sets are large. Thus, histograms have only been applied to split VQ structures. The high rate performance bound for a full dimensional LPC VQ has not yet been rigorously examined.

In this paper we present a new method to obtain high rate expressions that are numerically tractable for moderate dimensions. In practice, this means that we can handle dimensions of interest for spectrum quantization. We model the underlying pdf of the vectors in a database as a mixture of Gaussians and optimize the parameters of the model. The number of model parameters is sufficiently low for obtaining accurate high rate predictions of the performance of 10-dimensional unconstrained LPC quantization.

## 2. HIGH RATE QUANTIZATION

Consider a  $d$ -dimensional vector quantizer with  $N$  partitions  $\Omega_i$  with corresponding reconstruction vectors  $\mathbf{c}_i$ . Let the performance of the quantizer be expressed by the average distortion

$$D = E[\|\mathbf{x} - \tilde{\mathbf{x}}\|^r] \quad (1)$$

where  $\|\cdot\|$  denotes the  $L_2$  norm. According to asymptotic quantization theory [4, 5], the high rate distortion approximation can be expressed as

$$\begin{aligned} D &\approx N^{-r/d} \frac{d}{d+r} V_d^{-r/d} \int_{\mathcal{R}^d} f_{\mathbf{x}}(\mathbf{x}) \lambda(\mathbf{x})^{-r/d} d\mathbf{x} \\ &= \frac{d}{d+r} V_d^{-r/d} N^{-r/d} D_H \end{aligned} \quad (2)$$

where  $D_H$  is the integral,  $V_d$  is the volume of a  $d$ -dimensional sphere with unit radius and  $\lambda(\mathbf{x})$  is a continuous VQ point density function that integrates to one. A point density which minimizes  $D_H$  is

$$\lambda_{opt}(\mathbf{x}) = \frac{f_{\mathbf{x}}^{\frac{d}{d+r}}(\mathbf{x})}{\int_{\mathcal{R}^d} f_{\mathbf{x}}^{\frac{d}{d+r}}(\mathbf{x}) d\mathbf{x}} \quad (3)$$

Now consider the case where we estimate a model density  $f_{\mathbf{M}}(\mathbf{x})$  for the data and then design a high rate quantizer with optimal point density  $\lambda(\mathbf{x})$  for that model. If we employ this quantizer in the quantization of data having an unknown pdf  $f_{\mathbf{x}}(\mathbf{x})$ ,  $D_H$  can

be expressed

$$\begin{aligned}
D_H &= \int_{\mathcal{R}^d} f_{\mathbf{x}}(\mathbf{x}) \lambda(\mathbf{x})^{-r/d} d\mathbf{x} \\
&= \left( \int_{\mathcal{R}^d} f_{\mathbf{M}}^{\frac{d}{d+r}}(\mathbf{x}) d\mathbf{x} \right)^{r/d} \int_{\mathcal{R}^d} f_{\mathbf{M}}^{\frac{-r}{d+r}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\
&= \left( E_M [f_{\mathbf{M}}^{\frac{d}{d+r}}(\mathbf{x})] \right)^{r/d} E_X [f_{\mathbf{M}}^{\frac{-r}{d+r}}(\mathbf{x})] \\
&\approx \left( \frac{1}{N_y} \sum_{n=1}^{N_y} f_{\mathbf{M}}^{\frac{d}{d+r}}(\mathbf{y}_n) \right)^{r/d} \frac{1}{N_x} \sum_{n=1}^{N_x} f_{\mathbf{M}}^{\frac{-r}{d+r}}(\mathbf{x}_n) \quad (4)
\end{aligned}$$

where  $\{\mathbf{x}_n\}_{n=1}^{N_x}$  are vectors from the database we are modeling and  $\{\mathbf{y}_n\}_{n=1}^{N_y}$  are ‘‘synthetic’’ data generated from the model.

Thus, we have now obtained an expression which predicts the high rate performance of a vector quantizer designed from data having the probability density function  $f_{\mathbf{M}}(\mathbf{x})$  and evaluated on data having probability density function  $f_{\mathbf{x}}(\mathbf{x})$ .

Assuming perfect modeling, (2) is formally a lower bound valid in the domain of fine quantization. However, our experimental experience for standard pdfs is that results close to the bound are achievable with VQ training provided that the number of entries in the VQ codebook is sufficiently large. In other words, the rate of the VQ must be sufficiently large. For small codebooks, and in particular for low dimensions, the assumptions of the derivation are violated and, consequently, the bound will give misleading guidance for performance.

### 3. GAUSSIAN MIXTURE MODELS

One class of pdfs, which has been widely used for density estimations in a variety of applications, e.g [8] is the family of finite mixture densities, where the density function is a weighted sum of component densities. The case where the component densities are Gaussian is the most common and any continuous probability density function can be approximated arbitrarily closely by such a Gaussian Mixture (GM) density.

However, for some sources, such as LPC-parameters, the support of the source pdf is bounded and this boundary can be quite complex. To describe this boundary with a Gaussian mixture we would need a prohibitive large number of mixture components. Therefore we let our model density,  $f_{\mathbf{M}}(\mathbf{x})$ , consist of an unbounded GM density,  $f_{\mathbf{GM}}(\mathbf{x})$ , multiplied with a bounding function  $f_c(\mathbf{x})$ ,

$$f_{\mathbf{M}}(\mathbf{x}) = \frac{f_{\mathbf{GM}}(\mathbf{x}) f_c(\mathbf{x})}{\int_{\mathcal{R}^d} f_{\mathbf{GM}}(\mathbf{x}) f_c(\mathbf{x}) d\mathbf{x}} \quad (5)$$

where the denominator assures that the model pdf integrates to one. The function  $f_c$  is one inside the bounded support and zero outside. We can then express (4) as

$$D_H = \left( E_{GM} [f_{\mathbf{GM}}^{\frac{d}{d+r}}(\mathbf{x}) f_c(\mathbf{x})] \right)^{r/d} E_X [f_{\mathbf{GM}}^{\frac{-r}{d+r}}(\mathbf{x})]. \quad (6)$$

The gain of using a model pdf with a bounded support is illustrated in Figure 1.

We furthermore constrain the covariance matrices of the GM components to be diagonal, a constraint which increases the number of mixture components required for a given model accuracy but still reduces the total number of model parameters [8]. By rotating the coordinate system such that the data vectors are mapped

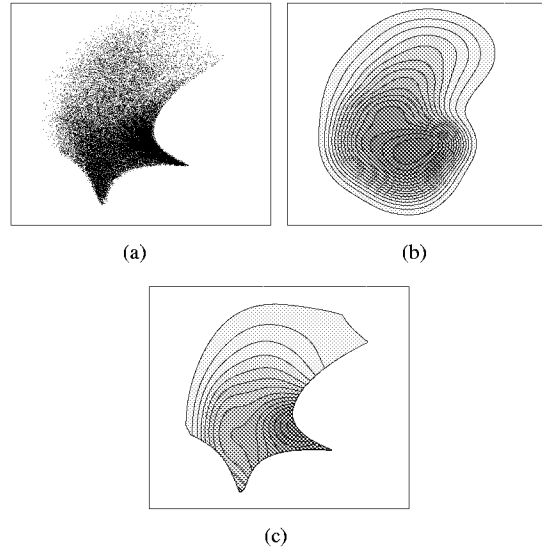
onto the eigenvectors of the covariance matrix of the data, a low loss in modeling capabilities occurs due to constraining to diagonal component covariance matrices

### 4. SEARCH FOR OPTIMAL VQ POINT DENSITY

The expectation-maximization (EM) algorithm [9] is widely used in the case of an unbounded GM model pdf. For the bounded model pdf an estimation scheme where the same criterion, i.e. the likelihood function, is maximized can easily be derived. In this case the log likelihood function normalized with the number of data points  $N$  can be seen as an approximation of an integral

$$\frac{1}{N} L(\Theta) = \frac{1}{N} \sum_{n=1}^N \ln f_{\mathbf{M}}(\mathbf{x}) \approx \int_{\mathcal{R}^d} \ln f_{\mathbf{M}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (7)$$

We can formulate an EM-like algorithm, referred to as EMbs, by first differentiating, setting the derivatives to zero and then formulate recursive update equations based on previous parameter settings.



**Figure 1:** (a): Cross-section of a three-dimensional scatter plot of cepstral parameters from a 3:rd order LPC analysis. (b): The corresponding cross-section for an 8-component EM-estimated unbounded Gaussian mixture density. (c): 8-component EMbs-estimated bounded Gaussian mixture density.

The EMbs algorithm optimizes the model according to the ML criterion but this is not what is really desired. For coding purposes we want to minimize  $D_H$ . One straight-forward approach to minimize the distortion is to utilize a gradient-based procedure. Then the model parameter vector,  $\Theta$ , is recursively given by

$$\Theta^{(k+1)} = \Theta^{(k)} - \eta \mathbf{T} \frac{\partial D_H}{\partial \Theta} \Big|_{\Theta=\Theta^{(k)}} \quad (8)$$

where  $\eta$  is a positive real-valued constant, i.e. the step-size determining the amount of movement, and  $\mathbf{T}$  is a matrix. The partial derivatives are straight-forward to express and details can be found in [10]. In our implementation  $\mathbf{T}$  is the unity matrix. We refer to

this algorithm as HRO. A few iterations of the EMbs algorithm gives the initial parameter set for HRO.

Hence, we have used two algorithms, EMbs and HRO, for optimizing the VQ point density. Both require a substantial amount of iterations to converge. In the following experiments, we essentially rely on HRO.

## 5. SPECTRUM CODING

Spectral distortion (SD) has become the standard measure for evaluating the performance of spectrum coding. SD is defined as an integral over a frequency region

$$SD_{\nu_0}^2 = \frac{20^2}{\nu_0} \int_0^{\nu_0} \left( \log_{10} \frac{|H(e^{j2\pi\nu})|}{|\tilde{H}(e^{j2\pi\nu})|} \right)^2 d\nu \quad (9)$$

using a normalized frequency scale  $|\nu| < 1/2$ , and where  $H(z)$  and  $\tilde{H}(z)$  are the original and the quantized synthesis filters, respectively, for the current frame. The spectral distortion is often calculated in a limited region of the spectrum  $\nu_0 < 1/2$ , therefore we have incorporated the argument in (9). The most common choice of region in the literature is  $\nu_0 = 3/8$ , corresponding to the 0–3 kHz band for a 8 kHz sampling frequency.

When evaluating over a database the performance is mostly presented as the average SD and the percentage of outlier spectra having SD greater than 2 % [1]. The average SD is most often calculated as

$$\overline{SD}_{\text{MRS}} = \frac{1}{N_D} \sum_{n=1}^{N_D} \sqrt{SD_n^2} \quad (10)$$

where  $N_D$  is the number of vectors in the database.

For the experiments we compiled a database of LP spectrum vectors from speech sampled at 8 kHz and low-pass-filtered at 3.4 kHz. A 10th order LP analysis using the stabilized covariance method with high-frequency compensation and error weighting (following [1]) was performed every 20 ms using a 25-ms analysis window. A fixed 10-Hz bandwidth expansion was applied to each pole of the LP coefficient vector. The database, which includes a large number of different speakers of both genders and a variety of languages, consists of 820 000 vectors. The evaluation set consists of an additional 48 000 vectors.

### 5.1. Cepstral Representation

In order to use the ideas presented in the previous sections to calculate a prediction of achievable performance in spectrum coding we have to choose an LPC parameter representation. The cepstral representation suits our purposes. For cepstral coefficients we have, as pointed out in [3, 11], by the Parseval relation, that for full-band spectral distortion ( $\nu_0 = 1/2$  in (9))

$$SD^2 = 2 \cdot 10^2 \cdot (\log_{10} e)^2 \sum_{n=1}^{\infty} (c_n - \tilde{c}_n)^2 \quad (11)$$

where  $c_n$  and  $\tilde{c}_n$  are the cepstral coefficients. Note that in order to minimize  $\overline{SD}_{\text{MRS}}$  we need to use  $r = 1$  in our estimation algorithms, c.f. (1). For a band limited distortion it can be shown that

$$SD_{\nu_0}^2 = 2 \cdot 10^2 \cdot (\log_{10} e)^2 (\mathbf{c} - \tilde{\mathbf{c}})^T \mathbf{B}_{\nu_0} (\mathbf{c} - \tilde{\mathbf{c}}) \quad (12)$$

where  $\mathbf{c}$  and  $\tilde{\mathbf{c}}$  are infinite-dimensional vectors containing the cepstral coefficients  $c_n$  and  $\tilde{c}_n$ , respectively, and where  $\mathbf{B}_{\nu_0}$  is a band limiting matrix having components

$$b_{i,j} = [\mathbf{B}_{\nu_0}]_{i,j} = \frac{\sin 2\pi(i+j)\nu_0}{2\pi(i+j)\nu_0} + \frac{\sin 2\pi(i-j)\nu_0}{2\pi(i-j)\nu_0} \quad (13)$$

The infinite dimension that is generally needed to represent the LPC parameters in the cepstral domain obviously imposes a practical problem. However, the magnitudes of the higher redundant cepstral coefficients decrease towards zero and for 10th order LP polynomials the sum can be truncated for coefficients  $n \geq N_{\text{max}}$ , where  $N_{\text{max}}$  is around 32 to 64, without significant loss of accuracy [3, 11]. Using the fact that the cepstral coefficients  $c_n$  for  $n > d$  are functions of the non-redundant coefficients  $\mathbf{c}_l = \{c_n\}_{n=1}^d$ , and the practical dimension limit  $N_{\text{max}}$ , we can for a high-resolution quantizer express the spectral distortion in terms of the  $d$  first non-redundant coefficients

$$SD_{\nu_0}^2 \approx 2 \cdot 10^2 \cdot (\log_{10} e)^2 (\mathbf{c}_l - \tilde{\mathbf{c}}_l)^T \mathbf{H}_{\nu_0}(\tilde{\mathbf{c}}_l) (\mathbf{c}_l - \tilde{\mathbf{c}}_l) \quad (14)$$

using a Taylor series expansion. This expression has a form similar to the high rate expressions in [5]. The weighting matrix has a structure

$$\mathbf{H}_{\nu_0}(\tilde{\mathbf{c}}_l) = \mathbf{B}'_{\nu_0} + 2\mathbf{B}''_{\nu_0} \mathbf{J} + \mathbf{J}^T \mathbf{B}'''_{\nu_0} \mathbf{J} \quad (15)$$

where the matrices have the components

$$\begin{aligned} [\mathbf{B}'_{\nu_0}]_{i,j} &= b_{i,j} & i, j &= 1, \dots, d \\ [\mathbf{B}''_{\nu_0}]_{i,j} &= b_{i+d,j} & i &= 1, \dots, d \\ & & j &= 1, \dots, N_{\text{max}} - d \\ [\mathbf{B}'''_{\nu_0}]_{i,j} &= b_{i+d,j+d} & i, j &= 1, \dots, N_{\text{max}} - d \\ [\mathbf{J}]_{i,j} &= \frac{\partial}{\partial c_j} c_{d+i}(\mathbf{c}_l) \Big|_{\mathbf{c}_l = \tilde{\mathbf{c}}_l} & i &= 1, \dots, N_{\text{max}} - d \\ & & j &= 1, \dots, d \end{aligned} \quad (16)$$

### 5.2. Adopting a Weighted Distortion Measure

A generalization of the high-rate formula (2) for a weighted distortion measure

$$\|\mathbf{x} - \mathbf{y}\|^r = \left( (\mathbf{x} - \mathbf{y})^T \mathbf{V}(\mathbf{x}) (\mathbf{x} - \mathbf{y}) \right)^{\frac{r}{2}} \quad (17)$$

with a data-dependent weight matrix  $\mathbf{V}(\mathbf{x})$ , has the same form as in (2) but  $D_H$  is in this case, with  $\mathbf{V}(\mathbf{x}) = \mathbf{H}_{\nu_0}(\mathbf{x})$

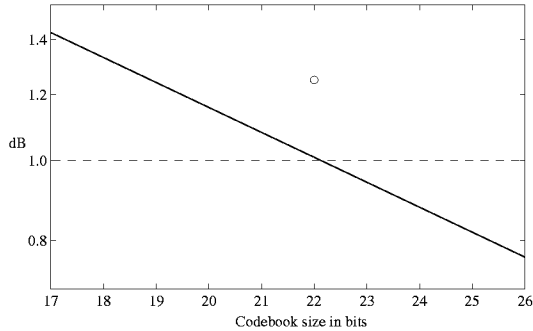
$$\begin{aligned} D_H &= \left( E_M \left[ (\det \mathbf{H}_{\nu_0}(\mathbf{x}))^{\frac{r}{2(d+r)}} f_{\mathbf{M}}^{\frac{-r}{d+r}}(\mathbf{x}) \right] \right)^{r/d} \\ &\cdot E_X \left[ (\det \mathbf{H}_{\nu_0}(\mathbf{x}))^{\frac{r}{2(d+r)}} f_{\mathbf{M}}^{\frac{-r}{d+r}}(\mathbf{x}) \right] \end{aligned} \quad (18)$$

Since the weight matrix  $\mathbf{H}_{\nu_0}(\mathbf{x})$  is independent of the density model parameters, the HRO algorithm must be only slightly modified to adopt a weighted distortion. The experiments presented in the paper use such a modified HRO algorithm.

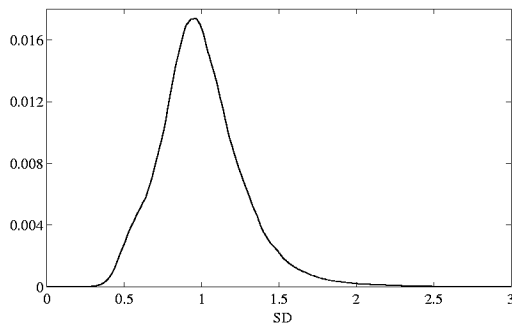
### 5.3. Minimum Required Number of Bits for Spectrum Coding

By combining the high rate expressions incorporating the effects of the higher cepstral coefficients and band-limited spectral distortion, we can now estimate the performance of single-stage VQ

of the LPC parameters. In Figure 2 we have plotted the estimated  $\overline{SD}_{MRS}$  in the 0–3 kHz band for 10-dimensional cepstrum VQ based on a 256-component bounded GM model. The model was estimated using the HRO algorithm and the higher cepstral coefficients were truncated at  $N_{max} = 64$ . According to the curve, a 22-bit quantizer will yield an  $\overline{SD}_{MRS}$  of 1 dB. The performance of a 22-bit 2-split LSF quantizer employing weightings according to [1] is also plotted for a comparison of the performance of typical contemporary VQ. The gap between the 2-split and the high rate prediction corresponds to approximately 3 bits.



**Figure 2:** Predicted  $\overline{SD}_{MRS}$  by the high rate approximation for 10-dimensional cepstrum VQ based on a 256-component bounded GM model. The model was estimated using the HRO algorithm. The performance of a 2-split LSF quantizer is marked by a circle.



**Figure 3:** Estimated pdf of the  $SD_{v_0}$  for a 22-bit 10-dimensional cepstrum quantizer based on a 256-component bounded GM model. The model was estimated using the HRO algorithm.

As previously mentioned, it is common for assessing spectrum quantization performance to report a spreading measure complementing the average distortion. It can be shown [10, 12], using high rate arguments, that the probability that the distortion is less than a value  $g$  can be expressed as

$$Pr(\|\mathbf{x} - \tilde{\mathbf{x}}\|^r \leq g) \approx \int_{\mathcal{R}^d} \min \left\{ \frac{g^{d/r} N \lambda(\mathbf{x}) V_d}{(\det \mathbf{H}_{v_0}(\mathbf{x}))^{1/2}}, 1 \right\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (19)$$

Note that (19) gives the distribution function of the high rate VQ distortion. The prevalent spreading measure for spectrum coding

is the percentage of outlier spectra having SD greater than 2 dB, which for the 22-bit quantizer is 0.7 %. Differentiation of (19) yields the pdf of the spectral distortion, which is depicted in Figure 3.

## 6. CONCLUSIONS

In this paper we model the underlying probability density function of vectors in a database as a Gaussian mixture model and estimate the parameters of the model. Having an analytical expression of the density, we then derived an expression for predicting the performance of vector quantization using high rate approximations. We apply the method to quantization of LPC-parameters and conclude that 22 bits are needed for meeting the standard requirements for spectrum coding. Further experiments are required to verify that such a 22-bit quantizer meets or exceeds subjective demands.

## 7. REFERENCES

- [1] K. K. Paliwal and B. S. Atal, “Efficient vector quantization of LPC parameters at 24 bits/frame,” *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 3–14, 1993.
- [2] K. K. Paliwal and W. B. Kleijn, “Quantization of LPC parameters,” in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., pp. 433–466. Elsevier Science Publishers, Amsterdam, The Netherlands, 1995.
- [3] P. Hedelin, “Single stage spectral quantization at 20 bits,” in *Proc. ICASSP*, Adelaide, Australia, 1994, vol. 1, pp. 525–528.
- [4] R. M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [5] W. R. Gardner and B. D. Rao, “Theoretical analysis of the high-rate vector quantization of LPC parameters,” *IEEE Trans. Speech Audio Processing*, vol. 3, no. 5, pp. 367–381, 1995.
- [6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.
- [7] P. Knagenhjelm, *Competitive learning in robust communication*, PhD dissertation, Chalmers University of Technology, 1993.
- [8] D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.
- [9] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. of Royal Stat. Soc.*, vol. 39, pp. 1–38, 1977.
- [10] J. Skoglund, *From Modeling to Perception - Topics in Speech Coding*, PhD dissertation, Chalmers University of Technology, 1998.
- [11] R. Hagen, “Spectral quantization of cepstral coefficients,” in *Proc. ICASSP*, Adelaide, Australia, 1994, vol. 2, pp. III3–III6.
- [12] D. H. Lee and D. L. Neuhoff, “Asymptotic distribution of the errors in scalar and vector quantizers,” *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 446–460, 1996.