# Rate Distortion Modeling of Spherical Vector Quantization Performance in Transform Audio Coding

**Wisarn Patchoo**[1] and **Thomas R. Fischer**[2], Non-members


## ABSTRACT

A block-based Gaussian mixture model (GMM) is used to model the distribution of the transform audio data encoded using spherical vector quantization and lossless coding. The expectation-maximization algorithm is used to design the GMM to model the marginal density of the transform coefficients and the block energy density. A GMM-based rate-distortion function is derived and shown to closely match the observed spherical VQ performance.


**Keywords**: Gaussian Distribution, Audio Coding, Rate Distortion Theory, and Vector Quantization

## 1. INTRODUCTION

Transform coding is an effective approach to audio coding. Such transformations include the discrete Fourier transform (DFT) [1, 2], the discrete cosine and modified discrete cosine transforms, and subband decomposition [3]. One recent example is DFT-based transform coded excitation (TCX) used in the adaptive multi-rate wideband audio coding algorithm [1]. The transform coding consists of three steps. First, the data sequence is divided into frames of size N and then a given transformation is performed on each frame. The second step is quantizing the transformed sequence subject to a fixed rate per frame constraint. The final step is encoding the quantized transformed sequence into a binary bitstream [4].

Spherical vector quantization (SVQ) has been shown to be an efficient way to quantize audio transform data [5, 6] and has been used in an audio coding standard [1, 2]. SVQ can be structured as a type of multi-rate, classified vector quantizer [7] that uses a product code to encode lattice codevectors as binary codewords. The product code consists of 1) a code for representing the codevector energy (squared radius), and 2) a code for representing a lattice codevector, conditioned on the codevector energy. The product code partitions the lattice codevectors into concentric

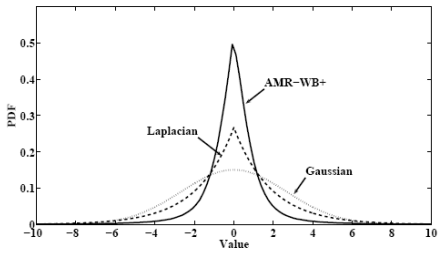



***Fig.1:*** *Empirical density of transform coefficients compared to that of memoryless Gaussian and Laplcian random variable with the same mean and variance.*

"shells" of codevectors. SVQ construction is motivated by the spherical geometry of the high probability volume of a memoryless Gaussian source probability density function [8, 9]. The lattice SVQ in [5, 6] is relatively simple to implement and remarkably effective, with an observed operational rate-distortion performance (for encoding audio transform data) significantly better than the memoryless Gaussian rate-distortion function.

As shown in Fig. 1, the audio transform coefficients are reasonably well modeled as having marginal Gaussian or Laplacian densities [5, 6]. However, memoryless Gaussian and Laplacian rate-distortion functions [10] are poor estimators of actual SVQ performance in transform audio coding, as will be shown later in this paper. This is due to the strong energy dependence in transform audio coefficient data. This can be seen in Fig. 2, which compares the empirical probability density function (pdf) of audio transform coefficient block squared radius (the block energy) to the block squared radius of memoryless Gaussian data, and Laplacian data, for block sizes $L = 4$, 8, 16, and 32. Clearly, for every block size the empirical density of the transform audio block squared radius differs significantly from that of memoryless Gaussian or Laplacian data of the same mean and variance. Since the correlation between coefficients is small, the empirical density is indicative of non-linear (energy) dependence in the transform coefficient data.

A Gaussian mixture model (GMM) [11, 12] is used in this paper to model vectors of audio transform co-

Manuscript received on August 1, 2010 ; revised on January 19, 2011.

This paper is extended from the paper presented in ECTI-CON 2010.

This work was supported by a Bangkok University faculty development program.

Portion of the results presented in this paper were reported in [29]

[1,2] The authors are with School of EECS, Washington State University, Pullman, WA, USA, E-mail:wpatchoo@eecs.wsu.edu, fischer@eecs.wsu.edu.

efficients. It is shown in [13] that any continuous pdf can be approximated by a Gaussian mixture density.

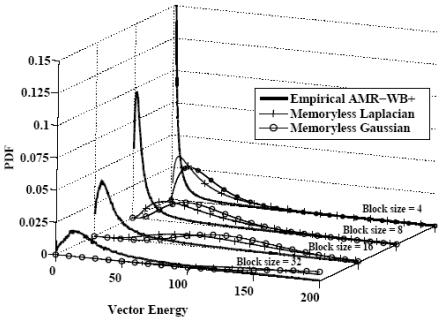



***Fig.2:*** *Empirical density of transform coefficient vector energy compared to that of memoryless Gaussian and Laplacian vector.*

A GM model has been successfully employed in several areas, e.g., speech recognization [14], speech coding [1517], image coding [18], etc. In [19], a generalized Gaussian (GG) mixture model is used to model image subband coefficients. A lattice VQ encodes lattice codevectors partitioned in "shells" matched to the GG shape parameters [20]. The experimental results in [19] focus on a mixture of Laplacian densities and report GG mixture modeling average relative mean squared error distortion within 6 to 20 percent of empirical rate-distortion LVQ performance, with maximum relative error as large as 32 percent.

In this paper, we focus on modeling the performance of lattice spherical vector quantization (SVQ) in transform audio coding. This is done in two steps. First, we model the transform audio coefficient data using a Gaussian mixture model. Then, a rate-distortion function based on the GM model is developed and used to estimate the performance of SVQ. GM model parameters are estimated using the Expectation-Maximization (EM) algorithm [11, 12, 22] and two alternative methods to estimate model parameters are proposed. As an application example of the proposed method, the model developed is shown to accurately describe the $RE_8$ lattice SVQ performance used in [1].

The outline of this paper is as follows. The Gaussian mixture model and EM algorithm are described in Section 2. In Section 3, a rate-distortion function based on the GM model is developed. The effectiveness of the model is evaluated in Section 4 by comparing the GM-based rate-distortion model to the cubic lattice SVQ performance, and to the RE8 lattice VQ performance used in the AMR-WB+ standard. Conclusions are discussed in Section 5.

## 2. GAUSSIAN MIXTURE MODEL

A $K$-class Gaussian mixture pdf for $L$-dimensional random vector $U$ is a parameterized function of the form

$$f_{mix}(u|\Theta) = \sum_{k=1}^{k} P(k) f_{\mathrm{U}|\Theta}(u|\theta_k) \tag{1}$$

where $P(k)$ denotes the prior probability or the probability that $U$ is generated by the $k^{th}$ class, and the component distribution, $f_{U|\Theta}(u|\theta_k)$, is a multivariate Gaussian distribution defined as

$$f_{\mathrm{U}|\Theta}(u|\theta_k) = \frac{1}{(2\pi)^{L/2}|\mathbf{C}_k|^{1/2}} e^{\{-\frac{1}{2}(x-\mu_k)^T \mathbf{C}_k^{-1}(x-\mu_k)\}} \tag{2}$$

where $\mu_k$ and $C_k$ are the mean vector and covariance matrice of the $k^{th}$ class, respectively. The mixture models parameters are defined as the set $\Theta = \{P(1), \ldots, P(K), \theta_1, \ldots, \theta_K\}$, where $\theta_k = \mu_k, \mathbf{C}_k$, for $k = 1, \ldots, K$.

### 2.1 Gaussian mixture model for audio transform coefficients

Let $X$ be a real-valued sequence of length $N$ to be quantized and encoded, formed from consecutive transform coefficients. Assuming $L$ divides $N$, partition $X$ into $N/L$ real-valued vectors (blocks) of size $L$, denoted as $Y$.The transformation is assumed to remove the linear dependence in $X$, and hence also in $Y$. Also, it is clear from Fig. 1 that $X$ has zero mean. We further assume a stationary property in each component class. So, now $\theta_k = \{0, \sigma_k^2\}$. Therefore, a $K$-class, $L$-dimensional Gaussian mixture model in (1) for vector $Y$ reduces to

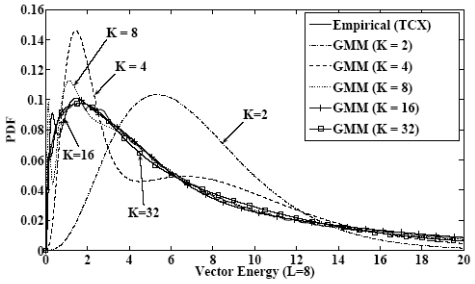
$$f_{mix}(y|\Theta) = \sum_{k=1}^{k} P(k) f_{\mathrm{Y}|\sigma_k^2}(y|\sigma_k^2) \tag{3}$$

where $f_{\mathrm{Y}|\sigma_k^2}(y|\sigma_k^2) = \frac{1}{(2\pi\sigma_k^2)^{L/2}} \exp(-\frac{1}{2\sigma_k^2}\sum_{l=1}^{L} y_l^2)$.

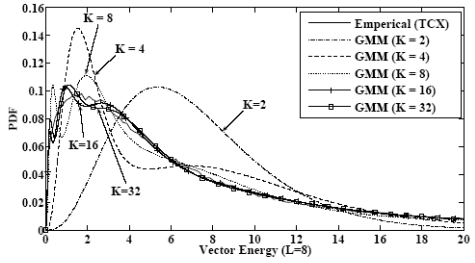
An alternative way to define the mixture model for audio transform coefficients is based on block or vector energy of $\mathbf{Y}$. Let the normalized energy be $Z = \frac{\epsilon}{\sigma_X^2}$, where $\epsilon = \sum_{l=1}^{L} y_l^2$ is the block energy (square radius) of the vector $\mathbf{Y}$ , and $\sigma_X^2$ is the variance of $X$. Suppose $\mathbf{Y}$ is generated from the $k^{th}$ class. Write the block energy as $Z = \frac{\sigma_k^2}{\sigma_X^2}\frac{\epsilon}{\sigma_k^2} = w_k Z_k$, where $w_k = \frac{\sigma_k^2}{\sigma_X^2}$ and $Z_k = \frac{\epsilon}{\sigma_k^2} k$ . The components of $\mathbf{Y}$ are independent and identically distributed and thus $Z_k$ is Chi-square distributed [21]. Then, the mixture model of block energy $Z$ can be expressed as

$$f_{mix}(z|\Theta) = \sum_{k=1}^{K} P(k) f_{z|\sigma_k^2}(z|\sigma_k^2) \tag{4}$$

where $f_{z|\sigma_k^2} = \frac{1}{w_k}\hat{f}\left(\frac{z}{w_k}\right)$ and $\hat{f}(z)$ is Chi-square distributed with degree of freedom $L$, defined by $\hat{f}(z) = \frac{1}{2^{L/2}\Gamma(L/2)} z^{L/2-1} e^{-z/2}, z \geq 0$.

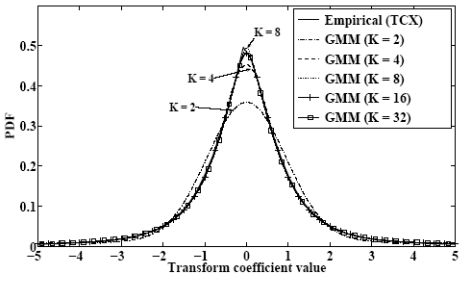



(a)

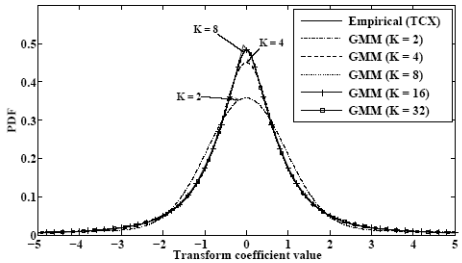



(b)

***Fig.3:*** *The Gaussian mixture model of transform coefficient block square radius compared to the empirical density: (a) Model based on (5) and (6); (b) Model based on (7) and (8)*




(a)




(b)

***Fig.4:*** *The Gaussian mixture model marginal density of transform coefficients compared to the empirical density (L = 8): (a) Model based on (5) and (6); (b) Model based on (7) and (8);*

The model parameters $P(k)$ and $\sigma_k^2$, for $k = 1, \ldots, K$, can be estimated by several methods such as Expectation-Maximization (EM) algorithm [11, 12], Markov-chain Monte Carlo algorithm [22], and Lloy clustering procedure [23]. In this paper, the EM algorithm is used since it is an algorithm widely used for finite mixture modeling.

Let $\mathbf{y}_m$ denote the $m^{th}$ block of $M$ total blocks and let $y_m, l$ denote the $l^{th}$ component of $\mathbf{y}_m$, for $l = 1, \ldots, L$. From [12], the EM algorithm requires introduction of auxiliary variables, $w_{m,k}$, that represent how likely block $\mathbf{y}_m$ is generated by the $k^{th}$ class, for blocks $m = 1, \ldots, M$ and classes $k = 1, \ldots, K$. From [12] and (3), the expectation and maximization steps of the EM algorithm are as follows.

**E-Steps**

$$E\left[w_{m,k}\right] = \frac{f_{\mathbf{Y}|\sigma_{\mathbf{k}}^{\mathbf{2}}}(y_m|\sigma_k^2)P(k)}{\sum_{j=1}^{M} f_{\mathbf{Y}|\sigma_{\mathbf{j}}^{\mathbf{2}}}(y_m|\sigma_j^2)P(j)} \tag{5}$$

$m = 1, \ldots, M$ and $k = 1, \ldots, K$.

**M-Steps**

$$\sigma_k^2 = \frac{\sum_{m=1}^{M} E\left[w_{m.k}\right]\left(\frac{1}{L}\sum_{l=1}^{L} y_{m,l}^2\right)}{\sum_{m=1}^{M} E\left[w_{m,k}\right]} \tag{6}$$

for $k = 1, \ldots, K$, where $E[\cdot\cdot]$ denotes expectation. Alternatively, let $z_m$ be the normalized block energy of $\mathbf{y}_m$, for $m = 1, \ldots, M$. Similar to above, by observing normalized block energy, the E-M steps for estimating the model parameters of the mixture model in (4) are as follows.

**E-Steps**

$$E\left[w_{m,k}\right] = \frac{f_{z|\sigma_k^2}(z_m|\sigma_k^2)P(k)}{\sum_{j=1}^{K} f_{z|\sigma_j^2}(z_m|\sigma_j^2)P(j)} \tag{7}$$

for $m = 1, \ldots, M$ and $k = 1, \ldots, K$.

**M-Steps**

$$\sigma_k^2 = \frac{\sum_{m=1}^{M} E\left[w_{m,k}\right] z_m \sigma_X^2}{k\sum_{m=1}^{M} E\left[w_{m,k}\right]} \tag{8}$$

The EM algorithm is used to estimate the mixture model parameters using transform coefficient vectors computed from a database of two minutes of wideband audio, 20% speech (two male and two female talkers) and 80% music (from nine different recordings). The resulting GMM energy density is compared to the empirical density of the transform coefficient vector energy in Fig. 3 for $K = 2$, 4, 8, 16, and 32 classes. The transform coefficient marginal density of the GMM is also compared to the empirical density in Fig. 4. It is clear that as the number of classes increases, the mixture models from both methods provide good approximations to both the vector energy density and the marginal density. However, one empirical observation is that GMM parameters based on (7) and (8) converge faster than those based on (5) and (6).

## 3. RATE DISTORTION FUNCTION OF GAUSSIAN MIXTURE MODEL

For small mean-squared error (MSE) distortion, $D$, the rate-distortion performance of entropy-coded quantization of a memoryless Gaussian source is

$$R(D) = \frac{1}{2}\log_2\left(\frac{2\pi e}{12\beta}\cdot\frac{\sigma^2}{D}\right) \tag{9}$$

where $\sigma^2$ is the source variance and $1 \le \beta \le (2\pi e/12)$ reflects the granular gain, also called the space filling advantage [24], of the quantization method. For uniform quantization, $\beta = 1$ and (9) is the Gish-Pierce asymptote [25]. For vector quantization using the $E_8$ lattice [26, 27], $\beta \approx 1.16$ (or 0.65 dB). For $\beta = 2\pi e/12$, (9) is the rate-distortion function for the memoryless Gaussian source [28].

Now, consider a $K$-class Gaussian mixture source model of vector (block) length L. Each class is modeled to have block components that are independent and identically distributed (i.i.d), as mention in Section 3. Using (9), the rate-distortion function for quantization and encoding the $k^{th}$ class can be expressed as

$$R_k(D) = \frac{1}{2}\log_2\left(\frac{2\pi e}{12\beta}\cdot\frac{\sigma_k^2}{D}\right) \tag{10}$$

where $\sigma_k^2$ is the $k^{th}$ class variance and D is assumed small compared to $\sigma_k^2$. One coding strategy is to first classify a source vector and then quantize and encode that vector conditioned on the class. Additional rate is necessary to specify the block class. As we assume an i.i.d sequence of source blocks, the minimum rate for encoding the class is the entropy, $H(k) = -\sum_{k=1}^{K} P(k)\log_2 P(k)$ bits/block, where $P(k)$ is the probability of class $k$. The average encoding rate for classification-based quantization and encoding is thus modeled as

$$R_{GMM}(D) = \frac{1}{L}H(K) + \sum_{k=1}^{K} P(k)\frac{1}{2}\log_2\left(\frac{2\pi e}{12\beta}\cdot\frac{\sigma_k^2}{D}\right) \tag{11}$$

$$= \frac{1}{L}H(K) + \frac{1}{2}\log_2\left(\frac{2\pi e}{12\beta D}\Pi_{k=1}^{K}(\sigma_k^2)^{P(k)}\right) \tag{12}$$

Define the first term in (12) as the classification rate, $R_K = \frac{1}{L}H(K)$, and the second term as the rate conditioned on the classification, $R_{class}(D)$.

## 4. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the mixture model rate-distortion function in (12), first we perform spherical VQ similar to [1]. However, for simplicity, the $Z_8$ lattice is used instead of the $RE_8$ lattice, and we compare the estimated encoding rate from (12) to the spherical VQ performance. Then, later in this section, we use (12) to estimate the encoding rate of $RE_8$ lattice spherical VQ in the AMR-WB+ standard [1].

***Table 1:*** *$R_{class}(D)$ corresponding to Table 2 (bits per sample)*

| $\mathbf{SVQZ_8}$ | $R_{class}(D)$ | | | | | |
|---|---|---|---|---|---|---|
| | ***K=1*** | ***K=2*** | ***K=4*** | ***K=8*** | ***K=16*** | ***K=32*** |
| $\Delta = 0.5$ | 4.46 | 3.44 | 3.19 | 3.14 | 3.14 | 3.15 |
| $\Delta = 1.0$ | 3.47 | 2.46 | 2.20 | 2.16 | 2.16 | 2.17 |
| $\Delta = 2.0$ | 2.54 | 1.53 | 1.28 | 1.28 | 1.27 | 1.29 |

***Table 2:*** *Estimated rate, $R_{GMM}(D)$ in (12), and empirical average encoding rate of $SVQZ_8$ at various step sizes*

| K | Average rate (bits/sample) | | | | | |
|---|---|---|---|---|---|---|
| | $\Delta = 0.5$ | | $\Delta = 1.0$ | | $\Delta = 2.0$ | |
| | $R_{GMM}(D)$ | $\mathbf{SVQZ_8}$ | $R_{GMM}(D)$ | $\mathbf{SVQZ_8}$ | $R_{GMM}(D)$ | $\mathbf{SVQZ_8}$ |
| **1** | 4.46 | | 3.47 | | 2.54 | |
| **2** | 3.53 | | 2.54 | | 1.61 | |
| **4** | 3.40 | 3.39 | 2.41 | 2.42 | 1.49 | 1.53 |
| **8** | 3.47 | | 2.50 | | 1.61 | |
| **16** | 3.56 | | 2.58 | | 1.69 | |
| **32** | 3.69 | | 2.72 | | 1.84 | |

The rate required for lossless coding of the $Z_8$ SVQ codevectors is determined as follows. Let $v$ be a $Z_8$ codevector with squared radius $r = \sum_{i=1}^{L} v_i^2 = ||v||^2$. A product code is used to encode $v$, consisting of two parts: 1) a code is used to specify the sphere of squared radius $r$, and 2) a code is used to specify the codevector on a given sphere. The ideal required rate can be expressed as

$$R_i = \frac{1}{L}\left[\log_2(1/P(r)) + \log_2 N(r)\right] \quad \text{bits/sample}, \tag{13}$$

where $P(r)$ is the probability that $v$ has squared radius $r$ and $N(r)$ is the number of $Z_8$ lattice points that lie on the sphere. $N(r)$ can be computed off-line from the theta function for the $Z_8$ lattice [26]. In practice the allowed range of lattice codevector radius can be truncated, and *overload* lattice codevectors losslessly encoded using the method of Voronoi extension (as in [1, 6]), or by simply partitioned the overload codevector into subblocks, and using a separate lossless code to encode the subblocks. In the experimental results to follow, the latter method is used, together with (13), to compute $Z_8$ SVQ encoding rates.

The source data are the spectrally pre-shaped and scaled transform coefficients from the AMR-WB+ encoding method. The transform coefficients are quantized using the scaled $Z_8$ lattice, and the average encoding rate computed using (13). The squared error distortion is controlled in the simulations by adjusting the $Z_8$ lattice step size, and is computed as

$$D = \frac{1}{L\cdot M}\sum_{m=l}^{M}\sum_{l=1}^{L}(y_{m,l} - \hat{y}_{m,l})^2, \tag{14}$$

where $M$ is the number of data vectors, $L = 8$ is the

***Table 3:*** *$R_{class}(D)$4 corresponding to Table 4 (bits per sample)*

| AMR-WB+ rate (kbps) | $R_{class}(D)$ K=1 | K=2 | K=4 | K=8 | K=16 | K=32 |
|---|---|---|---|---|---|---|
| 10.4 | 1.57 | 0.64 | 0.61 | 0.60 | 0.59 | 0.59 |
| 16.8 | 2.06 | 1.08 | 0.95 | 0.90 | 0.90 | 0.90 |
| 24.0 | 2.52 | 1.50 | 1.26 | 1.26 | 1.25 | 1.27 |

***Table 4:*** *Comparison between $R_{GMM}(D$ in (12) and average rate using $RE_8$ in AMR-WB+*

| K | 10.4 kbps $R_{GMM}(D)$ | AMR WB+ | 16.8 kbps $R_{GMM}(D)$ | AMR WB+ | 24.0 kbps $R_{GMM}(D)$ | AMR WB+ |
|---|---|---|---|---|---|---|
| 1 | 1.57 | | 2.06 | | 2.52 | |
| 2 | 0.72 | | 1.16 | | 1.58 | |
| 4 | 0.81 | 0.6 | 1.15 | 1.05 | 1.46 | 1.54 |
| 8 | 0.93 | | 1.23 | | 1.59 | |
| 16 | 1.00 | | 1.32 | | 1.67 | |
| 32 | 1.13 | | 1.43 | | 1.81 | |

***Table 5:*** *Results of the comparison based on the modified modeling*

| K | 10.4 kbps $R_{GMM}(D)$ | AMR WB+ | 16.8 kbps $R_{GMM}(D)$ | AMR WB+ | 24.0 kbps $R_{GMM}(D)$ | AMR WB+ |
|---|---|---|---|---|---|---|
| 1 | 0.75 | | 1.41 | | 2.12 | |
| 2 | 0.59 | | 1.02 | | 1.51 | |
| 4 | 0.58 | 0.60 | 1.00 | 1.05 | 1.45 | 1.54 |
| 8 | 0.62 | | 1.08 | | 1.54 | |
| 16 | 0.62 | | 1.10 | | 1.60 | |
| 32 | 0.66 | | 1.15 | | 1.67 | |

vector dimension coresponding to $Z_8$, and $\hat{\mathbf{y}}$ is the lattice SVQ codevector for $\mathbf{y}$. The distortion is thus the average squared error from lattice SVQ, and the rate, from (13) is the (idealized) spherical lattice VQ encoding rate. For a given lattice step size, the resulting simulation distortion, $D$, is used in (12) to determine the GMM estimate of encoding rate, $R_{GMM}(D)$.

The simulation results are summarized in Tables 1- 2, comparing the average rate required for spherical VQ using the scaled Z8 lattice ($SVQZ_8$) to the GMM rate-distortion function $R_{GMM}(D)$ in (12), with $\beta = 1$ (corresponding to $Z_8$ latice) and for several step sizes (equivalent to several signal-to-noise ratios, SNR). From Tables 1 and 2, it can be seen that the bit rate necessary to specify the block class, $R_K$, costs roughly 0.1 bits/dimension in classification rate for each doubling of the number of classes in the GMM. Examining Table 2 shows that for a single class (a memoryless Gaussian source model), the rate-distortion model over-estimates the rate by a significant margin. For effective rate-distortion modeling, $K = 4$ is a sufficient numbers of classes to capture the available classification gain, and increasing $K$ beyond 4 needlessly wastes rate in the modeling. This can be seen from Table 1 in which the conditional class encoding rate in (12), $R_{class}(D)$, saturates for $K \geq 4$. Note from Fig. 3 that the K-class mixture modeling estimate of the empirical block energy density continues to improve as K ranges from 1 to 32. For $K = 4$ the mixture model energy density is a rather coarse estimate of the empirical density. However, for modeling of rate-distortion performance, $K = 4$ classes is adequate. The 4-class GMM rate estimate is close to the (ideal) observed $Z_8$ VQ encoding rate, underestimating it by no more than 0.04 bits/sample.

The GMM rate-distortion function in (12) is used to estimate the average encoding rate of the RE8 lattice VQ in the AMR-WB+ algorithm. The value $= 1.16$ (corresponding with $RE_8$ lattice) is used in (12) [26, 27] for various distortions corresponding to different encoding modes of AMR-WB+ [1]. The results are shown in Tables 3-4. We note that the encoding in [1] uses a fixed rate per frame, whereas the GMM ratedistortion function in (12) does not impose this constraint. Hence, one expects the GMM rate-distortion mode to lower bound the observed $RE_8$ lattice VQ performance.

From Table 4, the rate-distortion modeling with $K = 4$ classes reasonably well models the AMR-WB+ rate at high rate (24.0 kbps), similar to the $SVQZ_8$ case. The results in Table 3 also demonstrate again that for rate-distortion modeling in (12), the number of classes, $K$, equal to 4 is enough to capture classfication gain of the mixture model.

At low and medium rates, however, the ratedistortion modeling in (12) does not adequately predict the AMR-WB+ encoding rate. The reason is that as the rate decreases, the frame gain (normalization factor) increases and the number of source vectors encoded as the zero codevectors increases. Thus, the overall distortion gets larger and the small distortion assumption in (12) is not valid. Some modifications have to be made in order to use rate-distortion modeling in (12) at low and medium rates.

A modification to the modeling approach is to use the GMM to model only significant source vectors, where significant means a source vector encoded using the AMR-WB+ $RE_8$ lattice VQ as a non-zero codevector. Using only significant source vectors, GMM parameters are again estimated using the EM algorithm, and the average encoding rate $R_{GMM}(D)$ is computed from (12). The total estimated rate, $R_{total}(D)$, is then computed by

$$R_{total}(D) = \frac{(\hat{R}_{GMM}(D) \times N_{nonzero}) + (R_{zero} \times N_{zero})}{N_{nonzero} + N_{zero}} \tag{15}$$

where $\hat{R}_{GMM}(D) = R_{GMM}(D) + 0.125$ is the estimated encoding rate for significant source vectors based on (12), plus an additional 0.125 bits/sample (or 1 bit/source vector) to distinguish the codeword as not having the same prefix as the zero vector codeword. $R_{zero}$ is the encoding rate for zero codevectors, which for AMRWB+ [1] is a fixed rate of 1 bit/vector (or 0.125 bits/sample). $N_{zero}$ and $N_{nonzero}$ are the number of zero codevectors and nonzero codevectors, respectively.

The estimated encoding rate based on the modified GMM approach is presented in Table 5. This

provides a better prediction of the observed AMR-WB+ encoding rates. Again, $K = 4$ is a sufficient number of classes.

## 5. CONCLUSION

This work accurately models the SVQ performance in transform audio coding using Gaussian mixture models. The Gaussian mixture model is used to model vectors of transform audio data and the EM algorithm is used to estimate GMM parameters. Two alternative methods are used to determine the mixture model parameters. A rate-distortion function based on GMM is developed and used to estimate the actual average encoding rate of SVQ. The effectiveness of the model is evaluated by comparing the estimated rate from the model with the average encoding rate of $Z_8$ lattice SVQ and with the $RE_8$ lattice VQ used in the AMRWB+ standard. The simulation results show that the estimated rate from the model with four classes reasonably well models SVQ performance, especially at high rate (small distortion). At low and medium rates, a modified model partitions source vectors into insignificant (encoded as zero codevector) and significant (encoded as non-zero codevector) classes. The GMM rate-distortion function is used only to estimate the encoding rate for nonzero source vectors since the encoding rate for zero source vector is fixed and predefined. With the modification, the model accurately estimates SVQ performance for low, medium, and high encoding rates. The results also indicate that GMM rate-distortion modeling with $K = 4$ classes is suffi- cient to capture the available classification gain of the mixture model for transform audio coding.

## References


[1] *Extended AMR Wideband Codec; Transcoding functions*, 3GPP TS 26.290, 2005.

[2] *G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729*, ITU-T G.729.1, 2006.

[3] *Information Technology Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s Part 3: Audio*, ISO/IEC 11172.3, 1993.

[4] K. Sayood, *Introduction to Data Compression*, 3rd ed., Morgan-Kaufmann, San Francisco, CA, 2006.

[5] M. Xie and J. P. Adoul, "Embedded algebraic vector quantization (EAVQ) with application to wideband audio coding," *Conf. Proceeding*, ICASSP, pp 240-243, 1996.

[6] S. Ragot, B. Bessette, and R. Lefebore, "Low-complexity multi-rate lattice vector quantization with application to wideband TCX speech coding at 32 kbps," *Conf. Proceedings*, ICASSP, pp. I501- I504, 2004.

[7] G. M. Gray, "Gauss mixture vector quantization," *Conf. Proceedings*, ICASSP, pp.1769-1772, 2001.

[8] D. J. Sakrison, "A geometric treatment of the source encoding of Gaussian random variable," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 481- 486, May 1968.

[9] T. R. Fischer and R. M. Dicharry, "Vector quantizer design for memoryless Gaussian, gamma, and Laplacian sources," *IEEE Trans. Commun.*, vol. COMM-32, pp. 1065-1069, Sep. 1984.

[10] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, New York, 1994.

[11] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195-239, 1984.

[12] J. H. Piater, Mixture Models and Expectation-Maximization, Lecture at ENSIMAG, May 2002, revised Nov. 2005.

[13] H. W. Sorenson and D. L. Aspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465-479, 1971.

[14] B. H. Juang, L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "Recent developments in the application of hidden markov models to speaker-independent isolated word recognition," *Conf. Proceedings*, ICASSP, pp. 9-12, 1985.

[15] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. on Speech and Audio Proc.*, vol. 8, no. 4, pp. 385- 401, July 2000.

[16] A. D. Subramaniam, W. R. Gardner, and B. D. Rao, "Low-complexity source coding using Gaussian mixture models, lattice vector quantization, and recursive coding with application to speech spectrum quantization," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 14, no. 2, pp. 524-532, Mar. 2006.

[17] D. Y. Zhao, J. Samuelsson, and M. Nilsson, "On entropy-constrained vector quantization using Gaussian mixture models," *IEEE Trans. Commun.*, vol. 56, no. 12, Dec. 2008.

[18] J. K. Su and R. M. Merserau, "Coding using Gaussian mixture and generalized Gaussian models," *Conf. Proceedings*, ICIP, pp. 217-220, 1996.

[19] L. Guillemot, Y. Gaudeau, S. Moussaoui, and J. M. Moureaux, "An analytical Gamma mixture based rate-distortionmodel for lattice vector quantization," *EUSIPCO*, 2006.

[20] P. Loyer, J. M. Moureaux, and M. Antonini, "Lattice codebook enumeration for generalized Gaussian source," *IEEE Trans. Inform. Theory*, vol. 49, No. 2 pp. 521-528, Feb. 2003.

[21] A. Papoulis and S. U. Pillai, *Probability, Random Variable, and Stochastic Processes*, 4th ed., McGraw-Hill 2002.

[22] D. Peel and G. Maclahlan, "Finite Mixture Models," Wiley interscience, 2000.

[23] Y. Z. Huang, D. B. OBrian, and R. M. Gray, "Classification of features and Images using Gauss mixtures with VQ clustering," *Conf. Proceedings*, Data Compression Conference (DCC), 2004.

[24] T. D. Lookabaugh and R. M. Gray, "Highresolution quantization theory and the vector quantization advantage," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1020-1033, Sept. 1989.

[25] N. Jayant and P. Noll, *Digital coding of waveforms: principles and applications to speech and video*, Prentice Hall Professional Technical Reference, 1990.

[26] J. H. Conway and N. J. A. Sloane, *Sphere Packing, Lattices, and Groups*, New York: Springer-Verlag, 1988.

[27] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211-226, Mar. 1982.

[28] T. M. Cover and J. A. Thomas, *Element of Information Theory*, 2nd ed., John Wiley & Son Inc., New Jersey, 2006.

[29] W. Patchoo and T. R. Fischer,"Gaussian-Mixture model of Lattice-based Spherical Vector Quantization Performance in Transform Audio Coding," *Conf. Proceedings*, ICASSP, pp. 197-200, 2010.

**Wisarn Patchoo** was born on August 23, 1978 in Bangkok, Thailand. He received Bachelor degree from Kasetsart University in 2000 and Master degree from King Mongkut's Institute of Technology Ladkrabang in 2003. Both are in Electrical Engineering. From 2002 to 2005, he was a Intelligent Network Engineer at the Huawei Technologies (Thailand). From 2006 until now, he is with School of Engineering, Bangkok University, Thailand. Currently, He is on leave to pursue a Ph.D atWashington State University. His interested researches are data compression, signal and image processing.

**Thomas R. Fischer** received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, and the Sc.B. degree magna cum laude from Brown University. From June 1975 until August 1976, he was a Staff Engineer at the Charles Stark Draper Laboratory, Cambridge, MA. From 1979 until 1988 he was with the Department of Electrical Engineering, Texas A&M University. Since January 1989 he has been a Professor in the School of Electrical Engineering and Computer Science at Washington State University. From 1999 until 2004 he was Director of the School of EECS. His current research interests include data compression, image and video coding, joint source/channel coding, digital communications, and digital signal processing. Professor Fischer has served as Associate Editor for Source Coding for the IEEE Transactions on Information Theory, a member of the Information Theory Society Board of Governors, Secretary of the Signal Processing and Communication Electronics Technical Committee of the IEEE Communications Society, and program evaluator for ABET. He is a regular reviewer for several journals, and has served on the Program Committee for several Workshops, Symposia, and Conferences. In 1987 Professor Fischer received an outstanding teaching award from the College of Engineering at Texas A&M University. Five times he has received departmental teaching awards at Washington State University. He was a co-recipient of the IEEE Signal Process- ing Society's 1993 Senior Award in the Speech Processing Area. In 1996 he was elected Fellow of the IEEE.