

# Predicting VQ Performance Bound for LSF Coding

Saikat Chatterjee, *Student Member, IEEE*, and T. V. Sreenivas, *Senior Member, IEEE*

**Abstract**—For vector quantization (VQ) of speech line spectrum frequency (LSF) parameters, we experimentally determine a mapping function between the mean square error (MSE) measure and the perceptually motivated average spectral distortion (SD) measure. Using the mapping function, we estimate the minimum bits/vector required for transparent quantization of telephone-band and wide-band speech LSF parameters, respectively, as 22 bits/vector and 36 bits/vector, where the distribution of LSF vector is modeled as a Gaussian mixture model (GMM).

**Index Terms**—Gaussian mixture model, line spectrum frequency (LSF) quantization, vector quantization.

## I. INTRODUCTION

FOR telephone-band speech coding, the problem of quantizing 10-dimensional LSF parameters has been thoroughly investigated in the literature [3], [4], [7], [10], [14]. Several studies also focus on determining the minimum bitrate required for “transparent quality” quantization of line spectrum frequency (LSF) parameters. Since it has been shown that 1 dB average spectral distortion performance (along with limited outliers) [3] will guarantee the transparency, there is much interest to determine the lowest bitrate required for transparent quality quantization performance. Such a rate-distortion (R/D) bound, determined for the full search vector quantization (VQ),<sup>1</sup> is very useful as a benchmark in the evaluation of different structured VQ methods used for LSF coding. Paliwal and Kleijn [7] have linearly extrapolated the operating (experimentally measured) R/D curve of a practical full search VQ and shown that at least 20 bits/vector is required for transparent quantization of telephone-band speech LSF parameters. Using the high rate quantization theory and Gaussian mixture model (GMM)-based analysis, Hedelin and Skoglund [8] have predicted the minimum bitrate as 22 bits/vector. This result is more conservative than the result of Paliwal and Kleijn. The issue of transparent quality LSF quantization becomes more important for wide-band speech, because of the higher quality concerns. Using the same engineering approach of [7], recently, So and Paliwal [12] have estimated the minimum bitrate required for transparent quantization of 16-dimensional wide-band speech LSF parameters as 31 bits/vector or 35 bits/vector, respectively, using either linear or exponential extrapolation. Since the high

rate VQ would provide an exponential R/D performance, the predicted bitrate of 35 bits/vector appears to be more acceptable from a practical point of view.

The approach of extrapolating the experimental R/D performance curve of a practical full search VQ [7], [12] is prone to significant variation using different extrapolating functions. Further, it is known that at lower bitrates, the quantization error is more signal dependent, and hence, the experimental results at lower rates may not be able to predict the VQ performance at higher rates through extrapolation. On the other hand, the GMM-based analysis in [8] requires the stochastic integration and an iterative algorithm for evaluating the high rate formulae to find the optimum distortion.

Recently, there is a growing interest in developing parametric VQ methods [8], [11] where the source PDF is modeled using finite mixture densities, such as the GMM. It can be shown that any continuous PDF can be approximated arbitrarily closely by a Gaussian mixture (GM) density [8]. The use of GMM provides a parametric and compact form of the source PDF which makes the VQ performance analysis task tractable. We have been able to determine a closed-form MSE performance expression of the VQ for quantizing a GM source [16]; we apply this result to predict the minimum bitrate required for transparent quality quantization of LSFs, both for telephone-band and wide-band speech cases. This is made possible by determining a mapping function between the MSE and the perceptually motivated average SD measure; the SD measure is commonly used for evaluating the LSF quantization performance in the literature. Thus, it becomes possible to map the theoretically derived MSE performance measures to the perceptually relevant distortion measures at any bitrate. Utilizing the 1 dB average SD threshold that has been used for both telephone-band and wide-band speech, we estimate the minimum bits/vector required for transparent quality LSF quantization. We then compare the estimated lower bound to the practical performances of established LSF quantization methods, such as the traditional split vector quantization (SVQ) [3] method and the recently proposed sequential SVQ (SeSVQ) [14] method.

## II. R/D PERFORMANCE OF VQ

We summarize here the basic high rate performance expression of VQ [6], [1] and mention the MSE performance expression for quantizing the GM source [16].

Let the PDF of  $h$ -dimensional vector  $\mathbf{g}$  be denoted by  $f_{\mathbf{g}}(\mathbf{g})$  and the source is quantized using  $b$  bits/vector. For quantization, we consider the weighted distortion measure as

$$\|\mathbf{g} - \hat{\mathbf{g}}\|_W^2 = [\mathbf{g} - \hat{\mathbf{g}}]^T \mathbf{W}(\mathbf{g}) [\mathbf{g} - \hat{\mathbf{g}}] \quad (1)$$

Manuscript received June 8, 2007; revised November 3, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Brian Kan-Wing Mak.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India (e-mail: saikat@ece.iisc.ernet.in; tvsree@ece.iisc.ernet.in).

Digital Object Identifier 10.1109/LSP.2007.914786

<sup>1</sup>In this letter, we use VQ and full search VQ interchangeably.

where  $\mathbf{W}(\mathbf{g})$  is a vector-dependent weighting matrix. Using the approach of asymptotic quantization theory, the high rate quantization distortion is given as [6], [8]

$$\begin{aligned} D_W &= \mathcal{E}\{\|\mathbf{g} - \hat{\mathbf{g}}\|_W^2\} \\ &\approx (2^b)^{-\frac{2}{h}} (V_h)^{-\frac{2}{h}} \frac{h}{h+2} \\ &\quad \times \left[ \int_{\mathcal{R}^h} \left[ |\mathbf{W}(\mathbf{g})|^{\frac{1}{h}} f_{\mathbf{g}}(\mathbf{g}) \right]^{\frac{h}{h+2}} d\mathbf{g} \right]^{h+2} \end{aligned} \quad (2)$$

where  $V_h$  is the volume of a  $h$ -dimensional sphere with unit radius, given by  $V_h = (\pi^{h/2})/((h/2)\Gamma(h/2))$ ;  $\Gamma(\cdot)$  is the usual Gamma function. Rigorously, (2) provides the lower bound of the distortion given the rate is high [1], [6], [8]. An important point to note is that deriving a closed-form expression for the standard PDFs using (2) is mathematically intractable, though the performance can be evaluated by the numerical method of stochastic integration.

In the case of an unweighted distance measure,  $\mathbf{W}(\mathbf{g}) = \mathbf{I}$  and (2) provides the MSE performance of VQ as [2]

$$\begin{aligned} D &= \mathcal{E}\{\|\mathbf{g} - \hat{\mathbf{g}}\|^2\} \\ &\approx (2^b)^{-\frac{2}{h}} (V_h)^{-\frac{2}{h}} \frac{h}{h+2} \left[ \int_{\mathcal{R}^h} [f_{\mathbf{g}}(\mathbf{g})]^{\frac{h}{h+2}} d\mathbf{g} \right]^{\frac{h+2}{h}}. \end{aligned} \quad (3)$$

For a standard PDF, such as multivariate Gaussian, the closed-form MSE performance expression can be found out in terms of the parameters of the respective PDF [15], and the experimental results closely match with the theoretical results [8], [16].

Let us evaluate the MSE performance of a VQ for quantizing a source which is modeled by a GM density. The GM density is a weighted sum of  $M$  Gaussian densities, given by

$$f_{\mathbf{g}}(\mathbf{g}) = \sum_{m=1}^M \gamma_m \mathbf{N}(\mu_{m,\mathbf{g}}, C_{m,\mathbf{g}}) \quad (4)$$

where  $\gamma_m$ ,  $\mu_{m,\mathbf{g}}$ , and  $C_{m,\mathbf{g}}$  are the prior probability, mean vector, and covariance matrix of the  $m$ th Gaussian component. Deriving a closed-form MSE performance expression for the GM density using (3) seems to be difficult. Hence, an expectation-maximization (EM) type of iterative algorithm is formulated in [8] to determine the optimum distortion using stochastic integration. Instead of the iterative approach, we have used the linearized approach of [11] to evaluate the MSE performance [16]. Thus, the MSE performance is given by a closed-form expression in terms of the GMM parameters as [16]

$$D_{\text{GM}} = (2^b)^{-\frac{2}{h}} K_h \left[ \sum_{m=1}^M \left[ \gamma_m |C_{m,\mathbf{g}}|^{\frac{1}{h}} \right]^{\frac{h}{h+2}} \right]^{\frac{h+2}{h}} \quad (5)$$

where  $K_h$  is a constant which is dependent on dimension  $h$  as [16]

$$K_h = 2 \left( \frac{h}{2} \Gamma \left( \frac{h}{2} \right) \right)^{\frac{2}{h}} \left( \frac{h+2}{h} \right)^{\frac{h}{2}}. \quad (6)$$

We mention that (5) is derived using the optimum bit allocation between the Gaussian components, and thus, it provides

the lower bound of the VQ performance for quantizing the GM source. The derivation of (5) is based on the assumption of minimally overlapping Gaussian components in the GM density. Thus, we use the well-known LBG algorithm instead of the standard EM algorithm to estimate the GMM parameters. The LBG algorithm is used to determine  $M$  number of non-overlapping Voronoi regions which are then modeled as mixture components of the GMM. Naturally,  $M$  needs to be sufficiently large for more accurate modeling of the source PDF.

### III. MINIMUM BITRATE REQUIRED FOR TRANSPARENT QUALITY LSF QUANTIZATION

#### A. Distortion Measure Mapping

LPC parameters represent important speech intelligibility information, and thus, the perceptually motivated objective measure, used to evaluate LSF quantization performance, is the spectral distortion (SD) [3], [6]. For the  $n$ th frame,  $SD_n$  in dB is defined as [3]

$$SD_n = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ 10 \log_{10} P_n(\Omega) - 10 \log_{10} \hat{P}_n(\Omega) \right\}^2 d\Omega \right\}^{\frac{1}{2}} \quad (7)$$

where  $P_n(\Omega)$  and  $\hat{P}_n(\Omega)$  are the original and quantized LP filter power spectra. The conditions for transparent quality coding of telephone-band speech LPC parameters are [3]

- 1) average SD is  $\leq 1$  dB;
- 2) no outlier having more than 4 dB SD;
- 3)  $< 2\%$  of outlier frames with 2–4 dB SD.

According to Guibe *et al.* [9], listening tests have shown that these conditions for transparency are also valid for the wide-band speech case, and hence, the conditions have been used in [12] and [13] also. In this letter, we will confine mainly to the condition (1) as in [7] and [12] and use the 1 dB average SD as the transparent quality quantization threshold for both telephone-band and wide-band speech. The other conditions require an estimate of the quantization error PDF unlike the mean error which has been investigated in this letter.

In LSF quantization, direct use of the SD measure is limited because of the difficulty of finding an expression for the generalized centroid in VQ design and also the high complexity to search the VQ codebook. As an approximation to SD, weighted square Euclidean distance (WSED) is used. The WSED measure is defined as

$$d_w(\mathbf{X}, \hat{\mathbf{X}}) = [\mathbf{X} - \hat{\mathbf{X}}]^T \mathbf{W}(\mathbf{X}) [\mathbf{X} - \hat{\mathbf{X}}] \quad (8)$$

where  $\mathbf{X}$  and  $\hat{\mathbf{X}}$  are the original and quantized LSF vectors, and  $\mathbf{W}(\mathbf{X})$  is a vector-dependent diagonal weighting matrix. Using high rate quantization theory, it has been shown [6] that the best weights are the spectral sensitivity coefficients.

For analytical evaluation, we note that neither SD nor WSED can be used to obtain the closed-form high-resolution R/D performance expression in terms of the PDF parameters [6], [1]. Thus, we use the square Euclidean distance (SED) measure<sup>2</sup> to estimate the R/D performance of LSF VQ. However, since the distortion threshold for transparent quality is available in terms

<sup>2</sup>MSE is the average SED measure.

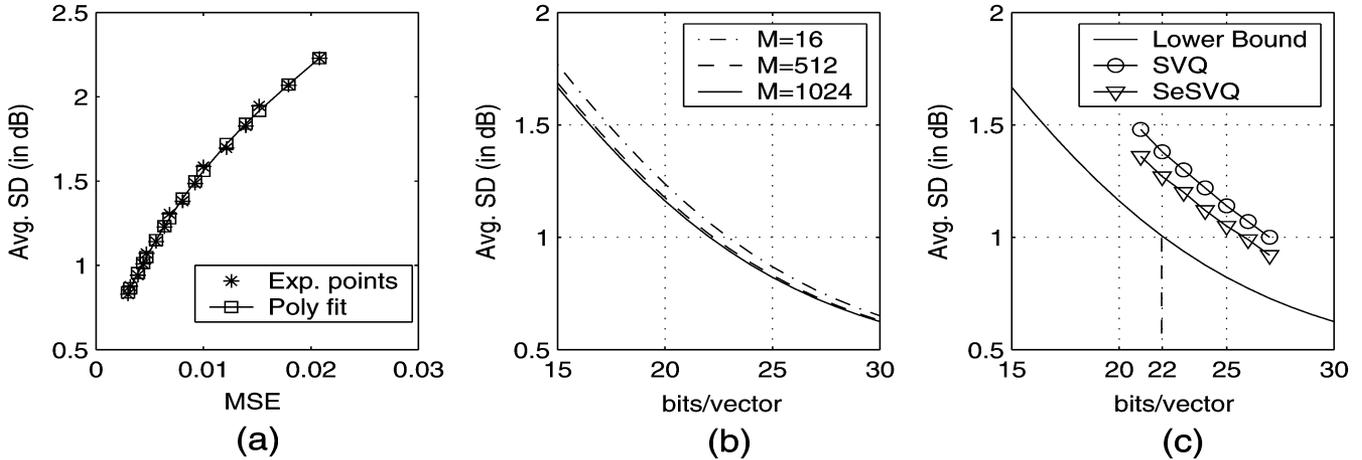


Fig. 1. Telephone-band speech: Estimated lower bound for quantizing the LSF parameters. (a) Mapping function using polynomial fitting between average SD and MSE. (b) Lower bounds using different model orders ( $M = 16$ ,  $M = 512$ , and  $M = 1024$ ). (c) Lower bound using  $M = 1024$  and comparison of the estimated lower bound with established quantization methods. Transparent quality quantization performance is achieved at 22 bits/vector.

of the average SD, we use an experimental approach to map the MSE (or average SED) measure to the average SD measure. The established SVQ method [3] is designed and tested using the WSED measure at different bitrates, providing a set of paired points of MSE and average SD. Using the paired points, we find the mapping function between MSE and average SD which is quite smooth but not linear. Hence, a third-order polynomial curve is fitted to provide the mapping of MSE to average SD. The two cases of telephone-band and wide-band speech are dealt with separately, determining two different mapping functions.

### B. Telephone-Band Speech

The speech data used in the experiments are from the TIMIT data base, and the analysis is similar to that reported in the literature. The 16 kHz speech is first low pass filtered to 3.4 kHz and then down sampled to 8 kHz. A tenth-order LPC analysis (i.e.,  $h = 10$ ) is used for a 20 ms Hamming windowed frame, based on the Burg method, with no successive frame overlap. In order to avoid sharp spectral peaks in the LPC spectrum, a fixed 10-Hz bandwidth expansion is applied as in [3] and the LPC parameters are converted to LSF parameters. The pooled data consisted of 368 815 LSF vectors for training and “out of training” 5000 LSF vectors are used for testing.

The three-split SVQ method of [14] is used to determine the mapping function between MSE and average SD. At 27 bits/vector, the SVQ method provides 1 dB average SD along with 1.32% of “2–4 dB” outliers and zero outlier “> 4 dB.” Thus, it can be said that the coding method providing 1 dB average SD threshold would also satisfy conditions (2) and (3) for transparent quality quantization requirements. Experimentally obtained paired points (of MSE and average SD) are shown in Fig. 1(a) as the discrete points, and the smooth third-order polynomial fit to the points is shown as the mapping function. The accuracy of polynomial fit is measured by the ratio of variance of experimental points to variance of estimation error, which is found to be 28.12 dB. The PDF of LSF source vector is modeled by GMM with different model orders. Using the estimated GMM parameters, the MSE performance of VQ is com-

puted employing the parametric expression of (5), and the mapping function is used to determine the corresponding average SD. Fig. 1(b) shows the estimated average SD performances for different model orders of the GMM. The different model orders indicate that higher  $M$  provides a tighter bound; this result is expected since a better model fit with higher number of Gaussian components is giving rise to lower distortion. The model order  $M = 1024$  can be viewed as a saturation level providing an estimation of the performance close enough to a full search VQ. Using the 1 dB average SD threshold, it is observed from Fig. 1(c) that 22 bits/vector is required for transparent quantization of telephone-band speech LSF parameters. We note that this result commensurates with the result of [8]. The experimental performances of two established LSF quantization methods are compared with the estimated lower bound. We can see that the practical performance of SVQ is inferior by 5 bits/vector. The recently proposed SeSVQ method [14] improves the performance by 1 bit/vector, but there is much scope for further improvement.

### C. Wide-Band Speech

In this case also, we use the TIMIT database for training and testing. We have used the specification of AMR-WB speech codec [17] to generate 16th order LP coefficients (i.e.,  $h = 16$ ), which are then converted to LSF parameters. Immittance spectral pairs (ISP) [5] are used in the AMR-WB speech codec, but we use LSF parameters in this letter which are claimed to provide similar performance. Like the telephone-band speech case, we use 368 815 LSF vectors for training and “out of training” 5000 LSF vectors for testing.

We resort to the five-split SVQ method [14] to determine the mapping function between the MSE measure and the average SD measure. At 47 bits/vector, the SVQ method provides 1.01 dB average SD along-with 0.36% of “2–4 dB” outliers and zero outlier “> 4 dB.” Thus, it can be said that the coding method providing 1 dB average SD threshold would also satisfy the transparent quality quantization requirements of minimum outliers. The nonlinear smooth mapping function is shown in Fig. 2(a). Like telephone-band case, the accuracy of the poly-

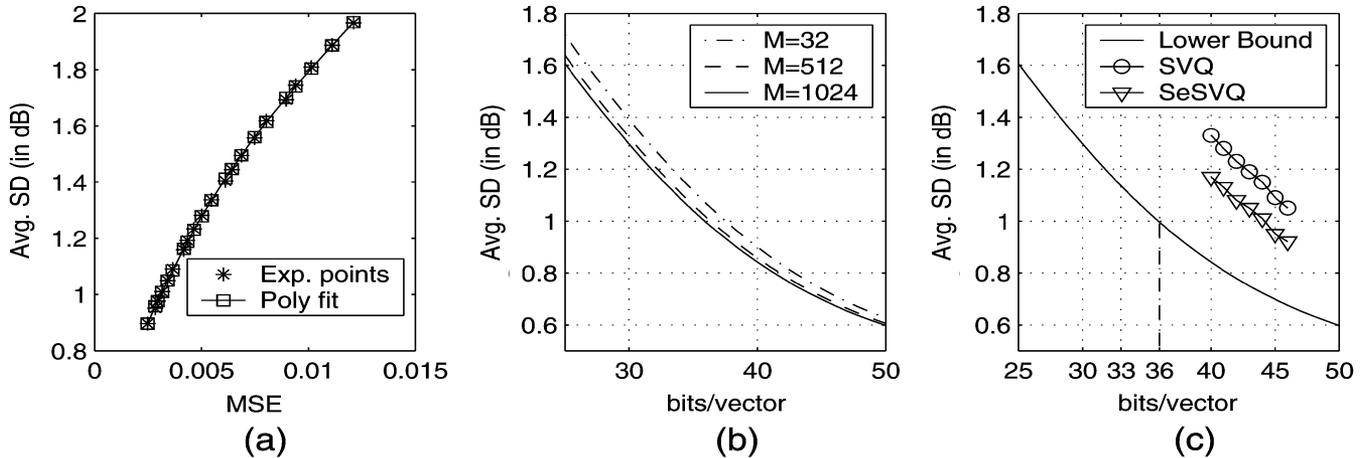


Fig. 2. Wide-band speech: Estimated lower bound for quantizing the LSF parameters. (a) Mapping function using polynomial fitting between average SD and MSE. (b) Lower bounds using different model orders ( $M = 32$ ,  $M = 512$ , and  $M = 1024$ ). (c) Lower bound using  $M = 1024$  and comparison of the estimated lower bound with established quantization methods. Transparent quality quantization performance is achieved at 36 bits/vector.

nomial fit is measured as 37.96 dB. The PDF of wide-band LSF data is again modeled by the GM density with different model orders. Using the estimated GMM parameters, the MSE performance of VQ is obtained using (5) at different bitrates and the average SD performances for LSF VQ are determined using the mapping function. The estimated results are plotted in Fig. 2(b) for different model orders. Again, the higher order models provide tighter lower bounds. Considering the model order  $M = 1024$  and the 1 dB threshold, it can be seen from Fig. 2(c) that the minimum bitrate for transparent quantization of wide-band speech LSF parameters is estimated to be 36 bits/vector. We note that this result is conservative by 1 bit/vector than the result of [12]. Comparing the estimated lower bound with the SVQ and SeSVQ methods, we can see that there is nearly 10 bits/vector performance gap to be bridged. The SeSVQ provides 2–3 bits/vector advantage over SVQ, but there is much scope for further improvement.

#### IV. CONCLUSION

Using the GMM-based parametric PDF framework, we estimate the minimum bitrate required for transparent quantization of the speech LSF parameters. For this, we use a functional mapping between the MSE measure and the perceptually motivated average SD measure. It may be noted that our experimental evaluation is quite generic and can be used for estimating the R/D performance bound of a full search VQ, through a mapping function between the MSE and the desired perceptual distortion measure.

#### REFERENCES

- [1] R. M. Gray, *Source Coding Theory*. Dordrecht, The Netherlands: Kluwer, 1990.
- [2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
- [3] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 1, pp. 3–14, Jan. 1993.
- [4] W. F. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 Kb/s speech coding," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 4, pp. 373–385, Oct. 1993.
- [5] Y. Bristritz and S. Pellerin, "Immitance spectral pairs (ISP) for speech coding," in *Proc. ICASSP*, 1993, pp. II-9–II-12.
- [6] W. R. Gardner and B. D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 5, pp. 367–381, Sep. 1995.
- [7] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam: Elsevier Science, 1995, pp. 433–466.
- [8] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 4, pp. 385–401, Jul. 2000.
- [9] G. Guibe, H. T. How, and L. Hanzo, "Speech spectral quantizers for wide-band speech coding," *Eur. Trans. Telecommun.*, vol. 12, no. 6, pp. 535–545, 2001.
- [10] L. Hanzo, F. C. A. Somerville, and J. P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*. New York: IEEE Press, 2001.
- [11] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 2, pp. 130–142, Mar. 2003.
- [12] S. So and K. K. Paliwal, "Empirical lower bound on the bitrate for the transparent memoryless coding of wideband LPC parameters," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 569–572, Sep. 2006.
- [13] S. So and K. K. Paliwal, "A comparative study of LPC parameter representations and quantisation schemes for wide-band speech coding," *Digit. Signal Process.*, vol. 17, no. 1, pp. 114–137, Jan. 2007.
- [14] S. Chatterjee and T. V. Sreenivas, "Sequential split vector quantization of LSF parameters using conditional pdf," in *Proc. ICASSP*, Honolulu, HI, Apr. 2007, vol. IV, pp. 1101–1104.
- [15] S. Chatterjee and T. V. Sreenivas, "Analysis of conditional PDF based split VQ," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 781–784, Nov. 2007.
- [16] S. Chatterjee and T. V. Sreenivas, "High rate analysis of conditional PDF based split VQ for Gaussian mixture source," *Signal Process.*, submitted for publication.
- [17] "AMR wide-band speech codec, transcoding functions (Release 5)," *3GPP TS 26.190 V 5.1.0*.