ANALYSIS-BY-SYNTHESIS BASED SWITCHED TRANSFORM DOMAIN SPLIT VQ USING GAUSSIAN MIXTURE MODEL

Saikat Chatterjee and T.V. Sreenivas

Department of Electrical Communication Engineering Indian Institute of Science, Bangalore, India 560012 Email: {saikat, tvsree}@ece.iisc.ernet.in

ABSTRACT

Using analysis-by-synthesis (AbS) approach, we develop a soft decision based switched vector quantization (VQ) method for high quality and low complexity coding of wideband speech line spectral frequency (LSF) parameters. For each switching region, a low complexity transform domain split VQ (TrSVQ) is designed. The overall rate-distortion (R/D) performance optimality of new switched quantizer is addressed in the Gaussian mixture model (GMM) based parametric framework. In the AbS approach, the reduction of quantization complexity is achieved through the use of nearest neighbor (NN) TrSVQs and splitting the transform domain vector into higher number of subvectors. Compared to the current LSF quantization methods, the new method is shown to provide competitive or better trade-off between R/D performance and complexity.

Index Terms— VQ, LSF quantization, GMM.

1. INTRODUCTION

Low complexity, but high quality vector quantization (VQ) of LSF parameters has attracted much attention in the current literature [5], [7], [8], [9], [10], [12], [13]. The complexity issues in VQ are more important to address for the applications where the requirement of higher perceptual quality is achieved through the allocation of higher bitrate. In these applications, such as in wideband speech coding, the complexity of LSF VQ is very high and hence, it is important to keep the complexity under check without sacrificing the R/D performance. One of the most cited low complexity VQ schemes is split VQ (SVQ) which was first proposed by Paliwal and Atal for telephone-band speech LSF coding [1] and then extended to wideband speech LSF coding [4]. SVQ is a product VQ scheme in which the LSF vector is split into subvectors and then quantized independently; this approach sacrifices the correlation between the subvectors and thus, leads to a coding loss, referred to as "split loss" [6]. To recover the split loss, So and Paliwal have recently proposed switched SVQ (SSVQ) method [12], [8] which is shown to provide better R/D performance than SVQ at lower computational complexity, but at the requirement of higher memory. In SSVQ, the vector space is divided into non-overlapping Voronoi regions¹ and a separate SVQ is designed for each region. An input vector to be quantized is first classified to a Voronoi region using NN criteria and then, the region specific SVQ is used for quantization. In a comparative study [11], So and Paliwal have shown the efficiency of SSVQ over several other structured VQ methods for quantizing the wideband LSF parameters. We also have proposed two stage VQ methods [10], [13] which are shown to provide a better trade-off between R/D performance and complexity. The SVQ, SSVQ and two stage VQ methods are non-parametric VQ methods whose R/D performances have been shown to be good using an experimental approach. Currently, there is much interest for designing and analyzing the VQ methods in a GMM based parametric framework [3], [5]. Using GMM based approach, we have recently addressed the R/D performance optimality of SSVQ method in [14] where the optimum SSVQ is shown to provide better performance than the non-parametric SSVQ.

In this paper, we seek for further reduction in complexity without sacrificing the R/D performance. For multi-variate Gaussian source coding, we have recently shown that the transform domain SVQ (TrSVQ) method [15] can recover the split loss of an SVQ through the use of a de-correlating transform. The transform domain approach provides the major advantages of low complexity and R/D performance optimality. Hence, we investigate the use of TrSVQ where the source PDF is modeled using GMM. Although this approach is effective for a general VQ problem, there is a specific limitation for LSF quantization, viz. the non-linear measure of spectral distortion $(SD)^2$ can not be used in the transform domain. We note that the analysis-by-synthesis (AbS) approach of Subramaniam and Rao [5] can be used for searching the optimum code-vector that produces the least SD. However, the AbS approach is computationally intensive than the direct quantization. We recognize that the refinement through AbS loop results in a code-vector which will be in the vicinity of input LSF vector and thus, this observation can be used for reducing the quantization complexity. In this paper, we explore the use of optimum TrSVQ in the AbS framework, which offers both low complexity as well as optimum R/D performance through the use of GMM. The new method is shown to provide competitive performance compared to the optimum SSVQ at lower complexity and better performance than the GMM based VQ (GMVQ) method of Subramaniam and Rao [5].

2. SOFT DECISION BASED ABS QUANTIZATION

In the AbS framework, we develop the soft decision based switched TrSVQ (SSTrSVQ) method and address its R/D performance optimality using a parametric model of the source PDF. The basis of the SSTrSVQ method is to divide the vector space into M number of non-overlapping Voronoi regions and design optimum TrSVQ for each region; the switching codebook consists of M mean vectors of Voronoi regions as the code-vectors. An input vector is quantized using L number of nearest neighbor (NN) TrSVQs (where L < M)

¹These Voronoi regions are referred to as "switching regions" in [12], [11]. In this paper, Voronoi region and switching region are used interchangeably.

²The perceptually motivated objective method used for evaluating the LSF quantization methods is the spectral distortion (SD) [1], [11].



Fig. 1. Soft decision based switched TrSVQ (SSTrSVQ)

and then the winning candidate, among the L quantized vectors, is selected through the AbS approach. If we choose L = 1, the SSTrSVQ method becomes a hard decision based switched quantizer like SSVQ. For SSTrSVQ, the increased computational complexity, due to the use of multiple TrSVQs (as L > 1), is kept under check through the use of low complexity TrSVQ method where the transform domain vector is split into higher number of subvectors. Fig. 1 shows the block diagram of SSTrSVQ method; the algorithmic steps are:

- 1. The input vector is compared with the *M* mean vectors of Voronoi regions (or switching regions) using square Euclidean distance (SED) measure and the SED distance values are rank ordered (sorted); according to the rank ordering, the *L* number of NN TrSVQs are chosen from *M* TrSVQs for quantization.
- 2. Quantize the input vector using the L NN TrSVQs.
- 3. Determine the best quantized vector using weighted square Euclidean distance (WSED) measure³.

2.1. Optimum R/D performance

The issue of R/D performance optimality is addressed using the GMM based framework. The source PDF of each switching region is modeled as a multi-variate Gaussian component of the GMM. Since the switching regions are nothing but Voronoi regions of the first stage quantizer, the mixture components of the GMM are assumed to be truncated and non-overlapping. Let X be the *p*-dimensional source whose PDF is modeled using a Gaussian mixture (GM) density of *M* components, given as

$$f_{\mathbf{X}}(\mathbf{x}) \approx \sum_{m=1}^{M} \alpha_m \, \mathcal{N}(\mu_{m,\mathbf{X}}, C_{m,\mathbf{X}}) \tag{1}$$

where α_m , $\mu_{m,\mathbf{X}}$ and $C_{m,\mathbf{X}}$ are the prior probability, mean vector and covariance matrix of the *m*th Gaussian component and $\sum_{m=1}^{M} \alpha_m = 1$. The approximate equality used in Eqn. 1 is due to the use of finite *M* and truncated components. While the direct R/D performance analysis of a GMM source is not easy to carry out [3], we consider a linearized approach [5] where the overall quantization distortion of the source is expressed as

$$\mathcal{D}_{\mathbf{X}} \approx \mathcal{D}_{GMM} = \sum_{m=1}^{M} \alpha_m \ \mathcal{D}_{G,m}(b_m) \tag{2}$$

where $\mathcal{D}_{G,m}(b_m)$ is the incurred distortion of the TrSVQ for the *m*th Gaussian component of GMM and b_m is the allocated bits/vector to the *m*th region specific TrSVQ. We need to solve for the optimum bit allocation, b_m , such that \mathcal{D}_{GMM} is minimized. Since, the source PDF of each Voronoi region is modeled as a multi-variate Gaussian with known parameters, we can use the high resolution quantization distortion of TrSVQ [15] for an overall distortion expression of Eqn. 2 and then optimize for b_m . The non-overlapping Voronoi regions (or switching regions) of SSTrSVQ method are determined using the LBG algorithm and we compute the Gaussian parameters for each Voronoi region. The switching codebook consists of $\{\mu_{m,\mathbf{X}}\}_{m=1}^{M}$ as the code-vectors.

2.1.1. Inter-cluster bit allocation

The optimality of the SSTrSVQ method is addressed by achieving the best R/D performance through the optimum bit allocation among the Gaussian components; this is referred to as inter-cluster bit allocation which can be carried out using the variable rate scheme [5].

In a variable rate scheme, the average bits/vector (b_{avg}) is fixed, but the actual bits/vector changes with the input vector. The first stage quantizer (switch) uses $b_c = \log_2 M$ bits, followed by the second stage quantization (TrSVQ) using b_m bits for the *m*th region. Thus, the coder becomes a variable rate coder because b_m varies according to the changing switching decision. The variable bitrate constraint is given as

$$b_{avg} = \sum_{m=1}^{M} \alpha_m \left(b_c + b_m \right) = b_c + \sum_{m=1}^{M} \alpha_m \, b_m, \tag{3}$$

where b_{avg} is a design parameter as in a fixed rate coder. The optimum inter-cluster bit allocation is solved by minimizing the overall distortion, given in Eqn. 2, to the constraint of average bits, shown in Eqn. 3. Evaluation of \mathcal{D}_{GMM} requires the best performance of $\mathcal{D}_{G,m}(b_m)$, which is given by the quantization distortion result of TrSVQ method as (using Eqn. (9) of [15])

$$\mathcal{D}_{G,m}(b_m) = \left(2^{b_m}\right)^{-\frac{2}{p}} \mathcal{K}_m \left|C_{m,\mathbf{X}}\right|^{\frac{1}{p}},\tag{4}$$

where

$$\mathcal{K}_m = \left[p \left[\prod_{i=1}^{S_m} \left(\frac{K_{q_{m,i}}}{q_{m,i}} \right)^{p_{m,i}} \right]^{\frac{1}{p}} \right] \text{ and}$$
$$K_{q_{m,i}} = 2 \left[\frac{q_{m,i}}{2} \Gamma\left(\frac{q_{m,i}}{2}\right) \right]^{\frac{2}{q_{m,i}}} \left[\frac{q_{m,i}+2}{q_{m,i}} \right]^{\frac{q_{m,i}}{2}}$$

Here, $\Gamma(.)$ is the usual Gamma operator and $q_{m,i}$ is the dimension of *i*th split subvector in transform domain for the *m*th cluster specific TrSVQ method $(1 \le i \le S_m)$. Note that, for the *m*th region, the number of splits (S_m) and the dimensions of split subvectors $(q_{m,i})$ are design parameters which can be different for different Voronoi regions according to the choice of the designer. The optimization is only for evaluating b_m given S_m and $q_{m,i}$; the optimization is carried out using Lagrange method.

The optimum inter-cluster bit allocation for SSTrSVQ method that minimizes the overall distortion of Eqn. 2, subject to the variable rate constraint of Eqn. 3 is given by

$$b_m = (b_{avg} - b_c) + \frac{p}{2} \times \log_2 \left[\frac{\kappa_m |C_{m,\mathbf{x}}|^{\frac{1}{p}}}{\prod_{j=1}^M \left[\kappa_j |C_{j,\mathbf{x}}|^{\frac{1}{p}}\right]}, \ 1 \le m \le M.$$
(5)

³The direct use of SD in VQ is complexity limited. It is shown in [2] that the SD can be approximated using WSED measure where the spectral sensitivity coefficients are the weighting values.

2.1.2. Intra-cluster bit allocation

For each cluster (switching region), the optimum bit allocation to the transform domain subvectors in TrSVQ method is given as (using Eqn. 8 of [15])

$$b_{m,i} = q_{m,i} \frac{b_m}{p} + \frac{q_{m,i}}{2} \times \log_2 \left[\frac{\frac{Kq_{m,i}}{q_{m,i}} |C_{m,\mathbf{Z}_i\mathbf{Z}_i}|^{\frac{1}{q_{m,i}}}}{\left[\prod_{k=1}^{S_m} \left(\frac{Kq_{m,k}}{q_{m,k}} \right)^{q_{m,k}} \right]^{\frac{1}{p}} |C_{m,\mathbf{X}}|^{\frac{1}{p}}} \right], \quad (6)$$

$$1 \le i \le S_m \text{ and } 1 \le m \le M$$

where $b_{m,i}$ is the number of bits allocated to the *i*th subvector for the *m*th region specific TrSVQ; for the *m*th region, $C_{m,\mathbf{Z}_i}\mathbf{z}_i$ is the covariance matrix [15] of the *i*th subvector in transform domain.

2.2. Complexity of SSTrSVQ

Let us consider L = 1; for this case, the computational steps associated with SSTrSVQ method are: switching to the best Voronoi region using SED measure, KLT transformation, SVQ of transform domain sub-vectors and inverse KLT for reconstruction. The average computational complexity/vector is: $(3p + 1)2^{b_c} + 2p^2 + \left[\sum_{m=1}^{M} \alpha_m \left[\sum_{i=1}^{S_m} (3q_{m,i} + 1)2^{b_{m,i}}\right]\right] + 2p^2$ flops⁴. For L > 1, the order of computational complexity is nearly L times of the above mentioned complexity. The required memory to store the switching codebook, KLT matrices and TrSVQ codebooks is: $p2^{b_c} + p^22^{b_c} + \sum_{m=1}^{M} \sum_{i=1}^{S_{t,m}} q_{m,i}2^{b_{m,i}}$ floats⁵.

3. QUANTIZATION RESULTS

The speech data used in the experiments is from the TIMIT database. We use the specification of AMR-WB codec [16] to compute the 16th order LPCs which are then converted to LSF parameters. In the experiments, 361,046 number of LSF vectors are used for training and 87,961 LSF vectors are used for testing (distinct from training data). To measure the wideband speech LSF quantization performance, we use the established measure of spectral distortion (SD) [1], [11]. A low average SD along-with minimum number of high SD outliers is considered necessary for good spectrum quantization performance.

The SSTrSVQ is implemented using M = 8 clusters (or switching regions) like the optimum SSVQ of [14]. For all the clusters, TrSVQs are designed for $S_m = 6$, $1 \le m \le 8$. The KLTs are so ordered that the eigen values are in descending order; accordingly, the 16-th dimensional transform vector is split into 6 parts of (2,2,2,3,3,4) dimensional sub-vectors such that it results in lower complexity. The inter and intra-cluster optimum bit allocations are carried out using, respectively, Eqn. 5 and Eqn. 6. We experimentally find the optimum value of L such that the SSTrSVQ method provides a reasonable trade-off between R/D performance and complexity. Fig. 2 shows the average SD (in dB) performance of the SSTrSVQ method for different L (at 42 and 44 bits/vector). It is observed that the performance becomes better as L increases and then saturates quickly. We choose L = 2 so that the SSTrSVQ method



Fig. 2. Average SD (in dB) performance of SSTrSVQ method for different L to choose the optimum L. (a) Performance at 42 bits/vector. (b) Performance at 44 bits/vector.

 Table 1. Performance of new soft decision based switched transform domain split VQ (SSTrSVQ) method

bits/vector (average)	Avg. SD (dB)	SD O 2-4 dB (in %)	utliers >4 dB (in %)	kflops/ vector (CPU)	kfloats/ vector (ROM)
42	1.110	1.47	0.017	15.20	16.44
43	1.068	1.15	0.018	16.00	17.31
44	1.026	0.88	0.009	18.08	19.84
45	0.988	0.69	0.006	19.82	21.79

has low complexity even though L = 3 would have given slightly better performance. Table 1 shows the performance of SSTrSVQ method using L = 2.

We compare the performance of SSTrSVQ method over the traditional SVQ, and recently proposed GMVQ [5], optimum SSVQ [14] and normalized two stage SVQ (NT_SSVQ) [13] methods. In the case of SVQ [4], the 16-th dimensional LSF vector is split into 5 parts of (3,3,3,3,4) dimensional sub-vectors⁶; the performance of SVQ is shown in Table 2. Comparing Table 1 and Table 2, it can be observed that the SSTrSVQ provides better R/D performance than SVQ, even at lower computational complexity, but with the requirement of higher memory. The SSTrSVQ method saves 3 bits/vector compared to the SVQ method. Table 3 shows the performance of GMVQ method using 8 Gaussian mixtures; note that the GMVQ functions with rate-independent complexity. Comparing Table 1 and Table 3, it is observed that the SSTrSVO provides 2 bits/vector advantage over GMVQ, but at the requirement of higher complexity. The performance of five part optimum SSVQ, with 8 switching directions, is shown in Table 4 where the bit allocations to the split subvectors and switching directions are carried out according to [14]. Comparing Table 1 and Table 4, we observe that the SSTrSVQ method provides competitive performance to the optimum SSVQ method, even at much lower computational complexity and memory. Table 5 shows the performance of NT_SSVO method where the bit allocation is carried out according to [13]. We note that the SSTrSVQ method provides better R/D performance than the NT_SSVQ method and saves nearly 1 bit/vector, but at the requirement of higher memory. Thus, considering the trade-off between R/D performance and complexity, the new SSTrSVQ method can be chosen as the potential solution for LSF quantization in wide-band speech coding.

⁴In the current literature [5], [9], [12], [11], it is a standard practice to assume that each operation like addition, subtraction, multiplication, division and comparison needs one floating point operation (flop). With this assumption, the codebook search complexity of a VQ with *h*-dimensional vector and *B*-bit allocation is $(3h + 1) \times 2^B$ flops.

⁵The "float" represents the required memory to store a real value.

⁶Five part SVQ is also implemented in [8], [11] to compare with five part SSVQ and several other VQ methods.

bits/vector	Avg.	SD Outliers		kflops/	kfloats/
	SD	2-4 dB	>4 dB	vector	vector
	(dB)	(in %)	(in %)	(CPU)	(ROM)
42	1.258	2.63	0.000	24.32	5.63
43	1.214	2.02	0.000	27.64	6.40
44	1.182	1.83	0.000	30.97	7.16
45	1.116	0.97	0.000	35.32	8.19
46	1 074	0.74	0.000	41 98	972

Table 2. Performance of split VQ (SVQ) method

Table 3. Performance of GMVQ method

bits/vector (average)	Avg. SD (dB)	SD O 2-4 dB (in %)	utliers >4 dB (in %)	kflops/ vector (CPU)	kfloats/ vector (ROM)
42	1.188	2.64	0.026	9.02	2.30
43	1.147	2.23	0.028	9.02	2.30
44	1.093	1.68	0.013	9.02	2.30
45	1.055	1.38	0.014	9.02	2.30
46	1.009	1.02	0.005	9.02	2.30

4. CONCLUSIONS

We develop an AbS based switched quantization method which is a deviation from the commonly used hard decision based switching. It is well-known that the hard decision based switching leads to the constraint of choosing the optimum code-vector. The use of soft switching circumvents the problem of hard switching and thus, leads to better R/D performance. For the SSTrSVQ method, the increased complexity due to soft switching is kept under check through the use of NN optimum TrSVQs and splitting the transform domain vector into higher number of subvectors. In a GMM based parametric framework, the optimum SSTrSVQ is shown to provide a better trade-off between R/D performance and complexity.

5. REFERENCES

- K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 1, pp. 3-14, Jan 1993.
- [2] W.R. Gardner and B.D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech and Audio Proc.*, vol. 3, No.5, pp. 367-381, Sept 1995.
- [3] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech and Audio Proc.*, vol. 8, No.4, pp. 385-401, July 2000.
- [4] G. Biundo, S. Grassi, M. Ansorge, F. Pellandini and P.A. Farine, "Design techniques for spectral quantization in wideband speech coding," *Proc. 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pp. 114-119, Budapest, Oct. 2002.
- [5] A.D. Subramaniam and B.D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech, Audio Processing*, vol. 11, No.2, pp. 130-142, March 2003.
- [6] F. Norden and T. Eriksson, "On split quantization of LSF parameters," *Proc. ICASSP*, vol.1, pp. I-157-160, May 2004.
- [7] S. Kang, Y. Shin and T.R. Fischer, "Low-complexity predictive trellis coded quantization of speech line spectral frequencies,"

Table 4. Performance of optimum SSVQ method

bits/vector	Avg.	SD Outliers		kflops/	kfloats/
(average)	SD	2-4 dB	>4 dB	vector	vector
	(dB)	(in %)	(in %)	(CPU)	(ROM)
42	1.087	0.97	0.004	29.83	53.92
43	1.044	0.72	0.004	33.44	60.60
44	1.004	0.57	0.002	44.42	80.51
45	0.968	0.43	0.002	49.31	89.08
46	0.929	0.33	0.002	55.43	100.99

Table 5. Performance of NT_SSVQ method

bits/vector	Avg.	SD Outliers		kflops/	kfloats/		
	SD	2-4 dB > 4 dB		vector	vector		
	(dB)	(in %)	(in %)	(CPU)	(ROM)		
42	1.157	2.38	0.001	13.82	3.84		
43	1.100	1.73	0.004	15.39	4.86		
44	1.063	1.34	0.003	17.05	5.24		
45	1.032	1.17	0.003	18.72	5.63		
46	0.972	0.62	0.002	20.89	6.14		

IEEE Trans. Signal Processing, vol. 52, No. 7, pp. 2070-2079, July 2004.

- [8] S. So and K.K. Paliwal, "Switched split vector quantisation of line spectral frequencies for wide-band speech coding," *Proc. EUROSPEECH (INTERSPEECH-2005)*, Lisbon, Portugal, pp. 2705-2708, Sept. 2005.
- [9] A.D. Subramaniam, W.R. Gardner and B.D. Rao, "Lowcomplexity source coding using Gaussian mixture models, lattice vector quantization, and recursive coding with application to speech spectrum quantization," *IEEE Trans. Audio, Speech, Language Processing*, vol. 14, Issue 2, pp. 524-532, March 2006.
- [10] S. Chatterjee and T.V. Sreenivas, "Two stage transform vector quantization of LSFs for wideband speech coding," *Proc. ICSLP* (*INTERSPEECH*), Pittsburgh, USA, pp. 233-236, Sept. 2006.
- [11] S. So and K.K. Paliwal, "A comparative study of LPC parameter representations and quantisation schemes for wide-band speech coding," *Digital Signal Processing*, Vol. 17, Issue 1, pp. 114-137, Jan 2007.
- [12] S. So and K.K. Paliwal, "Efficient product code vector quantisation using the switched split vector quantiser," *Digital Signal Processing*, Vol. 17, Issue 1, pp. 138-171, Jan 2007.
- [13] S. Chatterjee and T.V. Sreenivas, "Normalized two stage SVQ for minimum complexity wide-band LSF quantization," *Proc. EUROSPEECH (INTERSPEECH)*, Antwerp, Belgium, pp. 1657-1660, Aug. 2007.
- [14] S. Chatterjee and T.V. Sreenivas, "Optimum switched split vector quantization of LSF parameters," *Signal Processing*, vol. 88, Issue 6, pp. 1528-1538, June 2008.
- [15] S. Chatterjee and T.V. Sreenivas, "Optimum transform domain split VQ," *IEEE Signal Processing Letters*, vol. 15, pp. 285-288, 2008.
- [16] "AMR wide-band speech codec, transcoding functions (Release 5):" 3GPP TS 26.190 V 5.1.0