

IMPROVED QUANTIZATION STRUCTURES USING GENERALIZED HMM MODELLING WITH APPLICATION TO WIDEBAND SPEECH CODING*

Ethan R. Duni, Anand D. Subramaniam and Bhaskar D. Rao

Department of Electrical and Computer Engineering,
University of California, San Diego, CA, 92093-0407.
E-mail: {eduni, brao}@ucsd.edu, anands@qualcomm.com

ABSTRACT

In this paper, a low-complexity, high-quality recursive vector quantizer based on a Generalized Hidden Markov Model of the source is presented. Capitalizing on recent developments in vector quantization based on Gaussian Mixture Models, we extend previous work on HMM-based quantizers to the case of continuous vector-valued sources, and also formulate a generalization of the standard HMM. This leads us to a family of parametric source models with very flexible modelling capabilities, with which are associated low-complexity recursive quantization structures. The performance of these schemes is demonstrated for the problem of wideband speech spectrum quantization, and shown to compare favorably to existing state-of-the-art schemes.

1. INTRODUCTION

High quality vector quantization of high-dimensional sources, such as wideband speech, poses a number of challenges. The large codebook sizes needed to achieve high quality on these sources (on the order of 2^{40} codepoints) make general approaches such as full-search VQ impractical. Additionally, since sources such as speech signals display considerable memory, recursive quantization becomes attractive. In an ideal recursive coder, one would update the codebook at every time step to reflect the conditional density based on all previous data, imposing further constraints on the complexity of the codebook design and quantization processes.

To cope with these issues, model-based quantization techniques are attractive. These schemes work by building a parametric model of the source and then employing a closed-form quantization structure based on the model. Specifically, in [3], [4] and [6], Gaussian Mixture Models have been utilized in such a scheme and shown to perform well, as compared to conventional techniques such as MSVQ and split VQ. Notably, the complexity of GMM-based VQ schemes is rate-independent, and is low enough to permit codebook

updates at each time step. Here, we aim to increase the quality of these schemes through improved modelling techniques, while incurring minimal increase in complexity.

In Section 2, we extend the model-based quantizer structure to the case of a Hidden Markov Model of the source. The Hidden Markov Model is appealing in the case of recursive coding because it provides a simple, recursive formulation for the conditional density $f(x_n|x_{n-1}, \dots, x_0)$. Thus, we may exploit this model to design recursive coders that utilize all past data, provided a codebook for the conditional density can be either built on the fly or designed ahead of time and stored. Ott utilized this idea in [1] for the case that the data takes on discrete values, allowing a Huffman code to be built on the fly at each time step. Goblirsch presented a similar scheme in [2] for scalar-valued sources with arbitrary distributions for each state. In Goblirsch's scheme, a MAP estimate of the state is used, resulting in a switched-quantizer approach that allows the codebooks to be designed ahead of time. Here, we will utilize the quantization structure presented in [3] to extend Ott's scheme to the case of continuous vector-valued sources.

In Section 3, to better model the source, we develop a generalization of the Hidden Markov Model that is motivated by the joint-GMM source model presented in [4]. This new model generalizes both the usual HMM and the joint-GMM. We then show how to modify the HMM-based quantizer to use this Generalized HMM model. Finally, in Section 4, we evaluate the performance of these modelling/quantization techniques on a database of wideband speech. The new HMM-based quantizers are seen to outperform the corresponding GMM-based quantizers.

2. HMM-BASED RECURSIVE QUANTIZER

Hidden Markov Models are well known, and details can be found in [5]. Here, we briefly describe them to introduce notation relevant to the quantization problem. Let s_n be a Markov chain taking values in $\{1, \dots, M\}$ and denote the state transition matrix for s_n by $A \in R^{M \times M}$, with individual elements denoted by $a_{ij} = P(s_n = j | s_{n-1} = i)$. Let

*THIS RESEARCH WAS SUPPORTED BY MICRO GRANTS 02-062 AND 03-073 SPONSORED BY QUALCOMM INC.

the density associated with state i be denoted by $f_i(x) = N(x|\mu_i, \Sigma_i)$, where $x \in R^d$. It should be noted that the choice of a Gaussian Mixture Model for each state's density would also work, with minor modifications, but here we focus on the single Gaussian case for notational simplicity. Denote by λ the set of parameters (A, μ, Σ, π) , where π is the initial state distribution. We will assume that the Markov chain s_n is irreducible and so a stationary distribution $\hat{p}$ exists. Further, we will assume that $\pi = \hat{p}$, so that our model defines a stationary process.

Denote by α_n the *a priori* state distribution at time n :

$$\alpha_n(i) \equiv P(s_n = i | x_{n-1}, x_{n-2}, \dots, x_0) \quad (1)$$

Let β_n denote the *a posteriori* state distribution:

$$\beta_n(i) \equiv P(s_n = i | x_n, x_{n-1}, \dots, x_0) \quad (2)$$

Given β_{n-1} , α_n may be obtained as follows:

$$\alpha_n(i) = \sum_{j=1}^M a_{ji} \beta_{n-1}(j) \quad (3)$$

Similarly, given α_n and x_n , β_n is obtained as follows:

$$\beta_n(i) = \frac{\alpha_n(i) N(x_n | \mu_i, \Sigma_i)}{\sum_{j=1}^M \alpha_n(j) N(x_n | \mu_j, \Sigma_j)} \quad (4)$$

To initialize the recursion, α_0 is set to the initial state distribution. Finally, the density of x_n conditioned on all of the previous data is given by:

$$\begin{aligned} f(x_n | x_{n-1}, \dots, x_0) &= \sum_{i=1}^M p(x_n, s_n = i | x_{n-1}, \dots, x_n) \\ &= \sum_{i=1}^M \alpha_n(i) N(x_n | \mu_i, \Sigma_i) \end{aligned} \quad (5)$$

Thus, the density of the current data x_n , conditioned on all of the previous data, is an order- M Gaussian Mixture Model with mixture weights given by the state priors. Note that only the weights of the mixture components change with time, while the component means and covariances are fixed. In this sense, the HMM generalizes the GMM from a sequence of i.i.d. observations to a model with memory. In [3], a vector quantization structure based on Gaussian Mixture Models is presented. It is proposed to construct a recursive quantizer based on an HMM by using this GMM-based quantizer with mixture weights updated at each time step. As discussed in [4], the complexity of the GMM-based quantizer is low enough to permit updating the parameters in this way. In order to maintain synchronization between the encoder and decoder without sending side information, it is required that the update to the mixture weights depend

only on past data. Examining the recursions for α_n and β_{n-1} , we see that this is indeed the case, allowing us to base the current model update on the previously quantized vectors. The proposed recursive quantizer is seen in Fig. 1.

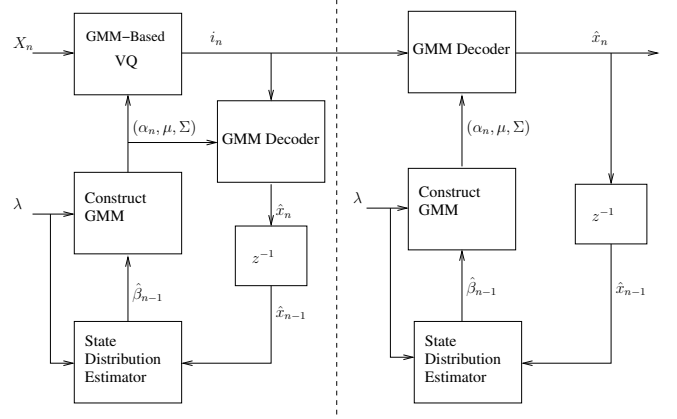


Fig. 1: HMM-Based Recursive Vector Quantizer

The box labelled "State Distribution Estimator" implements Equation (4), using $\hat{x}_{n-1}$, to estimate the *a posteriori* state distribution at time $n-1$. The box labelled "Construct GMM" then uses this distribution to produce the parameters of the Gaussian Mixture Model for the conditional density $f(x_n | \hat{x}_{n-1}, \dots, \hat{x}_0)$. In the basic case, this consists simply of finding the *a priori* state distribution at time n , α_n , using Equation (3). The means and covariances are fixed in the basic case. These parameters are then given to the "GMM-Based VQ" block, which operates as described in [3]. Note that the encoder side also includes a decoder to produce $\hat{x}_n$, in order to maintain synchronization with the decoder.

3. GENERALIZED HIDDEN MARKOV MODEL

In [4], a recursive vector quantizer was presented based on a jointly-Gaussian Mixture model of the source. This joint-GMM source model assumes that the source has a Markovity of 1 and provides a very detailed estimate of the dependence of the current sample on the previous one. However, it has the disadvantage that it ignores any dependence on samples farther back in time. Conversely, the usual HMM provides a simple model of the dependence on all previous data, but is not as flexible in describing the dependence on the previous sample, which may be significant. It is proposed to generalize the joint-GMM model by adding on a hidden Markov structure, much in the same way that the usual HMM generalizes the usual GMM. The idea here is that the joint-GMM structure can exploit the strong dependence on the previous sample, while the Markov structure will model longer-term dependency. Generally, there is a family of "hidden state" models with the property that the conditional density of the

current data given all previous data, $f(x_n|x_{n-1}, \dots, x_0)$, is a GMM. Such a model would have a hidden state sequence s_n that specifies the joint density of the current data x_n and the previous $D-1$ data points. If the data density is chosen to be Gaussian, this family includes the usual GMM (s_n is i.i.d. and $D = 1$), the joint-GMM (s_n is i.i.d. and $D > 1$) and the usual HMM (s_n is a Markov chain and $D = 1$). In [4], the joint-GMM with $D = 2$ was demonstrated. Here, we will use the model where s_n is a Markov chain and $D = 2$, and modify the HMM-based quantizer to use it.

Denote by x_n^D the $d * D$ -dimensional column vector formed by stacking $x_n, \dots, x_{n-D+1}$. In this case, the density associated with state i is: $f(x_n, \dots, x_{n-D+1}|s_n = i) = N(x_n^D|\mu_i, \Sigma_i)$. Here, the parameters μ and Σ are of dimensions $d * D$ and $(d * D) * (d * D)$. Note that, for a stationary source model, consistency of the marginal densities dictates a certain structure for the model parameters. This implies that the number of free parameters in this joint density model is not much greater than in the single-variable case. The only new free parameters in this case are the cross-covariance matrices. However, depending on the data dimension and model order, this can still amount to a significant increase in the number of free parameters.

Adopting this model changes the derivation of the conditional density only slightly. In particular, the definitions of α_n and β_n remain the same, and the update formula for α_n is unchanged. Let $N_c(x_n|\mu_i, \Sigma_i, x_{n-1}, \dots, x_{n-D+1})$ denote the conditional density of x_n given that $s_n = i$ and conditioned on the previous $D - 1$ pieces of data. As is well known, this density is also a normal whose mean depends on the previous data. The conditional covariance matrix is a constant. The update formula for β_n then becomes:

$$\begin{aligned} \beta_n(i) &= P(s_n = i|x_n, x_{n-1}, \dots, x_0) \\ &= \frac{p(s_n = i, x_n|x_{n-1}, \dots, x_0)}{f(x_n|x_{n-1}, \dots, x_0)} \\ &= \frac{\alpha_n(i)N_c(x_n|\mu_i, \Sigma_i, x_{n-1}, \dots, x_{n-D+1})}{\sum_{j=1}^M \alpha_n(j)N_c(x_n|\mu_j, \Sigma_j, x_{n-1}, \dots, x_{n-D+1})} \end{aligned} \quad (6)$$

This gives the the conditional density of the current data:

$$\begin{aligned} f(x_n|x_{n-1}, \dots, x_0) &= \\ \sum_{i=1}^M \alpha_n(i)N_c(x_n|\mu_i, \Sigma_i, x_{n-1}, \dots, x_{n-D+1}) \end{aligned} \quad (7)$$

Thus, the conditional density is again an order- M Gaussian Mixture density. Note that in this case, the mixture weights are adjusted by the HMM structure, as before, and that the component means are adjusted at each step by the jointly-Gaussian structure. Thus, use of this model requires some small changes to the quantizer structure. In particular, D previous data points must be retained for calculation of β_{n-1} . Additionally, the process of constructing the GMM

for use in the VQ now requires updating the means as well as the mixture weights.

3.1. Some Notes on Complexity

As described in [3], the simple GMM-based VQ has a low, rate-independent complexity. All of the recursive schemes discussed here will share this baseline complexity, and the added encoder-side complexity of operating a decoder for synchronization purposes. Thus, the differences in complexity arise from differences in how the update of the conditional density parameters is performed. In the simple HMM framework, the update consists solely of updating the mixture weights using equations (3) and (4). Equation (3) is a simple matrix-vector multiplication, while equation (4) requires evaluating M multidimensional Normal densities and normalizing. Utilizing the Generalized HMM structure adds to this the additional complexity of updating the conditional means at each time step. This requires M matrix-vector multiplications and $2M$ vector additions, as described in [4]. Note in particular that none of the schemes discussed here requires changing the covariance matrices, meaning that the matrix inverses and decompositions needed in the Gaussian density evaluations and in the quantization structure itself can be computed ahead of time, and do not contribute to run-time complexity.

4. PERFORMANCE ON SPEECH LSF QUANTIZATION

To demonstrate the utility of the above HMM models in a vector quantization context, several models were trained and tested on speech LSF data. The database used consisted of a training database of 350,000 16-dimensional LSF vectors derived from multispeaker, wideband speech data and an independent testing database of 15,000 LSF vectors. Four models were trained: a plain GMM, a joint-GMM, a plain HMM and a Generalized HMM. The plain GMM and plain HMM were trained with the usual EM and Baum-Welsh algorithms, respectively. The recursive GMM and Generalized HMM were also trained using EM and Baum-Welsh, but with the modification that successive vectors from the training set were "stacked" into vectors of size 32 prior to training. The orders of the models were selected so that they would all have an approximately equal number of free parameters. The resulting models were then used to quantize the test set, using the Log Spectral Distortion measure, using the techniques described in [4] and above. The resulting average distortion, as a function of bit rate, is shown in Figure 1. It can be seen that the plain HMM coder outperforms the plain GMM coder, and that the Generalized HMM coder outperforms the recursive GMM coder. The margin of improvement is around 1 bit in each case.

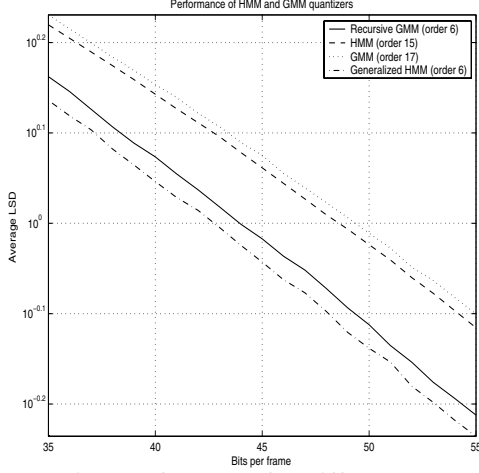


Fig. 1. Quantizer Performance for Different Source Models

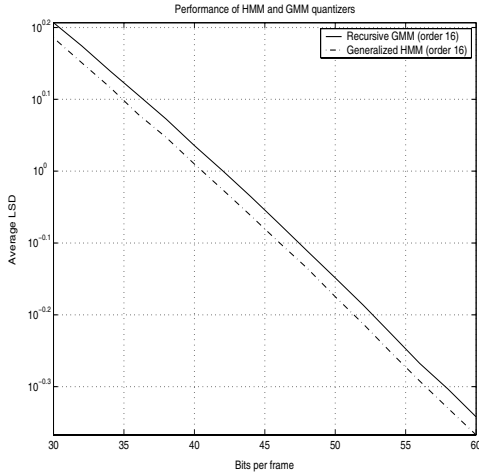


Fig. 2. Quantizer Performance for Recursive Source Models

This experiment was repeated for the two best coders, each with model order of 16. The results appear in Figure 2. Here, we see that the Generalized HMM coder can attain an average LSD of 1dB at a bit rate of 40.5 bits per frame, while the recursive GMM coder requires 42 bits per frame.

A histogram of the error for each coder at a rate of 40 bits per frame appears in Figure 3. The mean SD for the Generalized HMM coder was 1.03dB, while the mean SD for the recursive GMM coder was 1.09dB. As an outlier statistic, the percentage of quantized vectors with a SD greater than 2dB was used. The Generalized HMM coder produced 1.13% outliers, while the recursive GMM coder produced 1.80% outliers. Thus, we surmise that the Generalized HMM coder does a better job, both in terms of average distortion and spread of the distortion distribution.

5. CONCLUSIONS

In this paper, we have presented a general improved modelling scheme and shown its utility in the problem of vector quantization. In particular, we have shown how to ex-

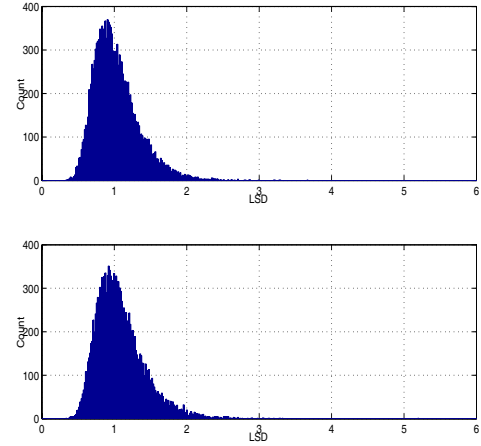


Fig. 3. Error Histograms for Generalized HMM coder (top) and Recursive GMM coder (bottom) at 40 bits/frame

tend GMM-based vector quantizers to the case of a Hidden Markov model of the source. This results in a high quality recursive vector quantizer with modest complexity. Also, we have presented a Generalized Hidden Markov model with improved modelling capabilities and shown how the quantizer can be further modified to utilize this model of the source. This recursive quantizer allows high quality encoding of continuous vector-valued sources with low complexity by efficiently exploiting memory in the source. The improved performance of these new schemes was illustrated for the case of wideband speech spectrum quantization. They were seen to outperform the previous GMM-based schemes, which have been shown in [3] to compare favorably with MSVQ and split-VQ schemes. Further issues to be investigated in this area include the mitigation of channel errors, the performance of models with deeper Markovity and the online adaptation of the model parameters.

6. REFERENCES

- [1] G. Ott, "Compact Encoding of Stationary Markov Sources", IEEE Trans. on Info. Theory, vol. IT-13, (no.1), Jan. 1967.
- [2] D. M. Goblirsch and N. Farvardin, "Switched Scalar Quantizers for Hidden Markov Sources", IEEE Trans. on Info. Theory, vol. 38, (no.5), IEEE, Sept. 1992.
- [3] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies", IEEE Trans. on Speech and Audio Proc., Vol 11, March 2003.
- [4] A. D. Subramaniam "Gaussian Mixture Models in Compression and Communication", Ph.D. Thesis, UCSD. <http://dsp.ucsd.edu/~anand/thesis.pdf>
- [5] L. Rabiner and B.-H. Juang, "Fundamentals of Speech Recognition", New Jersey: Prentice Hall, 1993.
- [6] P. Hedelin and J. Skoglund, "Vector Quantization Based on Gaussian Mixture Models", IEEE Trans. on Speech and Audio Proc., Vol.8, July 2000.