NOISE FEEDBACK STRUCTURE FOR NON-ORTHOGONAL TRANSFORM CODING

Shyam K.Raikar[1] *and Anamitra Makur*

Indian Institute of Science, Bangalore, India. Email:amakur@ece.iisc.ernet.in

ABSTRACT

A disadvantage of non-orthogonal transforms is the inevitable amplification of the quantization noise during synthesis. A feedback structure can be used to compensate this amplification. We try to generalize the idea of noise feedback to the case of non-orthogonal transforms. We explore the application of non-orthogonal transforms to the coding of quasi-stationary sources for which the KLT is, strictly speaking, not defined. Transforms which maximize the coding gain for the assumed model of the quasi-stationary source are found, which are generally non-orthogonal. The proposed transforms along with the feedback structure are seen to perform better than the average KLT for AR sources and real life speech signals.

1. INTRODUCTION

Transform coding belongs to a class of compression schemes that decompose a signal into several channels and quantize the decomposed signal. The decomposition is done in a way that will result in improvement in performance over that of PCM. The improvement over PCM is measured by a quantity called the *coding gain,* which is the ratio of the reconstruction error variance in PCM to that in transform coding. For a given input vector **x** and transform matrix **T**, the output vector θ is given by $\theta = \mathbf{Tx}$. Assuming the *additive noise model* for the quantizers and that the quantization noise in the different channels is mutually uncorrelated, the output, $\hat{\theta}$, of the quantizers is given by $\hat{\theta} = \theta + \mathbf{q}$, where **q** is the quantization noise vector. The output of the synthesis block is given by $\hat{\mathbf{x}} = \mathbf{T}^{-1}\hat{\theta}$. The overall reconstruction error vector is therefore given by

$$\mathbf{x} - \hat{\mathbf{x}} = \mathbf{T}^{-1}\mathbf{q}. \quad (1)$$

Using optimum bit allocation subject to bit rate constraint the coding gain can be shown to be

$$C_g = \frac{\varepsilon\sigma_x^2}{\left(\prod_{k=0}^{M-1} \varepsilon_k \sigma_k^2 \omega_k\right)^{\frac{1}{M}}} \quad (2)$$

where ε_k are the quantizer constants, σ_k^2 the signal variance of k^{th} channel, ω_k the squared norm of the k^{th} column of $\mathbf{T}^{-1}$, σ_x^2 is the total signal variance and M is the transform order. When the transform matrix is non-orthogonal the ω_k may not be unity. This causes amplification of the quantization noise during reconstruction. Scaling down the ω_k is of no consequence if we have perfect reconstruction constraint.

It was shown in [1] that for a given analysis transform the best choice for the synthesis transform is its inverse. The minimum mean squared error is achieved by a transformation that diagonalizes the covariance matrix of the input. This is called the KLT of the source. A general proof of the optimality of the KLT is given in [2]. In the context of non-orthogonal transforms a *prediction based lower triangular transform* (PLT) has been proposed in [3]. The transform is arrived at by using the *LU decomposition* and symmetry of the auto-correlation matrix. This transform being non-orthogonal causes amplification of quantization noise. To eliminate this amplification two minimum noise structures called MINLAB(I) and MINLAB(II) have been arrived at. The idea has been extended to the case of biorthogonal filter banks in [4]. A coloring filter matrix is used to shape the *psd* of the quantization noise in the different channels to compensate for the amplification taking place due to the non-orthogonal synthesis filter bank.

2. NOISE FEEDBACK STRUCTURE

The reconstruction error is dependent on $\mathbf{T}^{-1}$ as seen from (1). In order to compensate for the effect of $\mathbf{T}^{-1}$ we would require to pass the quantization noise through $\mathbf{T}$, but such a system may essentially have a delay free loop and cannot be realized. Hence the best we can do is to minimize the effect by using a premultiplying matrix **A**, which modifies the quantization noise through feedback. Assuming the channels to be quantized from top to bottom, the feedback in any channel should come from the channels already quantized and not the current channel or channels yet to be quantized. Hence **A** should be lower triangular. This ensures there are no delay free loops in the system and hence that the system is physically realizable. Fig 1 shows the proposed *noise feedback structure*. The quantization noise is fed back to the quantizers through the matrix **I-A**. With this structure the overall reconstruction error vector becomes $\mathbf{e} = \mathbf{T}^{-1}\mathbf{Aq}$. Since **T** is a non-singular matrix it can be decomposed as **T=LDQ**, where **L** is lower triangular, **D**

[1] Currently with PACE SOFT SILICON India. Email:sraikar@pace.stpp.soft.net

is diagonal and $\mathbf{Q}$ is unitary. Let $\widetilde{\mathbf{T}} = \mathbf{T}^{-1}\mathbf{A}$. Using the above decomposition $\widetilde{\mathbf{T}} = \mathbf{Q}^t \mathbf{D}^{-1}\mathbf{L}^{-1}\mathbf{A}$. As seen from (2), the optimum

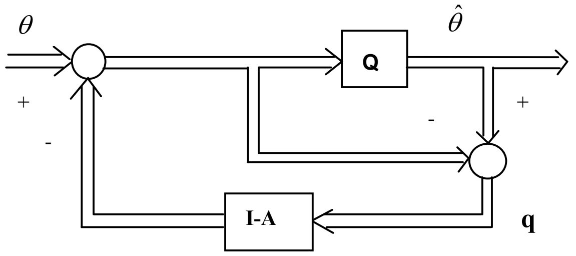


Fig 1. Noise feedback structure for non-orthogonal transforms

feedback matrix $\mathbf{A}$ can be obtained by minimizing

$$\prod_{k=0}^{M-1} \widetilde{\omega}_k = \prod_{k=0}^{M-1} (\widetilde{\mathbf{T}}^t \widetilde{\mathbf{T}})_{k,k} = \prod_{k=0}^{M-1} \left(\left((\mathbf{LD})^{-1}\mathbf{A} \right)^t \left((\mathbf{LD})^{-1}\mathbf{A} \right) \right)_{k,k} \quad (3)$$

From matrix theory it is known that for any matrix $\mathbf{P}$

$$\prod_{k=0}^{M-1} \mathbf{P}_{k,k} \geq Det(\mathbf{P}). \quad (4)$$

Equality holds when $\mathbf{P}$ is diagonal. Hence the optimal $\mathbf{A}$ should orthogonalize $(\mathbf{LD})^{-1}\mathbf{A}$. Since this is a lower triangular matrix, it will be orthogonal only if it is diagonal. Hence $(\mathbf{LD})^{-1}\mathbf{A} = D$. This implies $\mathbf{A} = \mathbf{LD}D$. Since the diagonal elements of $\mathbf{A}$ should be one, the optimal solution is $\mathbf{A}=\mathbf{L}$. The optimal $\mathbf{A}$ is unique since $\mathbf{L}$ is uniquely determined by $\mathbf{T}$. It is shown below that the feedback structure always reduces the reconstruction error variance. Equation (3) can be simplified as

$$\prod_{k=0}^{M-1} \widetilde{\omega}_k = \prod_{k=0}^{M-1} (\mathbf{D}^{-2})_{k,k} = \prod_{k=0}^{M-1} d_k^{-2}. \quad (5)$$

Without the feedback the above term is given by

$$\prod_{k=0}^{M-1} \omega_k = \prod_{k=0}^{M-1} \left(\mathbf{L}^{-1t} \mathbf{D}^{-2} \mathbf{L}^{-1} \right)_{k,k}. \quad (6)$$

Using (4) and the fact that the determinant of $\mathbf{L}$ is unity we get

$$\prod_{k=0}^{M-1} \omega_k \geq \prod_{k=0}^{M-1} d_k^{-2}. \quad (7)$$

Hence from (5) and (7) we have

$$\prod_{k=0}^{M-1} \widetilde{\omega}_k \leq \prod_{k=0}^{M-1} \omega_k \quad (8)$$

When the transform matrix is orthogonal, $\mathbf{L}$ and hence $\mathbf{A}$ is the identity matrix and feedback is inconsequential. For the case of a lower triangular transform, i.e. when $\mathbf{DQ} = \mathbf{I}$, the optimal $\mathbf{A}$ is the transform matrix itself and the feedback totally eliminates the amplification of quantization noise.

3. TRANSFORM CODING FOR QUASI-STATIONARY SOURCES

We consider the quasi-stationary model of the signal, i.e. the signal is comprised of N stationary processes. The signal at any instance is the outcome of any one of the processes chosen randomly, where each process occurs with some known probability α_i such that $\alpha_1 + \alpha_2 + \ldots + \alpha_N = 1$.

The problem is to find some transform $\mathbf{T}$ of dimension M which will optimize the actual coding gain for some bit allocation scheme and for a given fixed bit budget. The KLT in this case is not defined or defined only if the auto-correlation matrices of the constituting processes are simultaneously diagonalizable[5]. Generally, a fixed signal dependent transform which diagonalizes the long-term averaged auto-correlation matrix (which is sub-optimal) is most commonly used. Whether the average KLT is more towards an optimum for any process will depend on the auto-correlation matrices as well as the probabilities. Another solution for such non-stationary signals is a switching transform which changes according to the changing signal statistics. However the complexity of the encoder and decoder would be much more than that for a fixed transform. For the assumed quasi-stationary source model, among the class of fixed transforms, we try to find a solution better than the KLT of the averaged auto-correlation.

3.1. Bit Allocation

The actual coding gain should be the average of the coding gains of the given transform for each of the processes,

$$C_g = \sum_{i=1}^{N} \alpha_i C_{g_i} \quad (9)$$

where C_{g_i} is the coding gain of $\mathbf{T}$ for the i^{th} process. If the N processes are assumed to be zero mean and unit variance then (9) can be written as

$$C_g = \sum_{i=1}^{N} \frac{M \alpha_i \varepsilon_i 2^{-2R}}{\sum_{k=0}^{M-1} \varepsilon_{ik} 2^{-2R_{ik}} \sigma_{ik}^2 \omega_k} \quad (10)$$

where ε_i is the quantizer constant for the i^{th} process, ε_{ik} is the k^{th} channel quantizer constant for the i^{th} process, ω_k is the squared norm of the k^{th} column of either $\mathbf{T}^{-1}$ or $\mathbf{T}^{-1}\mathbf{A}$ depending on the structure used, R_{ik} is the bits allocated to the k^{th} channel and σ_{ik}^2 is the variance of the k^{th} channel for the i^{th} process.

3.1.1. Fixed bit allocation

The bit allocation is dependent on the statistics of the signals in the different channels, which are again non-stationary. Hence we need to find a bit allocation scheme that is optimal with respect to the average statistics of the signals. Let the average variance of each channel be

$$\bar{\sigma}_k^2 = \sum_{i=1}^{N} \alpha_i \sigma_{ik}^2. \quad (11)$$

The optimal bit allocation for this case is then given by

$$R_k = R + \frac{1}{2}\log_2 \frac{\omega_k \bar{\sigma}_k^2}{\left(\prod_{l=0}^{M-1} \omega_l \bar{\sigma}_l^2\right)^{\frac{1}{M}}} \qquad (12)$$

With this bit allocation scheme the cost function to be optimized i.e. the overall coding gain becomes

$$C_g = \frac{M}{\left(\prod_{l=0}^{M-1} \omega_l \bar{\sigma}_l^2\right)^{\frac{1}{M}}} \sum_{i=1}^{N} \frac{\alpha_i}{\sum_{k=0}^{M-1} \frac{\sigma_{ik}^2}{\bar{\sigma}_k^2}} \qquad (13)$$

3.1.2. Adaptive bit allocation

Adaptive bit allocation is one, which adapts to the changing signal statistics. With regard to our problem this would mean that we are using a separate optimal bit allocation for each of the *N* processes. The bit allocation in this case is given by

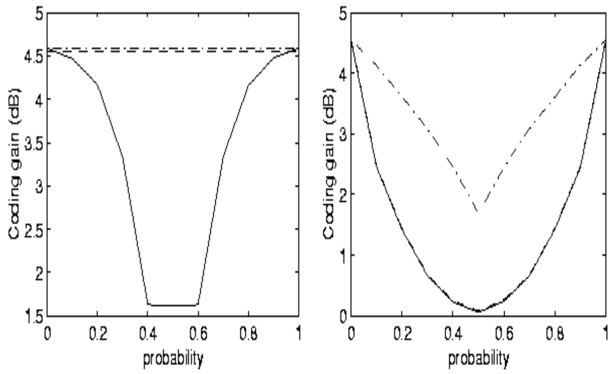
$$R_{ik} = R + \frac{1}{2}\log_2 \frac{\omega_k \sigma_{ik}^2}{\left(\prod_{l=0}^{M-1} \omega_l \sigma_{il}^2\right)^{\frac{1}{M}}}. \qquad (14)$$

The overall coding gain then becomes

$$C_g = \sum_{i=1}^{N} \frac{\alpha_i}{\left(\prod_{l=0}^{M-1} \omega_l \sigma_{il}^2\right)^{\frac{1}{M}}}. \qquad (15)$$

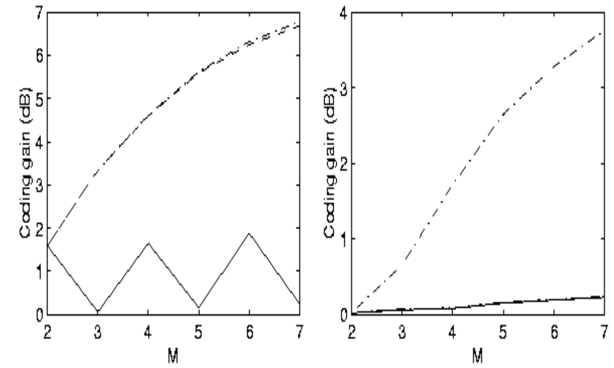
3.2. AR Source

We consider here the coding of a quasi-stationary source using the KLT, transform **T** without noise feedback and **T** with noise feedback for each of the two cases of fixed and adaptive bit allocation. The performance of the three, in terms of coding gain, is compared. The KLT is designed for the average auto-correlation matrix. The transform **T** with and without feedback is obtained by the optimization of the cost functions (13) and (15). The transform parameters enter the optimization in a complex form in the quantities σ_{ik}^2 and ω_k. We have used *Simulated Annealing* for optimization and it was found to give a better solution than other methods. The simulation results presented here are for the case of two AR(4) processes, with *psd*s that are the mirror images of each other, which form the constituting processes of the non-stationary source. Transform orders from two to seven are considered. Fig. 2(a) and 2(b) show the variation of the coding gain versus the probability α_1 in the case of adaptive and fixed bit allocation respectively. Transform of order four is used. In both figures the optimized transform with feedback performs superior to KLT. However, the optimized transform without feedback bettered the KLT only with adaptive bit allocation. In general the optimal transform is non-orthogonal and the feedback structure is seen to give up to 1.6 dB improvement over the case without feedback.



a) Adaptive bit allocation b) Fixed bit allocation

Fig 2. Coding gain Vs probability α_1. KLT(solid), Transform without feedback(dotted), transform with feedback(dash-dotted).



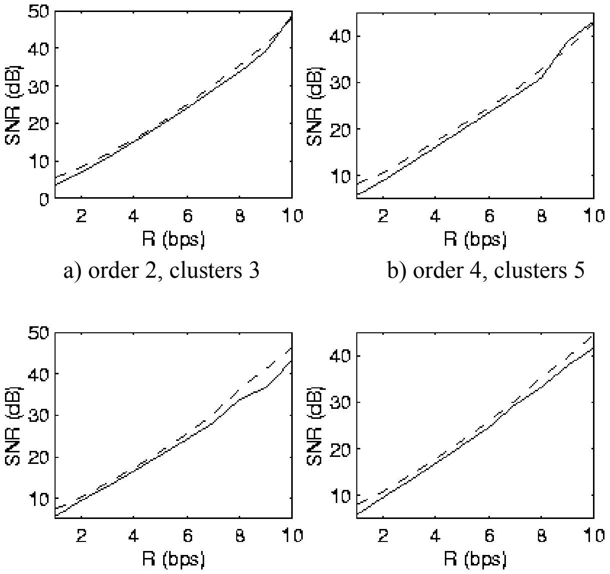
a) Adaptive bit allocation b) Fixed bit allocation

Fig 3. Coding gain Vs transform order. KLT(solid), Transform without feedback(dotted), transform with feedback(dash-dotted).

Fig. 3(a) and 3(b) show the variation of the coding gain with the transform order. Here we have considered the point α_1 =0.5 on the coding gain versus α_1 curves as the operating point and varied the transform order. In both cases we see considerable improvement in the coding gain over that of KLT.

3.3. Speech

Speech is known to be a *quasi-stationary* signal i.e. it is comprised of various small stationary segments. This fits approximately in our assumed model of a quasi-stationary source. Hence the above concept can be applied to speech. The **TIMIT** database was used in all the simulations presented in this section. All the speech signals used were 16 KHz sampled with 16 bits/sample representation. AR modeling was used to estimate a model for a general speech signal. A twelfth order AR model was used to get the spectral envelope of the phonemes. The lpc vectors so obtained were clustered using the *k-means method* using simple Euclidean distance as the distance measure. The number of processes constituting the speech signal is taken to be equal to the number of clusters into which the speech units, the *phonemes*, are segregated. The phonetic transcription available in the **TIMIT** database was used to identify and segregate the phonemes to be used for clustering. Different number of clusters were considered starting from two to twelve.



a) order 2, clusters 3 b) order 4, clusters 5

c) order 5, clusters 5 d) order 7, clusters 5

Fig 4 SNR Vs bit rate for single speaker (a,b) and multiple speakers (c,d). KLT (solid), non-orthogonal transform (dashed)

The probabilities of occurrence of the representative processes are obtained during the clustering, as the size of each cluster divided by the total size of all clusters. Uniform quantization was used in all the channels. We have also used the numerically obtained optimum values for the overload points, assuming the Laplacian *pdf* for the speech signal. The SNR versus bit rate curves are compared with those of the KLT designed for the average statistics of the signal. All simulations were carried out only for female speakers. For the case of a single speaker eight spoken sentences of the speaker were taken as the training data and two as test data. The performance in all the cases is better than the KLT. We have next considered the across speaker variability. Six spoken sentences from each of the five different speakers were used as training data and four sentences from each speaker were used as the test data. In this case it was found that the best performance is obtained for the transform designed with five clusters. Better performance is obtained for lower transform orders as seen from table 1 because the optimization converges to the optimum solution when the number of free variables is less. The proposed transform consistently performs better than the KLT at all rates. An average improvement in SNR of about 1.5 dB over the KLT at all bit rates from 1 to 10 bits/sample is observed in the best case and a peak improvement of around 4.5 dB at higher bit rates. Figure 4 shows the performance comparison for the test data.

TABLE I
Average Improvement in SNR

	Transform order M	2	3	4	5	6	7
Single Speaker	SNR (dB) improvement	1.1	1.5	1.3	0.4	0.1	0.4
	Number of clusters N	2	3	5	5	6	5
Multiple Speakers	SNR (dB) improvement	1.2	1.3	1.0	1.0	0.3	0.9
	Number of clusters N	5	5	5	5	5	5

4. CONCLUSION

To minimize the quantization noise amplification occurring due to non-orthogonal transforms, a noise feedback structure has been found. The feedback matrix is obtainable from the QR factorization of the transform matrix and is unique for a given invertible transform matrix. It always reduces the reconstruction error variance for a non-orthogonal transform. Next we have considered the coding of *quasi-stationary* signals with an assumed model. Transforms have been found, by numerical optimization, which outperform the KLT designed for the average signal statistics. The two cases of fixed and adaptive bit allocation have been considered. Results on AR signals and speech show that in both cases the transform gives a higher coding gain than the average KLT. The transforms obtained were highly non-orthogonal, and the feedback structure gave a further increase in the coding gain.

5. REFERENCES

[1] J.J.Y.Huang and P.M.Schultheiss, ‘‘Block Quantization of Correlated Gaussian Random Variables,’’ *IEEE Transactions on Communication Systems*, vol.11, pp. 289-296, Sep 1963.

[2] V.Goyal, ‘‘High-Rate Transform Coding: How High is High, and Does it Matter?,’’ *Proc. IEEE Int. Symp. on Information Theory* (Sorrento, Italy), June 2000.

[3] S.M.Phoong and Y.P.Lin, ‘‘Prediction-Based Lower Triangular Transform,’’ *IEEE Transactions on Signal Processing* vol.48, No.7, pp. 1947-1955, July 2000.

[4] Arunkumar M and A. Makur, ‘‘Optimal Biorthogonal Filter Banks with Minimization of Quantization Noise Amplification,’’ *Proc. of IEEE workshop on Statistical Signal Processing* (Singapore), Aug 2001.

[5] S.Raikar, ‘‘Transform Coding using Non-Orthogonal Transforms,’’ M.E. Thesis, Indian Institute of Science [2002].