A Low Complexity Audio Coding Scheme for Wideband Audio*

Ram M. Peddibhotla, Robert C. Maher, and Khalid Sayood
Department of Electrical Engineering
University of Nebraska
Lincoln, NE 68588-0511
Tel: (402) 472-3771
Fax: (402) 472-4732

Abstract

Differential encoding is a well known low complexity coding technique. Its use in the coding of wideband audio is limited by its inability to follow rapid changes in the signal. This is a serious drawback when coding high fidelity audio where this inability can seriously degrade the perceptual quality of the reconstruction. This overload problem can be remedied by using a recursively indexed quantizer. In this paper we present some empirical results for the differential coding of audio signals.

1 Introduction

With the increased popularity of multimedia, and the increasing computational power of personal computers, schemes for compressing wideband audio signals have been attracting increasing attention. While several highly efficient audio compression techniques have been developed in recent years [1], these are for the most part relatively complex schemes. In this paper, we present a low complexity scheme for coding wideband audio signals. The proposed scheme uses differential encoding to remove the redundancy from the signal, and a recursively indexed quantizer to encode the residual signal. We describe the basic system in the following.

1.1 Differential Pulse Code Modulation

Differential Pulse Code Modulation (DPCM) is a popular speech coding technique. The DPCM system consists of two main blocks, the predictor and the quantizer.

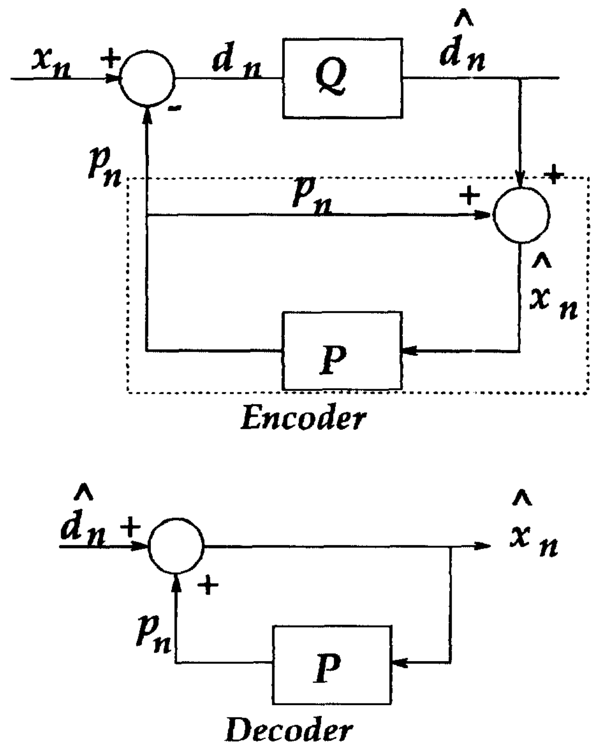

Fig. 1: The DPCM system

For a given input sequence $\{x_n\}$, the predictor generates the prediction sequence $\{p_n\}$ using the past reconstructed values $\{\hat{x}_n\}$

$$p_n = f(\hat{x}_{n-1}, \hat{x}_{n-2}, \cdots, \hat{x}_0) \qquad (1)$$

The difference between the input sequence and the predicted sequence d_n is quantized and transmitted to the receiver. The quantizer can be viewed as a partitioning of the input space into intervals, with each interval being represented by a binary codeword. This binary codeword is translated into a representation value at the receiver. The difference between the in-

*This work was supported in part by the NASA Goddard Space Flight Center under Grant NAG 5-1612.

1058-6393/95 $4.00 © 1995 IEEE

put and the representation value is called the quantization noise. In order to cover the entire input space, the outer intervals are, at least in theory semi-infinite. This results in the division of the quantization noise into two types. The quantization noise resulting from an input falling in the inner (bounded) levels is called the granular noise, while the quantization error resulting from inputs falling in the outer (unbounded) intervals is called the overload noise.

Representing the quantizer as a source of quantization noise the differencing and quantization operations can be represented as follows

$$d_n = x_n - p_n \quad (2)$$
$$\hat{d}_n = d_n + q_n \quad (3)$$

At the receiver this quantized difference is added to the prediction value generated by the receiver. If there is no error introduced in the transmission channel the transmitter and receiver predictor are operating on the same input and therefore produce the same prediction value. Therefore the reconstructed value is given by

$$\begin{aligned} \hat{x}_n &= p_n + \hat{d}_n & (4)\\ &= p_n + d_n + q_n & (5)\\ &= p_n + x_n - p_n + q_n & (6)\\ &= x_n + q_n & (7) \end{aligned}$$

Thus the only distortion contained in the reconstruction is the quantization noise. However, the quantization noise depends on the magnitude of the difference d_n which in turn depends on the accuracy of the prediction p_n. If the prediction is far from the input d_n will be large. A large value of d_n would fall in the outer intervals of the quantizer, resulting in a large overload error. This error would then become part of the reconstruction signal which would result in a decrease in the accuracy of the next prediction. Such a situation would generally occur when the input signal would be changing too rapidly for the predictor to keep track. As this is not an unusual situation for many wideband audio signals, differential encoding can result in objectionable distortion in the reconstructed signal. The recursively indexed quantizer (RIQ) was originally developed to deal with similar problems in image compression [3]. We have since shown that the use of the RIQ in a differential encoding system results in optimum performance for synthetic sources [4] and that an adaptive version of the RIQ also performs well in differential encoding of audio [2].

1.2 The Recursively Indexed Quantizer

In [5] a Recursively Indexed quantizer (RIQ) was presented. The RIQ algorithm is briefly described as follows.

For a given quantizer stepsize Δ and a positive integer K, define x_l and x_h as follows:

$$x_l = -\lfloor \frac{K-1}{2} \rfloor \Delta$$

$$x_h = x_l + (K-1)\Delta$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x. A recursively indexed quantizer of size K is a uniform quantizer with step size Δ (the uniform spacing both between the thresholds and between the output levels) and with x_l and x_h being its smallest and largest output levels (Q defined this way always has 0 as an output level). The quantization rule Q is given as follows:

For a given input value x if x falls in the interval $(x_l + (\Delta/2), x_h - (\Delta/2))$, then $Q(x)$ is the nearest output level. If x is greater than $x_h - (\Delta/2)$, see if $x_1 \triangleq x - x_h \in (x_l + (\Delta/2), x_h - (\Delta/2))$.
If so, $Q(x) = (x_h, Q(x_1))$.
If not, form $x_2 = x - 2x_h$ and do the same as for x_1.

This process continues until for some $m, x_m = x - mx_h$ falls in $(x_l(\Delta/2), x_h - (\Delta/2))$, in which case x will be quantized into

$$Q(x) = (\underbrace{x_h, x_h, \cdots, x_h}, Q(x_m))$$

If x is smaller than $x_l + (\Delta/2)$, a similar procedure to this is used, i.e., $x_m = x - mx_l$ is formed so that it falls in $(x_l + (\Delta/2), x_h - (\Delta/2))$, and is quantized to $(x_l, x_l, \cdots, x_l, Q(x_m))$.

In summary, the quantizer operates in two modes: it operates in one mode when the input falls in the range $(x_l + \frac{\Delta}{2}, x_h - \frac{\Delta}{2})$, and another when the input falls outside of the specified range. The distortion per sample is always bounded by $\frac{\Delta}{2}$.

Let θ be the ratio of the number of output symbols from the RIQ for a given number of input symbols. The rate of the quantizer is given by [5]

$$R = \theta \lceil \log_2(K) \rceil \quad (8)$$

if the output of the RIQ is encoded using a fixed rate code, and

$$R = \theta H_{RQ}(K) \quad (9)$$

when the output of the RIQ is encoded using an entropy coder. $H_{RQ}(K)$ is the entropy of the representation alphabet.

2 The Proposed System

The proposed system is a DPCM system similar to the one shown in Figure 1. The only difference is that the quantizer is replaced by a recursively indexed quantizer. This structure yields granular distortion only and with the error being bounded by $\Delta/2$. Thus even if the prediction is highly inaccurate resulting in a large input for the quantizer, the output of the quantizer is at most $\Delta/2$ away from the input. This means that the reconstructed value $\hat{x_n}$ will differ from the input x_n by at most $\Delta/2$, thus preventing the error in prediction from propagating.

In the case of small Δ, and a smooth input density, we can show that

$$\sigma_q^2 \approx \frac{\Delta^2}{12} \tag{10}$$

This means that specifying the value of the stepsize Δ specifies the quantization noise power and hence the signal to noise ratio. This property allows us to "fine tune" the noise power at any time. Modification of the stepsize has been used as a method of rate control in some applications. While in these applications the modification of the stepsize also affected the distortion, the control over the distortion was not precise; modifying the stepsize changes the granular noise, but it also affects the probability of overload. In the DPCM-RIQ system, as there is no overload noise, modification of the stepsize allows us to precisly control the distortion.

The recursively indexed quantizer can be totally specified by specifying the stepsize Δ and the number of levels K. We have briefly discussed the effect of the selection of Δ, we look now at the effect of the selection of K. In [5] we had noted that if the output of the RIQ was to be entropy coded, as long as the value of K was moderately large (> 16), there was no significant effect of the actual size of K. However if the output of the RIQ was to be encoded using a fixed length code, the size of K could significantly affect the rate. For fixed rate encoding the average rate is going to be at least as large as $\log_2(M)$, therefore we would like to keep K small. This is exactly the opposite of the situation in the entropy coded case, where increasing the size of the quantizer alphabet can only decrease the average rate. However, making K small also increases the number of recursions and therefore the expansion factor θ.

In [4] we had shown that for Gauss-Markov and Laplace-Markov sources, if the output of the RIQ system is entropy coded, the DPCM-RIQ performed at or close to the optimum entropy constrained DPCM system. In this paper, we will examine the performance of the DPCM-RIQ system when it is used to encode audio sequences. We have a number of objectives. First, we would like to see if the earlier results obtained for synthetic sources hold true for real sources. Second, we would like to see the effect of the size of the alphabet on the average rate for a fixed rate coding system. Finally we would like to see if there is much advantage to be gained from entropy coding the output of the RIQ in terms of both objective and subjective criteria.

3 Simulation Results

We implemented the DPCM-RIQ system using a first order predictor and RIQs with different numbers of levels. The inputs consisted of three different audio signals, an orchestra piece (*moz*), a rock and roll piece (*cohn*), and a solo soprano (*vega*). The stepsize Δ was adjusted in order to provide two compression ratios, 4 : 1, and 5 : 1. The results are tabulated in Tables 1- 6. The rates were computed assuming fixed rate coding but without using the ceiling function in (8). Therefore, unless we were to use some form of extended code, the compression ratios have been over estimated. However, for seven level quantizer which seems to give the best or close to the best performance in all cases, this overestimation is very slight. The SNR is given by

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{\sigma_q^2} \tag{11}$$

where σ_x^2 is the variance of the input, and σ_q^2 is the variance of the reconstruction error. For the estimated values of SNR, the value of σ_q^2 was obtained by using equation (10). For the computed value of SNR, the value of σ_q^2 was computed from the simulation. Notice the extremely close agreement between the computed and estimated values.

In both the 4 : 1 and 5 : 1 case the SNR peaks at around 7 levels. This means that if we represent the output of the DPCM-RIQ system using three bits per sample we will be operating close to the optimum.

To see how much gain could be had by using entropy coding on the output, we used a twenty one level RIQ with entropy coding. The stepsize was adjusted to obtain the desired rate. The results are tabulated in Table 7.

As can be seen from Table 7 except for the 4 : 1 *Moz* input, there is about a 3 dB gain with entropy coding. Based on informal listening tests, this 3 dB gain was clearly perceptible. However, whether the improvement is sufficient to warrant the added complexity would depend on the application.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.047	37.34	37.34
5	.037	39.42	39.44
7	.035	39.91	39.91
9	.034	40.16	40.16
11	.034	40.16	40.16
13	.036	39.66	39.67
15	.050	36.81	36.81

Table 1. Performance of the 4 : 1 DPCM-RIQ system with the *Cohn* input.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.055	35.98	35.97
5	.043	38.12	38.12
7	.038	39.19	39.19
9	.035	39.91	39.90
11	.034	40.16	40.16
13	.036	39.66	39.66
15	.050	36.81	36.81

Table 2. Performance of the 4 : 1 DPCM-RIQ system with the *Moz* input.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.077	33.06	33.08
5	.063	34.80	34.81
7	.059	35.37	35.38
9	.055	35.98	35.99
11	.058	35.52	35.52
13	.063	34.80	34.81
15	.080	32.73	32.72

Table 3. Performance of the 4 : 1 DPCM-RIQ system with the *Vega* input.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.070	33.88	33.89
5	.060	35.22	35.22
7	.067	34.27	34.26
9	.150	27.27	27.28

Table 4. Performance of the 5 : 1 DPCM-RIQ system with the *Cohn* input.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.083	32.41	32.41
5	.069	34.01	34.01
7	.069	34.01	34.01
9	.110	29.96	29.96

Table 5. Performance of the 5 : 1 DPCM-RIQ system with the *Moz* input.

Levels	$\frac{\Delta}{\sigma_x}$	SNR(est.)	SNR(comp.)
3	.117	29.42	29.39
5	.105	30.36	30.27
7	.115	29.57	29.55
9	.270	22.16	22.65

Table 6. Performance of the 5 : 1 DPCM-RIQ system with the *Vega* input.

Input	Compression Ratio	Best Fixed Rate SNR	Entropy Coded SNR
Cohn	4 : 1	40.16	43.56
Cohn	5 : 1	35.22	38.33
Moz	4 : 1	40.16	41.53
Moz	5 : 1	34.01	36.64
Vega	4 : 1	35.99	39.18
Vega	5 : 1	30.27	34.02

Table 7. Comparison of entropy coded and fixed rate systems.

4 Conclusion

The results for the DPCM-RIQ system for audio signals seems to agree with the earlier results for synthetic sources. For the fixed rate coding scheme, there is a clear dependence on the number of levels. Based on these rather limited results, a seven level RIQ seems to give the best, or close to the best performance. The utility of this scheme depends on the tradeoffs required in a particular application. The system is extremely simple to implement, and can easily perform at real-time on a 486 Personal computer using an implementation in *C*. The perceptual quality of the 4 : 1 coded system, while not transparent, is quite good. If transparent reconstruction is required and computational complexity is not an issue, systems such as those described in [1] would probably be preferred. However, where complexity becomes important, this approach is certainly worth looking at.

References

[1] J.D. Johnston and K. Brandenburg. Wideband Coding - Perceptual Considerations for Speech and Music. In S. Furui, editor, *Advances in Speech Signal Processing*, chapter 4, pages 109–140. Marcel Dekker Inc, 1992.

[2] R.C. Maher. Computationally Efficient Compression of Audio Material by Means of RIQ-DPCM. In *Proc. 1993 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY, October 1993. IEEE.

[3] M.C. Rost and K. Sayood. An Edge Preserving Differential Image Coding Scheme. *IEEE Transactions on Image Processing*, 1:250–256, April 1992.

[4] K. Sayood and S. Na. Recursively Indexed Differential Pulse Code Modulation. In R. Calderbank, G.D. Forney, and N. Moayeri, editors, *Coding and Quantization*, pages 253–263. DIMACS, American Mathematical Society, October 1992.

[5] K. Sayood and S. Na. Recursively Indexed Quantization of Memoryless Sources. *IEEE Transactions on Information Theory*, IT-38:1602–1609, November 1992.