COMPUTATIONALLY EFFICIENT COMPRESSION OF AUDIO SIGNALS BY MEANS OF RIQ-DPCM

Robert C. Maher

Department of Electrical Engineering and Center for Communication and Information Science
University of Nebraska
Lincoln, NE 68588-0511
(402) 472-2081
rmaher@unl.edu

ABSTRACT

The need to transmit large amounts of data over limited bandwidth channels has resulted in many methods for digital data compression. The common approach is to identify and remove redundancy from the input data stream using knowledge of the source characteristics. In the case of signals intended for human observers (speech, music, pictures, etc.) it is also useful to consider the strengths and weaknesses of the human sensory systems in order to achieve a greater degree of data compression. Unfortunately, achieving perceptually transparent compression requires considerable computational resources. For situations requiring extremely low computational complexity without strictly transparent coding, such as multimedia applications on personal computer platforms, a new adaptive differential pulse code modulation (DPCM) data compression scheme is proposed. Although standard DPCM structures are widely used in single-talker speech coding systems, the models and statistical assumptions well-known for speech signals are not applicable to arbitrary audio signals such as music. The new DPCM formulation presented here includes a recursively indexed quantizer (RIQ) to eliminate the problem of overload distortion, a simple predictor structure to take advantage of the short-term correlation present in wideband audio signals, and an adaptation strategy to optimize the system to the local statistics of the input signal. Thus, the new RIQ-DPCM formulation is presented as a computationally efficient means of wideband audio compression.

1. INTRODUCTION

Audio signal compression is useful in many communications situations where it is necessary to transmit digitized speech and music. Most general-purpose *lossless* data compression schemes perform relatively poorly on audio signals because of the ill-defined statistical character of these signals. However, since human beings are typically the receivers of audio information it is possible to develop *lossy* compression schemes which retain the perceptual quality of the original signals by forcing the lossy distortion to occur in such a way as to be less detectable by the human ear-brain system. Thus, the incorporation of human psychoacoustical information in the design of audio data compression systems is a significant area of research.

Existing audio data compression systems specifically intended for high quality applications can achieve approximately 6:1 compression of monophonic audio material with results nearly indistinguishable from the original signal to human listeners [6, 7, 8, 9, 10]. Achieving this "transparent" compression at a data rate below 128 kbits/s requires significant computational complexity. In situations where the available computational capability is insufficient to support the goal of transparent coding, such as general purpose personal computers, one must consider the available tradeoffs between bit rate, distortion, and computation.

1.1 Efficient Lossy Coding with DPCM

The extremely well-known technique of differential pulse-code modulation (DPCM) is a computationally efficient means for waveform coding. Many common audio signals exhibit a long-term lowpass spectral characteristic which results in a significant correlation between successive samples of the input signal [1]. Since the adjacent sample difference has a smaller variance than x[n] it is clear that a more efficient coding of the audio signal can be obtained by storing or transmitting the sample difference signal rather than the signal itself. The basic DPCM framework is depicted in Figure 1.

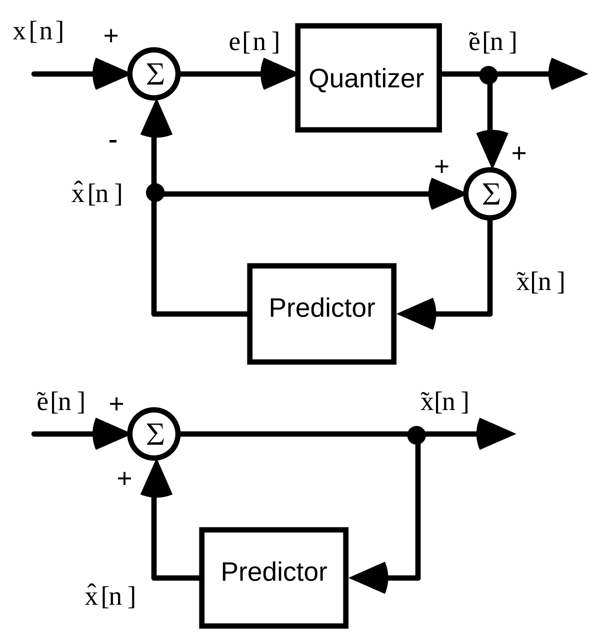

Figure 1: Basic DPCM structure.

In order to address some of the perceptual requirements of audio data compression, several well-known alterations to standard DPCM systems can be employed. One fundamental example is Noise Feedback Coding (NFC) [2, 3, 4].

Despite the desirable simplicity of standard DPCM the scheme has several drawbacks. In particular, standard DPCM has difficulty tracking rapidly varying, non-stationary signals, which are typical of wideband audio. Also, since it is necessary to minimize the bit rate in order to achieve useful compression, it is desirable to design the quantizer module with a small number of encoding bits [2]. When the prediction error e[n] is small it will fall in one of the inner levels of the quantizer. The resulting *granular* quantization noise, q[n], is commonly broadband and strictly bounded by the size of the quantization interval. However, choosing a small number of available quantizer levels to achieve a lower bit rate can result in increased distortion, referred to as overload noise, if e[n] ever falls in the outer levels of the quantizer. Quantizer overload results in undesirable signal correlated distortion which may persist for many samples due to the recirculation of error in the predictor feedback structure. However, avoiding overload noise typically requires an increase in quantizer step size and a corresponding increase in granulation noise, or an undesirable increase in the bit rate.

1.2 RIQ-DPCM: Avoiding Overload

The proposed coding technique involves a novel extension of adaptive DPCM through the use of a recursively indexed quantizer (RIQ), as described in the research description below. As will be shown, the RIQ allows a data rate increase only when the quantizer input is large, meaning the average rate can be kept low. In this way the RIQ-DPCM scheme provides a convenient tradeoff between quantizer step size and output bit rate.

2. RESEARCH DESCRIPTION

The major features of this research are:

- A low-complexity compression scheme capable of realtime operation on conventional microprocessor and DSP hardware.
- A system that eliminates DPCM quantizer overload distortion caused by predictor mismatch, thereby reducing the penalty for using a simple predictor structure.
- A compression scheme which can be adjusted easily to operate efficiently and with low distortion at different bit rates depending on the available storage or transmission bandwidth and the available computational resources.

2.1. RIQ Formulation

To handle the problem of efficiently encoding wideband audio signals, we have investigated the use of a quantizer called a recursively indexed quantizer (RIQ) [11]. The RIQ concept was proposed for use in image coding applications where the problem of preserving abrupt boundaries, or "edges", must be confronted. Since a similar situation occurs in wideband audio as sound sources start and stop it is rational to consider RIQ for efficient audio coding.

The main idea behind RIQ is to provide a guaranteed distortion level on a *per sample* basis. This is accomplished by allowing an increase in bit rate when the quantizer encounters large magnitude inputs. Since the increase in rate only occurs when the quantizer input is large, the effect on the average bit rate is generally minimal [11, 12]. A brief description of RIQ is given next.

For a given quantizer stepsize Δ and a positive integer K, the smallest and largest output levels, x_L and x_H, are defined:

$$x_L = -\left\lfloor \frac{K-1}{2} \right\rfloor \Delta, \qquad x_H = x_L + (K-1)\Delta$$

where $\lfloor y \rfloor$ is the largest integer not exceeding y. In other words, a recursively indexed quantizer of size K comprises a uniform quantizer with stepsize Δ and always has 0 as one of the output levels. The recursive aspect of the quantization rule Q is defined next. For a given input value x:

(i) If x falls in the interval $(x_L+(\Delta/2)),\ x_H-(\Delta/2))$, then $Q(x)$ is the output level nearest to x.

(ii) If x is greater than $x_H-(\Delta/2)$, determine whether the difference $x_1 \equiv x - x_H \in (x_L+(\Delta/2)),\ x_H-(\Delta/2))$.

If so, the quantized representation $Q(x) = (x_H, Q(x_1))$. If not, form the difference $x_2 = x - 2x_H$, then do the same as for x_1.

This recursive process continues until for some m, $x_m = x - mx_H$ falls in the interval $(x_L+(\Delta/2)),\ x_H-(\Delta/2))$, in which case the quantized representation becomes the sequence

$$Q(x) = \Big(\underbrace{x_H, x_H, \cdots, x_H}_{m}, Q(x_m)\Big).$$

(iii) If x is smaller than $x_L+(\Delta/2)$, a similar procedure is used: $x_m = x - mx_L$ is formed so that it falls in the interval $(x_L+(\Delta/2)),\ x_H-(\Delta/2))$, and is quantized into the sequence $(x_L, x_L, \cdots, x_L, Q(x))$.

The important properties of the RIQ for use in wideband audio data compression are emphasized here:

- The RIQ operates in two modes: a nonrecursive mode when its input falls in the interval $(x_L+(\Delta/2)),\ x_H-(\Delta/2))$, and a recursive mode when the input falls outside this range. In either mode the quantization error for any input is bounded by $\Delta/2$, since the quantizer effectively contains an infinite number of steps. If Δ is chosen to be smaller than the resolution of the input data stream the quantizer can operate in a lossless fashion, while a larger Δ results in lossy coding. Thus, the RIQ-DPCM structure is operable at a variety of bit rates to meet the acceptable level of distortion *simply by changing the stepsize*, Δ, of the RIQ.

- The output sequences produced when the quantizer is in the recursive mode consist of a stream of identical values (either x_L or x_H) followed by a quantized value, and since this number of different output symbols is limited, subsequent lossless coding can be applied to the output stream to further reduce this redundancy.

- Because overload distortion is eliminated and the absolute error of the quantizer ($\Delta/2$) is small, the DPCM predictor structure recovers immediately from a mismatched condition, unlike the error propagation that can occur in standard DPCM [12].

- Finally, the RIQ quantization noise is entirely granular in nature and is found to be uniformly distributed and uncorrelated with the input signal in most practical situations. Thus, the RIQ-DPCM structure is particularly well-suited to noise feedback modifications compared to standard DPCM that is susceptible to signal-correlated overload noise.

Sayood and Na [13] have studied the behavior of the RIQ-DPCM system with first-order Gauss-Markov and Laplace-Markov input sources. The results show that the RIQ-DPCM system performs at or close to the optimum entropy constrained DPCM system, but without the relatively complex iterative procedure typically required for quantizer design [14].

2.2. Performance Simulation

To demonstrate the proposed system on actual audio material several simulation experiments were conducted. Five arbitrary examples (approximately 10 seconds each) of 16-bit, 44.1 kHz sample rate audio material were obtained by digital transfer from Compact Disc (CD) sources. The example sources were a solo soprano singer, solo castanets, symphony orchestra, synthesizer pop-rock, and pop vocal-instrumental. For each source segment the signal was coded using four low-complexity techniques with similar computational requirements at a variety of bit rates:

- Simple DPCM with uniform quantizer and fixed first-order predictor (non-adaptive).

- Simple RIQ-DPCM with fixed first-order predictor (non-adaptive).

- DPCM with adaptive quantizer and fixed first-order predictor.

- DPCM with adaptive RIQ and fixed first-order predictor.

No noise shaping was used. For each of these techniques the average signal-to-error ratio (dB) for the five examples was computed. For the two techniques with fixed quantizers the step size, Δ, was chosen empirically to achieve the best measured signal-to-error performance across the five examples. The quantizer adaptation strategy was kept simple to minimize computation: $\Delta = 0.9\Delta$ when the quantizer input was not overloaded, and $\Delta = 1.5\Delta$ when overload occurred. The example bit rates represent the raw output of the DPCM and RIQ-DPCM systems, which could be further reduced somewhat in practice by subsequent entropy coding if sufficient computational resources were available. Furthermore, because the computational complexity was deliberately minimized for the purposes of this simulation, additional performance increases would be expected with a simple adaptive predictor in addition to a more sophisticated adaptive quantizer algorithm if sufficient resources were available.

The simulation results are summarized in Figure 2. The rate-distortion performance of the RIQ-ADPCM structures exceeds the corresponding performance of the non-RIQ techniques at all rates simulated in this test. *The reader is reminded* of the customary caveat that the SNR figures are based on the *average* over a specific set of input signals and are used here only for comparison among the techniques, not in the place of formal perceptual testing that will be done in the future. The results appear to indicate that the problem of mismatch between the predictor and the input signal can be diminished quite effectively by the recursively indexed quantizer DPCM approach.

3. CONCLUSIONS

We have described some of the theoretical and practical aspects of a new data compression framework for audio and other types of wideband signals that is appropriate for use when limited computational resources are available. The new DPCM formulation includes a recursively indexed quantizer (RIQ) to eliminate the problem of overload distortion, a simple predictor structure, and provision for basic noise-shaping features. Thus, this new formulation achieves a high degree of compression while exploiting the computational efficiency of DPCM.

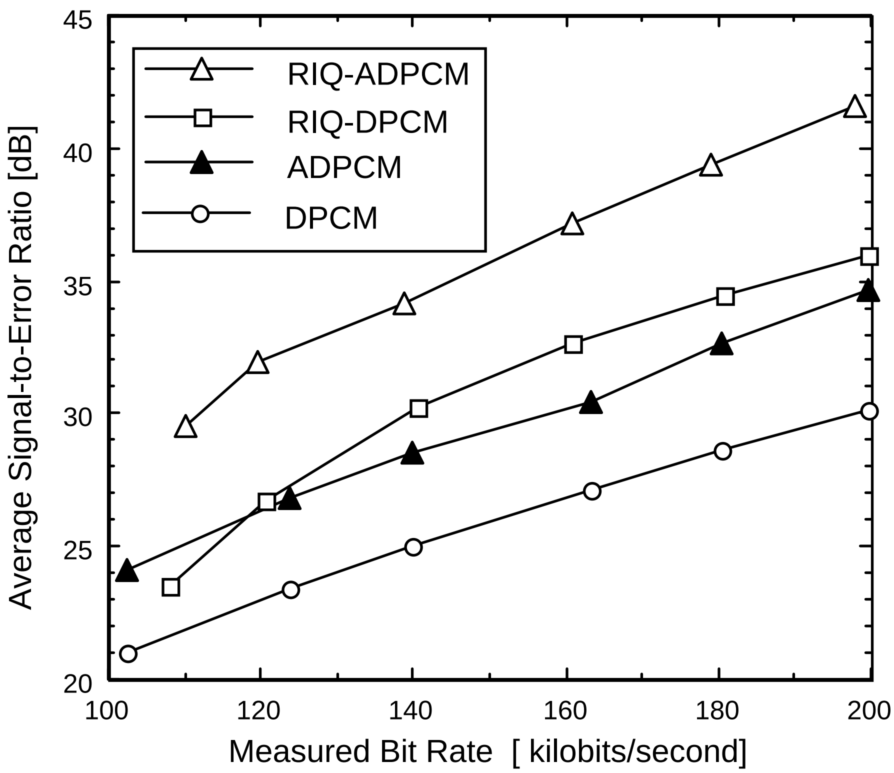

Figure 2: Simulation results and performance comparison of DPCM, DPCM with adaptive quantization, RIQ-DPCM, and RIQ-DPCM with adaptive quantization.

REFERENCES

1. N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM, and DM quantizers," *Proc. IEEE*, vol. 62 (May, 1974), pp. 611-632.
2. N. S. Jayant and P. Noll, Digital Coding of Waveforms, Englewood Cliffs, NJ: Prentice-Hall, 1984.
3. B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27 (June 1979), pp. 247-254.
4. J. C. Asmuth and J. D. Gibson, "Sequential noise spectral shaping in DPCM," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 2 (1984), pp. 228-235.
5. J. D. Gibson, "Adaptive prediction in speech differential encoding systems," *Proc. IEEE*, vol. 68, no. 4 (1980), pp. 488-525.
6. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Selected Areas Comm.*, vol. 6, no. 2 (1988), pp. 314-323.
7. J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding," *Proc. 1992 IEEE ICASSP*, vol. 2, (1992), pp. II-569 - II-572.
8. L. D. Fielder, "Low complexity transform coder for music applications," *Proc. 1989 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
9. G. Stoll and D. Wiese, "Optimizing of a sub-band coder for high quality audio signals by considering the ear's masking threshold," *Proc. 1989 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
10. G. Davidson, W. Anderson, and A. Lovrich, "A low-cost adaptive transform decoder implementation for high-quality audio," *Proc. 1992 IEEE ICASSP*, vol. 2, (1992), pp. II-193 - II-196.
11. K. Sayood and S. Na, "Recursively indexed quantization of memoryless sources," *IEEE Trans. Inf. Theory*, vol. 38, no. 5 (1992), pp. 1602-1609.
12. M. C. Rost and K. Sayood, "An edge preserving differential image coding scheme," *IEEE Trans. Image Processing*, vol. 1, no. 2 (1992), pp. 250-256.
13. K. Sayood and S. Na, "Recursively indexed differential pulse code modulation," *Joint DIMACS/IEEE Workshop on Coding and Quantization*, Rutgers Univ., Piscataway, NJ, Oct. 1992.
14. N. Farvardin and J. W. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Inf. Theory*, vol. IT-31 (1985), pp. 402-418.