PERFORMANCE OF PREDICTIVE CODERS OVER NOISY CHANNELS WITH FEEDBACK

Masoud Khansari

Hewlett-Packard Laboratories, Palo Alto, California, 94304

ABSTRACT

Predictive coding methods such as DPCM used for lossless coding of images or motion compensated hybrid video coders MPEG family are shown to compress the input signal well with a reasonable complexity. The performance of these coders, however, degrades considerably when the transmission channel is not error-free. This is due to the error propagation at the decoder where a single error can have catastrophic consequences. A low-rate feedback channel is shown to improve the overall performance. In this paper, we consider two such methods and provide the analysis and investigate different trade-offs.

1. INTRODUCTION

Traditionally, the goal of a image or a video compression scheme has been to compress the input information as much as possible for a given acceptable distortion. Compensating for the loss caused by transmission errors was assumed to be the responsibility of another functional unit, namely the channel encoder. As we have started using new time-varying transmission channels (e.g. radio fading links and packet data networks) for transmission, this separation of responsibility does not seem to provide an adequate performance. It is therefore of no surprise to see some of the new compression schemes being designed with channel errors in mind. For example, an error-resilient video compression algorithm tries to compress the input signal, organize and packetize the generated bit stream such that the impact of channel errors are contained to a minimum level.

Some of the most popular image and video compression algorithms are based on prediction where the previous symbols or frames are used to predict and encode the current symbol or frame. These methods are quite efficient and provide good compression. The effect of channel errors can however be catastrophic and without a powerful channel encoder or resynchronization of the source encoder the quality of the reconstructed signal will be unacceptable. The reason is the mitigation of errors where a single error can effect all the other signal units which directly or indirectly have used that symbol or frame for prediction. Many proposals are used to remedy this problem. One is to use leaky predictor which causes the error to eventually dies out with a rate which is depended on the leak factor. Another method is to re-start the encoder by coding a unit without prediction hence limiting any past errors which may have affected the previous units.

One of the methods which is currently being considered is to use a *thin* low-rate back channel from the receiver to the transmitter [1][2]. This back channel can provide the status of the transmitted information as well as the current state of the decoder. Using this back channel, the encoder can then try to contain and minimize the effect of the channel errors or retransmit the erroneously received information. Clearly, if the Round-Trip-Delay (RTD) is too long, the effectiveness of the back channel is marginal and the added complexity may not be justified. For a reasonable RTD then the main problem is to identify the best encoding method which achieves the best end-to-end performance. This is currently an open problem. Also, note that for different values of RTD the best encoding method may be different. In this paper, we analyze two proposed methods, investigate different trade-offs involved and provide a guideline for their selection.

2. PROBLEM STATEMENT

Figure 1 shows the general set-up for the prediction-based source encoder and decoder where the encoder emulates the operation of the decoder. Note the unit of the input signal x can be a pixel (as in DPCM) or a video frame (as in motion-compensated hybrid coders). The encoder emulates the operation of the decoder, i.e. the prediction is based on the reconstructed sequence $\tilde{x}$. The transmitted sequence q is the quantization of the prediction error e. The prediction filter P is not time-invariant and different filter can be used for encoding different input units. Sequence d - also transmitted to the receiver - indicates what predictor is being used. At the receiver the sequences p and d' are used to construct the sequence y and the reconstruction error is the different between the sequences x and y.

We assume that the channel is memoryless with the cross-over probability ε. Therefore, at the receiver, each sample of the transmitted sequence q and d are correctly received with probability $1 - \varepsilon$. We also assume that the receiver can always detect the errors and when an error is detected the last correctly decoded sample is frozen until the decoder is able to correctly decode a new unit. A unit is decoded correctly only if the transmitted residual q is received correctly and all the units used for the prediction of that unit are also decoded without error. Using the back channel the decoder informs the transmitter of the status of the decoded sequence by sending ACK or NACK signals. An ACK signal is sent when the transmitted information is received without error otherwise a NACK is sent. We assume that the back-channel is error free and all the ACK and NACK signals are received correctly. The encoder is then able to emulate the operation of the decoder and know

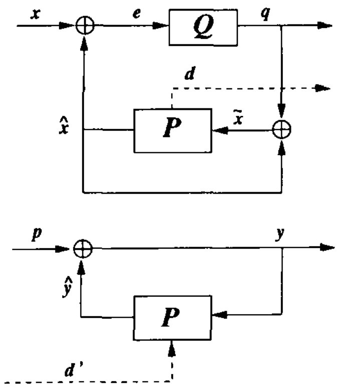

Figure 1: Encoder and Decoder Set-up

which units are reconstructed correctly. Note that, at the transmitter, the status of the signal transmitted at time n can be known no earlier than $n + RTD$ where RTD is the round-trip-delay.

As we stated previously, the predictor used is time-varying and the encoder can use the back channel to determine what predictor to use. Clearly, it is unwise to use any units that is not decoded correctly since the decoder will not be able to decode it. For simplicity, we assume that the prediction support is only one unit (or only one unit is used for prediction) and investigate the following two different strategies considered for error-resillient operation mode of MPEG-4 [1]:

1. ACK Method: In this method, the encoder uses the last acknowledged unit to predict the current unit. The advantage of this method lies on the fact that there will not be any error propagation and if an error occurs it will only affect that unit and the error will be contained within that unit. Note, however, that the last acknowledged unit is RTD units away and when this value is too large(i.e. the round-trip-delay is too long), the quality of the prediction will be poor and as a result the overall performance of the system will not be satisfactory.

2. NACK Method: In this mode, the last adjacent unit is always used for prediction unless a NACK is received in which case the transmitter uses the last acknowledged unit and it will then switch back to the normal mode of operation which is to use the last unit for prediction. Note that when a NACK is received the transmitter ignores the subsequent ACKs and NACKs for the duration of $RTD-1$ units as they are of no significance. The advantage of this method over the previous method is that the prediction is of better quality. The disadvantage is that for any single channel error the subsequent $RTD-1$ units cannot also be decoded correctly.

The main trade-off is therefore between the error propagation at the decoder and the prediction quality at the encoder. Qualitatively, one would expect that when the channel error rate is relatively high then the ACK method to have a better performance and for the low error-rate environment the NACK method to outperform the ACK method.

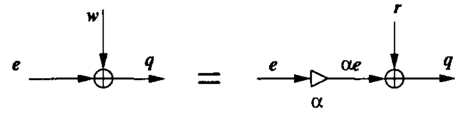

Figure 2: Quantization Noise Model

3. ANALYSIS

We model the quantization operation as a gain plus additive noise operation [3]. Figure 2 shows this model where $w = r - (1-\alpha)e$ is the added quantization noise where the component r is independent from e and $0 < \alpha \leq 1$. One can show that the noise variance σ_w^2 is equal to $\sigma_r^2 + (1-\alpha)^2\sigma_e^2$ and $\mathbf{E}[ew] = -(1-\alpha)\sigma_e^2$. Also, note that the variance of the quantization noise is depended on the variance of the prediction error. The usual high-rate additive white noise modeling of the quantization operation fails to capture the effect of prediction unless its variance is assumed to be a function of the input signal variance.

We assume the input process to be a first order autoregressive process ($AR(1)$) with correlation coefficient ρ, i.e. $x(n) = \rho x(n-1) + z(n)$ where z is the innovation process. Also, we use first order predictor

$$\hat{x}(n) = \rho^D \tilde{x}(n-D) \tag{1}$$

where D is the random variable corresponding to the distance (in terms of signal unit) between the current unit and the unit used for its prediction. Note that in the case of ACK scheme $D \geq RTD$. This predictor, which is not optimum, has been shown to provide a performance close to an optimum predictor.

Let K be the random variable corresponding to the number of consecutive channel errors counted from the time index n backward. For example, $K = 0$ if $q(n)$ is received correctly and $K = 1$ if $q(n)$ is received in error but $q(n-1)$ is received correctly, etc. Then the reconstructed sample for $x(n)$ is $y(n-K) = \tilde{x}(n-K)$ and, as shown in Appendix A, the energy of the reconstruction error can be closely approximated by

$$\sigma_s^2 = \frac{\mathbf{E}[A]}{1-\rho^2}\sigma_z^2 + \sigma_r^2, \tag{2}$$

where A is defined as

$$\begin{aligned} A &\triangleq 2(1-\rho^K) + (1-\alpha)^2(1-\rho^{2D}) \\ &\quad + 2(1-\alpha)(1-\rho^K)(1-\rho^{2D}). \end{aligned} \tag{3}$$

A is a two-dimensional function of random variables D and K. Clearly, the lower the expected value of A, the lower the variance of the reconstruction error hence the better the performance of the overall system. Note that D and K are not independent. For example, for the signal unit $x(n)$ and the system operating in the ACK mode, $0 \leq K \leq D$. This is because D signifies the last acknowledged unit (i.e. the signal unit $n-D$ is correctly decoded at the receiver). For long round-trip-delay RTD or low channel error ε, however, it is safe to assume that the processes D and K are independent. Under this assumption, to evaluate the overall performance of the system through equation 3, it would be sufficient to find $\mathbf{E}[1-\rho^K]$ and $\mathbf{E}[1-\rho^{2D}]$ which is addressed in the following two sub-sections.

3.1. ACK Method

In this method, the encoder uses the last acknowledged unit for prediction. Since it take at least one RTD to send information and receive its feedback from the receiver, the last acknowledged unit is at least one RTD away from the current unit or $D \geq RTD$. Since the transmission channel is memoryless D has a geometric distribution given by

$$P[D = l] \;=\; \varepsilon^{l-RTD}(1-\varepsilon) \qquad l \geq RTD, \qquad (4)$$

and it is straightforward to show

$$\mathbf{E}[1-\rho^{2D}] = \frac{1-\rho^{2RTD} - \varepsilon(\rho^2 - \rho^{2RTD})}{1-\varepsilon\rho^2}. \qquad (5)$$

In the case of the channel error, the concealment method used at the decoder is to freeze the last correctly decoded frame till a unit can be decoded without error. In the case of the ACK method there is no error propagation, as a result, the probability of each unit being decoded incorrectly is independent of the other units. This mean K is also geometrically distributed and is given by

$$P[K = l] \;=\; \varepsilon^{l}(1-\varepsilon) \qquad l \geq 0 \qquad (6)$$

and

$$\mathbf{E}[1-\rho^{K}] = \frac{\varepsilon(1-\rho)}{1-\varepsilon\rho}. \qquad (7)$$

Substituting (5) and (7) into (3) gives $\mathbf{E}[A]$.

3.2. NACK Method

In this method, the encoder always uses the last unit for prediction unless a NACK signal is received which signifies that all the last RTD transmitted units cannot be decoded correctly at the receiver. Therefore, to stop the error propagation, the encoder uses the last acknowledged unit for the prediction since it knows that a correct copy of it is available at the receiver. It then returns to the normal mode of operation until it receives the status of this transmitted unit from the receiver. For example, if at time n a NACK is received, the unit $n - RTD - 1$ is used unless that unit was also decoded incorrectly in which case $n - 2\ RTD - 1$ is used, etc.

The analysis of this method is more involved than the ACK method and in Appendix B we show that:

$$\mathbf{E}[1-\rho^{2D}] = 1-\rho^2\frac{\varepsilon(RTD-1)(1-\varepsilon\rho^{2RTD})+1-\varepsilon}{(\varepsilon RTD+1-\varepsilon)(1-\varepsilon\rho^{2RTD})}, \qquad (8)$$

and

$$\mathbf{E}[1-\rho^{K}] = 1-\frac{(1-\varepsilon)(1-\rho+\varepsilon(\rho-\rho^{RTD}))}{(\varepsilon RTD+1-\varepsilon)(1-\rho)(1-\varepsilon\rho^{RTD})}. \qquad (9)$$

Substituting (8) and (9) into (3) provides $\mathbf{E}[A]$.

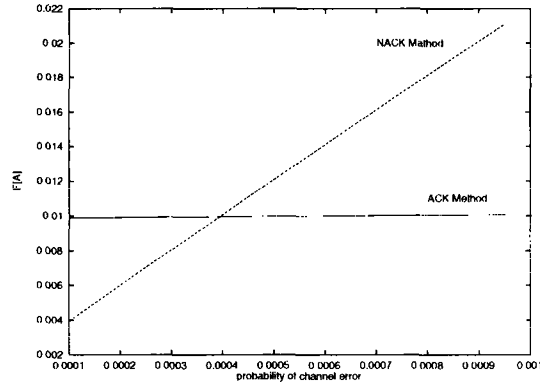

Figure 3: Performance of ACK and NACK methods ($RTD = 20$)

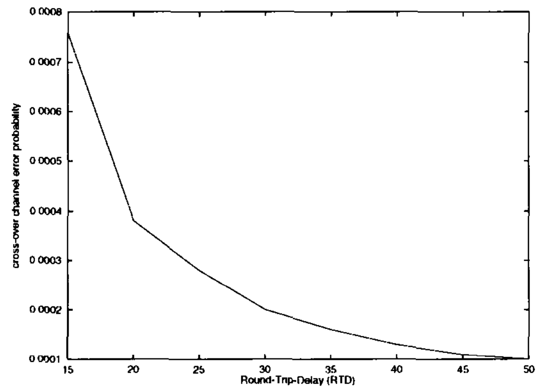

Figure 4: Cross-over probability between ACK and NACK methods

4. PERFORMANCE

In this section, we compare the performance of the two methods analyzed in this paper. For all the results presented in this section the input process is an $AR(1)$ process with correlation coefficient $\rho = 0.9$. We also assume that the gain factor of the quantization model used is $\alpha = 0.9$. As a result, the performance of both ACK and NACK methods are functions of the round-trip-delay RTD and the channel error probability ε.

Figure 3 shows the value of $\mathbf{E}[A]$ for both methods as a function of ε for $RTD = 20$. Note that the lower the value of $\mathbf{E}[A]$, the lower σ_s^2 hence the better the overall performance of the system. From the figure, we can say that for lower probability of errors the NACK method has better performance. There is a cross-over probability where for higher error probability the ACK method is better. Figure 4 shows this cross-over probability as a function of RTD. Note that for longer round-trip delay the cross-over probability is lower or there is a bigger region where the ACK method, in comparison, provides a better performance.

5. REFERENCES

[1] Ad-hoc group on core experiments on error resillience aspected of MPEG-4 video, ISO/IEC JTC1/SC29/WG11 N1586 MPEG97, March 1997.

[2] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environment," To appear in IEEE Trans. on CAS for Video Technology.

[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.

APPENDIX A

The reconstruction error for signal unit n is $s(n) = x(n) - y(n)$ where $y(n)$ is the same as $\tilde{x}(n)$ if it can be decoded correctly otherwise sample $y(n-1)$ is used as a replacement. But, the sample $y(n-1)$ itself could have been in error and sample $y(n-2)$ was used for its reconstruction, etc. Therefore, $y(n) = \tilde{x}(n-K)$ where K is the number of consecutive incorrect reconstruction from time index n backward, and

$$\begin{aligned} s(n) &= x(n) - \tilde{x}(n-K) \\ &= -(1-\rho^K)x(n-K) \\ &\quad + \sum_{l=0}^{K-1} \rho^l z(n-l) + w(n-K), \end{aligned} \tag{10}$$

and the variance of the reconstruction error is given by

$$\begin{aligned} \sigma_s^2 &= \mathbf{E}[(x(n) - \tilde{x}(n-K))^2] \\ &= \frac{2(1-\rho^K)}{1-\rho^2}\sigma_z^2 + (1-\alpha)^2\sigma_e^2 + \sigma_r^2 \\ &\quad -2(1-\rho^K)\mathbf{E}[x(n-K)w(n-K)]. \end{aligned} \tag{11}$$

The main assumption in the development of this analysis is that the processes x and $\tilde{x}$ are statistically equivalent or $\mathbf{E}[x^2] = \mathbf{E}[x\tilde{x}] = \mathbf{E}[\tilde{x}^2]$. It is then straightforward to show that

$$\begin{aligned} \sigma_e^2 &= \mathbf{E}[(x(n) - \rho^D\tilde{x}(n-D))^2] \\ &\cong \mathbf{E}[(x(n) - \rho^D x(n-D))^2] \\ &= \frac{1-\rho^{2D}}{1-\rho^2}\sigma_z^2, \end{aligned} \tag{12}$$

and

$$\begin{aligned} \mathbf{E}[x(n)w(n)] &= -(1-\alpha)\mathbf{E}[x(n)(x(n) - \rho^D\tilde{x}(n-D))] \\ &\cong -(1-\alpha)\mathbf{E}[x(n)(x(n) - \rho^D x(n-D))] \\ &= -(1-\alpha)\frac{1-\rho^{2D}}{1-\rho^2}\sigma_z^2. \end{aligned} \tag{13}$$

substituting (12) and (13) into (11), we arrive at (3).

APPENDIX B

The decision about the value of K is made at the decoder whereas for D, it is made at the encoder. Let us first consider K. The key to the analysis is to notice that when the first channel error is detected, not only that unit but also the next $RTD-1$ units cannot be decoded correctly. Therefore the duration of the error is at least RTD. When the NACK signal for the first unit arrives at the transmitter, the encoder uses the last positively acknowledged unit for prediction. This transmission can itself become corrupted in which case another RTD units are decoded incorrectly. Therefore the duration of consecutive errors is a multiple of RTD units with the average length of $RTD/(1-\varepsilon)$. One can also show that the average length for consecutive correctly received unit is $1/\varepsilon$ and the probability of a sample being decoded incorrectly is therefore given by

$$P_b = \frac{RTD/(1-\varepsilon)}{RTD/(1-\varepsilon) + 1/\varepsilon}. \tag{14}$$

Now Given that the unit is decoded incorrectly, the probability of it being in the first RTD units is

$$(1-\varepsilon)P_b = \frac{\varepsilon(1-\varepsilon)RTD}{\varepsilon RTD + 1 - \varepsilon}. \tag{15}$$

Similar expressions can be written for the second, third,... RTD units and we can show that:

$$P[K = l] = \frac{(1-\varepsilon)\varepsilon^{\lfloor \frac{l}{RTD} \rfloor}}{\varepsilon RTD + 1 - \varepsilon}, \tag{16}$$

where $\lfloor x \rfloor$ is the largest integer which is smaller than x, and after some algebraic manipulations, we have

$$\begin{aligned} \mathbf{E}[1-\rho^K] &= \sum_{l=0}^{\infty}(1-\rho^l)P[K=l] \\ &= 1 - \frac{(1-\varepsilon)(1-\rho+\varepsilon(\rho-\rho^{RTD}))}{(\varepsilon RTD + 1 - \varepsilon)(1-\rho)(1-\varepsilon\rho^{RTD})} \end{aligned} \tag{17}$$

Let us now consider D. When the transmitter receives a NACK from the receiver, it terminates the usual mode of the operation and uses the last acknowledged unit for prediction. Clearly this unit is $RTD+1$ unit away unless that unit itself was not decoded correctly in which case the last acknowledged unit is $2RTD+1$ unit away, etc. After this, the encoder ignores the next RTD feedback signals since it knows that the corresponding units are not correctly decoded at the receiver and are not reliable. If then the transmitter receives an ACK, it goes back to its normal operation mode otherwise again it uses the last acknowledged unit. Therefore, D can only have values of the form $kRTD+1$ for $k \geq 0$. We can then show the following:

$$\begin{aligned} P[D=1] &= \frac{\varepsilon RTD + 1 - 2\varepsilon}{\varepsilon RTD + 1 - \varepsilon} \\ P[D = kRTD+1] &= \frac{\varepsilon^k(1-\varepsilon)}{\varepsilon RTD + 1 - \varepsilon} \qquad k = 1, 2, \ldots \end{aligned} \tag{18}$$

and zero for the other values. Using the above, we arrive at:

$$\begin{aligned} \mathbf{E}[1-\rho^{2D}] &= \sum_{l=0}^{\infty}(1-\rho^{2lRDT+2})P[D = lRDT+1] \\ &= 1 - \rho^2\frac{\varepsilon(RTD-1)(1-\varepsilon\rho^{2RTD}) + 1 - \varepsilon}{(\varepsilon RTD + 1 - \varepsilon)(1-\varepsilon\rho^{2RTD})}. \end{aligned} \tag{19}$$