Presented at the 22nd Annual Conference on Information Science and Systems, March 1988, Princeton, New Jersey.

MAXIMUM APOSTERIORI JOINT SOURCE/CHANNEL CODING

N92-23417

74221

p. 12

Khalid Sayood*
Department of Electrical Engineering
University of Nebraska
Lincoln, Nebraska 68588-0511
(402) 472-6688

Jerry D. Gibson
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843
(409)845-7441

1. INTRODUCTION

One of Shannon's many fundamental contributions was his result that source coding and channel coding can be treated separately without any loss of performance for the overall system [1]. The basic design procedure is to select a source encoder which changes the source sequence into a series of independent, equally likely binary digits followed by a channel encoder which accepts binary digits and puts them into a form suitable for reliable transmission over the channel. However, the separation argument no longer holds if either of the following two situations occur:

i. The input to the source decoder is different from the output of the source encoder, which happens when the link between the source encoder and source decoder is no longer error free, or

ii. when the source encoder output contains redundancy.

Of course, case (i) occurs when the channel coder does not achieve zero error probability and case (ii) occurs when the source encoder is suboptimal. These two situations are common occurrences in practical systems where source or channel models are imperfectly known, complexity is a serious issue, or significant delay is not tolerable. Various approaches have been developed for such situations. They are usually grouped under the general heading of joint source/channel coding.

Most of the various joint source channel coding approaches can be classified in two main categories; (A) approaches which entail the modification of the source coder/decoder structure to reduce the effect of channel errors, and (B) approaches which examine the distribution of bits between the source and channel coders. The first set of approaches can be divided still further into two classes. One class of approaches examines the modification of the overall structure, while the other deals with the modification of the decoding procedure to take advantage of the redundancy in the output of the source coder.

To the first class belongs the work of Dunham & Gray [2] who proved the existence of joint source channel trellis coding systems for certain fidelity criteria, and a design of a joint source channel trellis coder presented by Ayanoğlu and Gray [3], where the design procedure is the generalized Lloyd algorithm. Further, Massey [4] and Ancheta [5] showed that for distortionless transmission of the source using linear joint source channel encoders, equivalent performance can be obtained with a significant reduction in complexity. Chang and Donaldson [6] propose modifications to the DPCM system to reduce the effect of channel errors, while Kurtenbach and Wintz [7] and Farvardin and Vaishampayan [8] study the problem of optimum quantizer design for noisy channels. Goodman and Sundberg [9,10] propose an embedded DPCM system which consists of a two bit DPCM and a two bit PCM system in parallel.

In the second class of category A, we include the work of Reininger and Gibson [11], who use the fact that coefficients in neighboring blocks in a transform coding scheme will not vary greatly, and thus use coefficients from neighboring blocks to correct a possible error, and the work of Steele, Goodman and McGonegal [12,13], who propose a difference detection and correction scheme for broadcast quality speech. In this scheme the receiver infers an error whenever an individual sample to sample difference is greater than the mean squared difference of a 21 sample sliding block. When an error is detected, the received sample is replaced by the output of a smoothing circuit. Ngan and Steele [14] use a similar method for recovering from errors in an image transmission system. Sayood and Borkenhagen [15,16] use the redundancy at the source coder output to perform sequence estimation.

*Supported by NASA Lewis Research Center (NAG 3-806) and NASA Goddard Space Flight Center (NAG 5-916)

The work of Modestino, Daut and Vickers [17] belongs to category B. In their study of transform coding they examine tradeoffs between allocating bits for source and channel coding. Comstock and Gibson [18] extend this work and provide an explicit mechanism for allocating bits between a source coder and a Hamming channel coder. Additionally, Moore and Gibson [19] study the allocation of bits between a DPCM coder and self orthogonal convolutional coding.

In this paper we present a maximum aposteriori probability (MAP) approach to joint source/channel coder design, which belongs to category A, and hence we explore a technique for designing joint source/channel coders, rather than ways of distributing bits between source coders and channel coders. We assume that the two nonideal situations referred to earlier are present. Our approach is as follows. For a nonideal source coder, we use MAP arguments to design a decoder which takes advantage of redundancy in the source coder output to perform error correction. Once the decoder is obtained, we analyze it with the purpose of obtaining "desirable properties" of the channel input sequence for improving overall system performance. We then propose an encoder design which incorporates these properties.

2. THE MAP DESIGN CRITERION

For a discrete memoryless channel (DMC), let the channel input alphabet be denoted by $A = \{a_0, a_1, \ldots, a_{M-1}\}$, and the channel input and output sequences by $Y = \{y_0, y_1, \ldots, y_{L-1}\}$ and $\hat{Y} = \{\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{L-1}\}$, respectively. If $\mathcal{A} = \{A_i\}$ is the set of sequences $A_i = \{a_{i,0}, a_{i,1}, \ldots, a_{i,L-1}\}$, $a_{i,k} \epsilon A$, then the optimum receiver (in the sense of maximizing the probability of making a correct decision) maximizes $P[C]$, where

$$P[C] = \sum_{A_i} P[C|\hat{Y}]P[\hat{Y}].$$

This in turn implies that the optimum receiver maximizes $P[C|\hat{Y}]$. When the receiver selects the output to be A_k, then $P[C|\hat{Y}] = P[Y = A_k|\hat{Y}]$. Thus, the optimum receiver selects the sequence A_k such that

$$P[Y = A_k|\hat{Y}] \geq P[Y = A_i|\hat{Y}]\,\forall_i.$$

Lemma 1

Let y_i be the input to a DMC. Given y_{i-1}, y_i is conditionally independent of $y_{n-k}, k > 1$. If $\hat{y}_0 = y_0$ then the optimum receiver selects a sequence A_i to maximize $\Pi_{i=1}^{L-1} p(y_i|y_{i-1}, \hat{y}_i)$.

Proof:

From the preceding result, the receiver tries to maximize $P(Y|\hat{Y})$. Using the chain rule we can write this as

$$\begin{aligned} P(Y|\hat{Y}) &= P(y_0, y_1, \ldots, y_{L-1}|\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{L-1}) \\ &= P(y_{L-1}|y_{L-2}, y_{L-3}, \ldots, y_0, \hat{y}_0, \ldots, \hat{y}_{L-1}) \\ &P(y_{L-2}|y_{L-3}, \ldots, y_0, \hat{y}_0, \ldots, \hat{y}_{L-1}) \ldots P(y_0|\hat{y}_0, \ldots, \hat{y}_{L-1}) \end{aligned}$$

The last factor on the right hand side (RHS) is equal to one. Using the assumption of the DMC, we obtain

$$P(Y|\hat{Y}) = \Pi_{i=1}^{L-1} p(y_i|y_{i-1}, \hat{y}_i). \quad (1)$$

□

The lemma addresses the situation in case (ii), i.e., the situation in which the source coder output (which is also the channel input sequence) contains redundancy. Using this lemma, we can design a decoder which will take advantage of dependence in the channel input sequence.

3. DECODER DESIGN

The lemma of the previous section provides the mathematical structure for the decoder. The physical structure can be easily obtained by examining the quantity to be maximized. The decoder maximizes $P(Y|\hat{Y})$ or equivalently $\log P(Y|\hat{Y})$, but

$$\log P(Y|\hat{Y}) = \sum \log P(y_i|\hat{y}_i, y_{i-1}) \qquad (2)$$

and various solutions exist for the maximization of additive path metrics. To implement this decoder we need to be able to compute the path metric. This task is considerably eased by the following lemma.

Lemma 2 Let A be the channel input alphabet and $\{y_i\}$ and $\{\hat{y}_i\}$ be the input and output sequences of a DMC. Then

$$P[y_i = a_j | y_{i-1} = a_m, \hat{y}_i = a_n] =$$

$$\frac{P[\hat{y}_i = a_n | y_i = a_j] P[y_i = a_j | y_{i-1} = a_m]}{\sum_l P[y_i = a_l | y_{i-1} = a_m] P[\hat{y}_i = a_n | y_i = a_l]} \qquad (3)$$

Proof: See [16].

The expression on the RHS of (3), while it looks more complicated, is actually a much more tractable form of the desired conditional probability. Note that this expression is a function of two distinct sets of transition probabilities, the channel transition probabilities and the source coder output transition probabilities. As the channel transition probabilities depend only on the channel, and the source coder and source transition probabilities depend only on the source coder and source probabilities, the two sets of transition probabilities can be estimated independently. The two can then be combined according to (3) to construct a M x M x M lookup table for use in decoding. If the source coder or source changes, the only parameters to be modified are the source coder output transition probabilities. Results using a DPCM source coder with an image as the source are presented in Section 6.

4. DECODER ANALYSIS

In the previous section we developed a scheme for providing error correction using the redundancy in the channel input sequence, or the source coder output. We looked at the design of the decoder given a source coder or channel input sequence with some rather general statistical properties. In this section we examine the reverse problem. That is, given the decoder obtained in the previous section we look for "desired properties" of the channel input sequence and, hence, the source coder.

To obtain the desired properties we need to examine the factors involved in the error correcting capability of the decoder. Toward this end let us examine the following situation. Referring to Figure 1, assume that the correct sequence of transmitted codewords is $a_0a_0a_0$. An error occurs if the path metric for $a_0a_ja_0$ is greater than the path metric of $a_0a_0a_0$. Assume $\hat{y}_1 = a_n$ and $\hat{y}_2 = a_0$. An error occurs if the following quantity is positive.

$$\log \frac{P[\hat{y}_1 = a_n | y_1 = a_j] P[y_1 = a_j | y_0 = a_0]}{\sum_l P[y_1 = a_l | y_0 = a_0] P[\hat{y}_1 = a_n | y_1 = a_l]}$$

$$+ \log \frac{P[\hat{y}_2 = a_0 | y_2 = a_0] P[y_2 = a_0 | y_1 = a_j]}{\sum_l P[y_2 = a_l | y_1 = a_j] P[\hat{y}_2 = a_0 | y_2 = a_l]}$$

$$- \log \frac{P[\hat{y}_1 = a_n | y_1 = a_0] P[y_1 = a_0 | y_0 = a_0]}{\sum_l P[y_1 = a_l | y_0 = a_0] P[\hat{y}_1 = a_n | y_1 = a_l]}$$

$$- \log \frac{P[\hat{y}_2 = a_0 | y_2 = a_0] P[y_2 = a_0 | y_1 = a_0]}{\sum_l P[y_2 = a_l | y_1 = a_0] P[\hat{y}_2 = a_0 | y_2 = a_l]} \qquad (4)$$

Defining

$$d_{ij} = \text{Hamming distance between } a_i \text{ and } a_j$$

$$g_{lk} = P[y_i = a_l | y_{i-1} = a_k] \qquad (5)$$

then using (5) and simplifying (4), an error occurs if

$$(d_{nj} - d_{n0}) \log\left(\frac{p}{1-p}\right) + \log \frac{g_{j0}}{g_{00}} + \log \frac{g_{0j}}{g_{00}} - \log \frac{\sum_l \left(\frac{p}{1-p}\right)^{d_{nl}} g_{lj}}{\sum_l \left(\frac{p}{1-p}\right)^{d_{nl}} g_{l0}} > 0 \quad (6)$$

Defining $\alpha = \frac{1-p}{p}$, (6) can be rewritten as

$$(d_{n0} - d_{nj}) \log \alpha + \log \frac{g_{j0}}{g_{00}} + \log \frac{g_{0j}}{g_{00}} - \log \frac{\sum_l (\alpha^{-d_{nl}} g_{lj})}{\sum_l (\alpha^{-d_{nl}} g_{l0})} > 0 \qquad (7)$$

or

$$d_{n0} - d_{nj} > \frac{1}{\log \alpha} \left(\log \frac{g_{00}}{g_{j0}} + \log \frac{g_{00}}{g_{0j}} + \log \frac{\sum_l \alpha^{-d_{nl}} g_{lj}}{\sum_l \alpha^{-d_{nl}} g_{l0}} \right) \qquad (8)$$

The left hand side is maximized when $j = n (d_{nj} = 0)$. Thus an error occurs when the number of bit errors, which in this case is d_{n0}, is greater than the quantity on the LHS of (8) or

$$d_{n0} > \frac{1}{\log \alpha} \left(\log \frac{g_{00}}{g_{j0}} + \log \frac{g_{00}}{g_{0j}} + \log \frac{\sum_l \alpha^{-d_{nl}} g_{lj}}{\sum_l \alpha^{-d_{nl}} g_{l0}} \right). \qquad (9)$$

The alternative path shown in Figure 1 is only one of possible paths. Another longer alternative path is shown in Figure 2.
In this case the number of errors required to take the alternative path is given by

$$d_{n0} > d_{m0} + \frac{1}{\log \alpha} \left(\log \frac{g_{00}}{g_{mn}} + \log \frac{g_{00}}{g_{0m}} \right)$$

$$+ \frac{1}{\log \alpha} \left(\log \frac{\sum \alpha^{-d_{jl}} g_{l0}}{\sum \alpha^{-d_{jl}} g_{ln}} + \log \frac{\sum \alpha^{-d_{sl}} g_{l0}}{\sum \alpha^{-d_{sl}} g_{lm}} \right) \qquad (10)$$

Notice that the number of errors required to take a longer incorrect path (a path with more branches) is larger than for a shorter incorrect path. To make our statements more concrete, we define a parameter we call the error correction capability I as

$$I = 1 - H(Y_n | Y_{n-1}) / \log M =$$

$$1 - \frac{1}{\log M} \sum_{l,k} p(y_n = a_l, y_{n-1} = a_k) \log \frac{1}{P(y_n = a_l) | y_{n-1} = a_k)}$$

$$= 1 - \frac{1}{\log M} \sum_{l,k} p(y_n = a_l, y_{n-1} = a_k) \log \frac{1}{g_{lk}} \qquad (11)$$

where M is the size of the channel input alphabet. We immediately note the following properties of I

(i) I is a convex cup function of the conditional probabilities $\{p(y_n | y_{n-1})\}$.
(ii) $0 \leq I \leq 1$.

Further properties of I are developed in the following lemmas.

Lemma 3: If I is zero for a particular channel input sequence, the decoder will not correct any errors.

Proof:

I is zero when

$$\sum p(y_n, y_{n-1}) \log \frac{1}{g_{lk}} = \log M$$

This is true when $g_{lk} = \log \frac{1}{M}$ for all l, k. In this condition the right hand side of (9) is zero giving the desired result.

□

Lemma 4: If I is one for a particular input sequence, the decoder obtains the correct sequence with probability one.

Proof:

For I to be one, $H(y_n | y_{n-1})$ has to be zero. This is true if for each k_0 there exists an l_0 such that

$$g_{lk_0} = \begin{cases} 1, & l = l_0 \\ 0, & l \neq l_0. \end{cases}$$

This in turn implies that there exists some i_0 such that

$$p(A_i) = \begin{cases} 1, & i = i_0 \\ 0, & i \neq i_0. \end{cases}$$

Thus

$$P(Y = A_i | \hat{Y}) = \begin{cases} 1, & i = i_0 \\ 0, & i \neq i_0 \end{cases}$$

and the decoder will pick the correct sequence with probability one. $\square$

The above two lemmas provide a relationship between the value of I and the error correcting capability of the decoder, for the extreme values of I. To obtain an insight into the relationship for other values of I we look at a simplified version of (9). Assume that the size of the channel input alphabet is two, then (9) simplifies to

$$d_{10} > \frac{1}{\log \alpha}\left(\log\frac{g_{00}}{g_{10}} + \log\frac{g_{00}}{g_{01}} + \log\left(\frac{g_{01} + \alpha^{-d_{01}} g_{11}}{g_{00} + \alpha^{-d_{01}} g_{10}}\right)\right). \qquad (12)$$

Noting that $g_{00} = 1 - g_{10}, g_{01} = 1 - g_{11}$, and for the right hand side to be positive $g_{00} > \frac{1}{2}$ and $g_{01} < \frac{1}{2}$ we can rewrite (12) as

$$d_{10} > \frac{1}{\log \alpha}\left(\log\frac{g_{00}}{g_{10}} + \log\left(\frac{1 + \alpha^{-d_{01}}\frac{g_{11}}{g_{01}}}{1 + \alpha^{-d_{01}}\frac{g_{10}}{g_{00}}}\right)\right). \qquad (13)$$

In (13) the larger the right hand side the greater is the error correcting capability of the receiver. The right hand side can be increased by decreasing g_{01} below $\frac{1}{2}$. Thus the error correcting capability increases as g_{01} decreases below $\frac{1}{2}$. If we examine I we find that I increases as g_{01} decreases below $\frac{1}{2}$. This is because

$$\begin{aligned} H(Y_n|Y_{n-1}) &= p_0\left((g_{00})\log\tfrac{1}{g_{00}} + g_{10}\log\tfrac{1}{g_{10}}\right) \\ &+ p_1\left(g_{01}\log\tfrac{1}{g_{01}} + p_1(1-g_{01})\log\tfrac{1}{1-g_{01}}\right) \qquad (14) \end{aligned}$$

decreases with g_{01} decreasing below $\frac{1}{2}$. Thus for this simple example, an increase in I means an increase in the error correcting capability of the decoder.

5. ENCODER DESIGN

In the previous section we obtained desirable properties for the channel input/encoder output sequence. In this section we examine ways of incorporating these desireable properties into the encoder. We wish to do this without decreasing the redundancy removal capability of the source coder, and if possible, without increasing the transmitted bit rate. To see how to approach this problem let us first examine the source coder for noiseless channels in some detail.

In general, a source coder consists of two operations, data compression and data compaction [20]. The data compression operation usually consists of redundancy removal and involves some loss of information. Examples of data compression schemes are DPCM, transform coding and vector quantization. The data compaction schemes are information preserving. They may result in a variable rate out. Examples include Huffman coding and runlength coding. Generally in discussions of joint source/channel coder design, the data compaction operations are not included. The reason for this is that due to the variable rate output, the data compaction schemes are highly vulnerable to channel noise and, therefore, are not considered for noisy channel applications.

A possible way of achieving our objectives is to insert another operation between the data compression and data compaction steps as shown in Figure 3.
To satisfy our objectives, the Π operator should have the following properties.

(a) The Π operator should perform distortionless encoding.

(b) The Π operator should increase the error correcting capability.

(c) The Π operator should not increase the bit rate. For the case where the data compaction scheme is a Huffman coder, this is equivalent to the condition that the output entropy not be greater than the input entropy.

An example of the Π operator which satisfies (a) and (b) and which can be modified to satisfy (c) functions as follows. Let the input to Π be selected from the alphabet

$$A = \{a_0, a_1, \ldots, a_{N-1}\},$$

and let the output alphabet be denoted by

$$S = \{s_0, s_1, \ldots, s_{N^2-1}\}.$$

Then the input/output mapping is given by

$$x_n = a_i, x_{n-1} = a_j \Longrightarrow y_n = s_{jN+i}. \qquad (15)$$

The effect of the Π operator is to increase the distance between alternative sequences. To see this, let us construct a simple example. Let $A = \{a_0, a_1\}$ and $S = \{s_0, s_1, s_2, s_3\}$ then

$$\begin{array}{llll} y_n = s_0 & \text{if} \quad x_n = a_0 & \text{and} & x_{n-1} = a_0, \\ y_n = s_1 & \text{if} \quad x_n = a_1 & \text{and} & x_{n-1} = a_0, \\ y_n = s_2 & \text{if} \quad x_n = a_0 & \text{and} & x_{n-1} = a_1, \\ y_n = s_3 & \text{if} \quad x_n = a_1 & \text{and} & x_{n-1} = a_1. \end{array}$$

In this case if $y_n = s_0$, y_{n+1} cannot be s_2 or s_3 because $y_n = s_0$ means $x_n = a_0$, and $y_{n+1} = s_2$ or s_3 means $x_n = a_1$. Thus a decoded sequence cannot have s_2 or s_3 following s_0.

For simplicity let us ignore the Huffman coder and assign fixed length codewords to the s_i as

$$s_0 : 00,\ s_1 : 01,\ s_2 : 10,\ s_3 : 11$$

Now suppose the transmitted sequence was the all zero sequence, the metric used was the Hamming distance, and the received sequence is 00001000000000; that is, there is an error in the fifth bit. If the receiver decoded the first four bits as $s_0 s_0$ then it cannot decode the fifth and sixth bits as s_2 for the reason noted above. The only two options are decoding them as s_0 or s_1. If we decoded them as s_0, we could continue decoding the rest of the sequence as $s_0 s_0 \ldots$, and the Hamming distance between the received and decoded sequence would be one. If we decoded them as s_1, we would have to decode the next set of two bits as s_2 or s_3 because s_0 cannot follow s_1. Decoding as s_2 gives the smallest Hamming distance so we decode the seventh and eighth bit as s_2. This gives a total Hamming distance of two for the incorrect path. Thus the receiver will select the correct path (the path with the smallest Hamming distance).

6. SIMULATION RESULTS

We present the results of simulating two different systems in this section. The first set of results were obtained using a nonideal source coder with the decoder proposed in Section 3. The second set of results pertain to the system proposed in the previous section. In both cases the data compression scheme is a DPCM system with a fixed one tap predictor and a nonuniform Lloyd-Max quantizer.

The source for the first set of results is the USCGIRL image. The source coder output transition probabilities were obtained using a training set. The training image was the USC COUPLE image. The performance measure was the Peak-signal-to-noise-ratio (PSNR) defined as

$$PSNR(dB) = 10\log_{10}\left(\frac{\sum(255)^2}{\sum(x_i - \hat{x}_i)^2}\right)$$

where x_i is the input to the source coder while $\hat{x}_i$ is the output of the source decoder. Figure 4 shows the performance comparison for a two bit per pixel system. Most of the performance improvement is available at high probabilities of error. At these probabilities of error, however, the improvement is substantial. Figure 5 shows the same kind of results for a four bit per pixel system. The performance improvement for this case are even more substantial than those for the two-bit system. Two things are especially noteworthy in these results. The first one is that the performance improvement does not really become significant until the channel is very noisy. The other is that the performance curve in the high noise region is relatively flat. This means that even very noisy channels may be usable for image transmission. Further results including perceptual results can be found in [16].

The second set of results were obtained using the approach proposed in Section 5. The source encoder was replaced by the proposed joint source/channel coder. The Π operator used is the one described in the previous section. The source again was the USCGIRL image, and end of line synchronization was assumed. The performance comparison is shown in Figure 6. Note that unlike the previous case, the performance improvement occurs at both low and high error probabilities. This makes the scheme especially useful for transmission at low error rates.

7. CONCLUSIONS

In this paper we have presented a MAP approach to joint source/channel coder design. The approach is based in part on the fact that source coders are, in general, nonideal and, therefore, cannot remove all redundancy from a source. This nonideality is taken advantage of, by a MAP decoder, to correct errors. The decoder is analyzed to obtain desired properties for the encoder output sequence. A joint source/channel encoder design approach is presented which incorporates the desired properties, and examples are given which show that considerable performance improvements can be obtained with the proposed approach.

8. REFERENCES

1. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
2. J. G. Dunham and R. M. Gray, "Joint Source and Channel Trellis Encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 516–519, July 1981.
3. E. Ayanoğlu and R. M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 855–865, November 1987.
4. J. L. Massey, "Joint Source and Channel Coding," in *Communication Systems and Random Process Theory*, J. K. Skwirzynski, ed., Sijthoff and Nordhoff: Netherlands, pp. 279–293, 1978.
5. T. C. Ancheta, Jr., "Bounds and Techniques for Linear Source Coding," Ph.D. dissertation, Dept. of Electrical Engineering, Univ. of Notre Dame, August 1977.
6. K-Y Chang and R. W. Donaldson, "Analysis, Optimization, and Sensitivity Study of Differential PCM Systems Operating on Noisy Communication Channels," *IEEE Trans. Commun.*, vol. COM-20, pp. 338–350, June 1972.
7. A. J. Kurtenbach and P. A. Wintz, "Quantizing for Noisy Channels," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 291–302, April 1969.
8. N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An approach to Combined Source-Channel Coding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827–838, November 1987.
9. D. J. Goodman and C. E. Sundberg, "Combined Source and Channel Coding for Variable Bit Rate Speech Transmission," *Bell Syst. Tech. J.*, vol. 62, pp. 2017–2036, September 1983.
10. D. J. Goodman and C. E. Sundberg, "Transmission Errors and Forward Error Correction in Embedded Differential Pulse Code Modulation," *Bell Syst. Tech. J.*, vol. 62, pp. 2735–2764, November 1983.
11. R. C. Reininger and J. D. Gibson, "Soft Decision Demodulation and Transform Coding of Images," *IEEE Trans. Commun.*, vol. COM-31, pp. 572–577, April 1983.
12. R. Steele, D. J. Goodman, and C. A. McGonegal, "A Difference Detection and Correction Scheme for Combatting DPCM Transmission Errors," *IEEE Trans. Commun.*, vol. COM-27, pp. 252–255, January 1979.
13. R. Steele, D. J. Goodman, and C. A. McGonegal, "Partial Correction of Transmission Errors in DPCM without recourse to Error Correction Coding," *Elec. Lett.*, vol. 13, pp. 351–353, June 1977.
14. K. N. Ngan and R. Steele, "Enhancement of PCM and DPCM Images Corrupted by Transmission Errors," *IEEE Trans. Commun.*, vol. COM-30, pp. 257–269, January 1982.
15. K. Sayood and J. C. Borkenhagen, "Utilization of Correlation in Low Rate DPCM Systems for Channel Error Protection," *Proceedings IEEE International Conference on Communications*, June 1986, pp. 1888–1892.
16. K. Sayood and J. C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," submitted to *IEEE Transactions on Communications*.
17. J. W. Modestino, D. G. Daut, and A. L. Vickers, "Combined Source-Channel Coding of Images Using the Block Cosine Transform," *IEEE Trans. Commun.*, vol. COM-29, pp. 1262–1274, September 1981.
18. D. Comstock and J. D. Gibson, "Hamming Coding of DCT Compressed Images Over Noisy Channels," *IEEE Trans. Commun.*, vol. COM-32, pp. 856–861, July 1984.
19. C. C. Moore and J. D. Gibson, "Self-Orthogonal Convolutional Coding for the DPCM-AQB Speech Encoder," *IEEE Trans. Commun.*, vol. COM-32, August 1984.
20. R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, 1987.

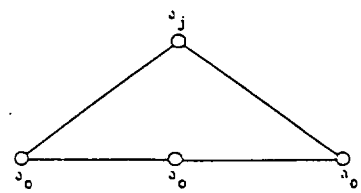

Figure 1. Alternative Paths at the Receiver

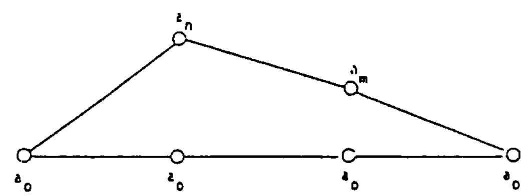

Figure 2. Longer alternative paths

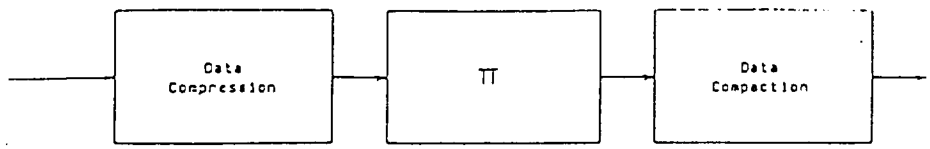

Figure 3. Proposed Joint Source/Channel Coder

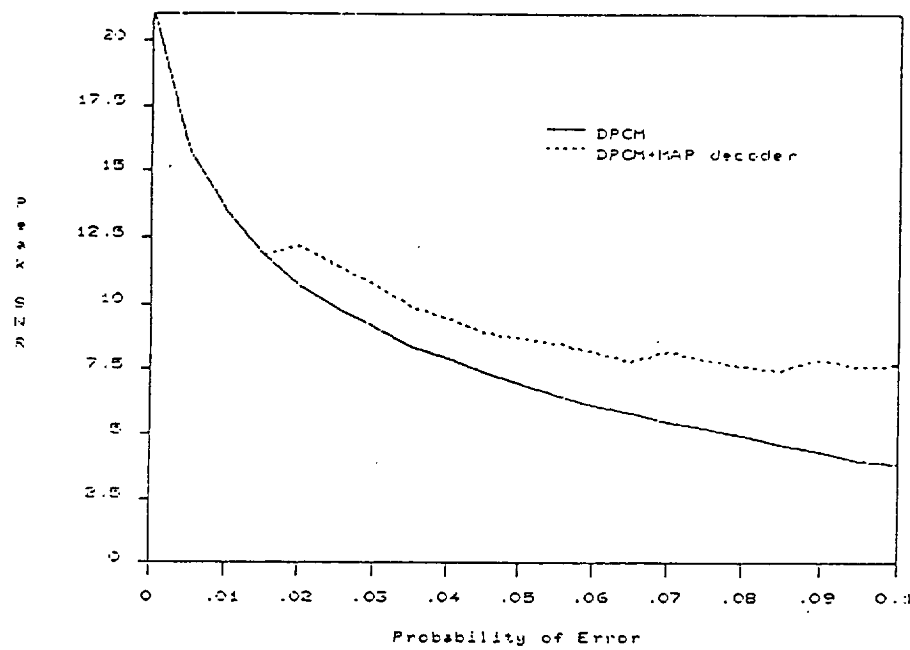

Figure 4. Performance of two-bit DPCM with MAP decoder

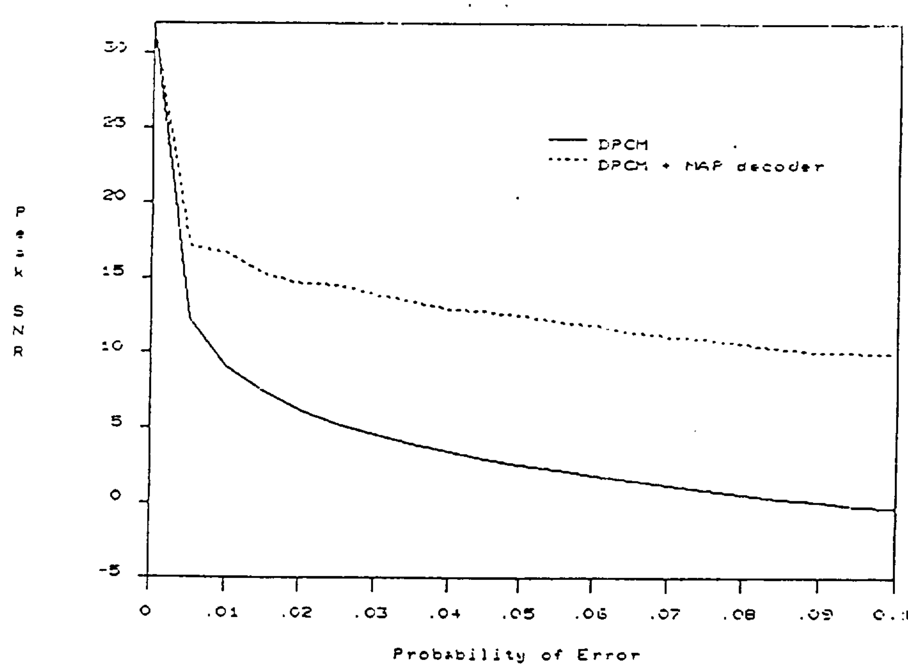

Figure 5. Performance of four-bit DPCM with MAP decoder

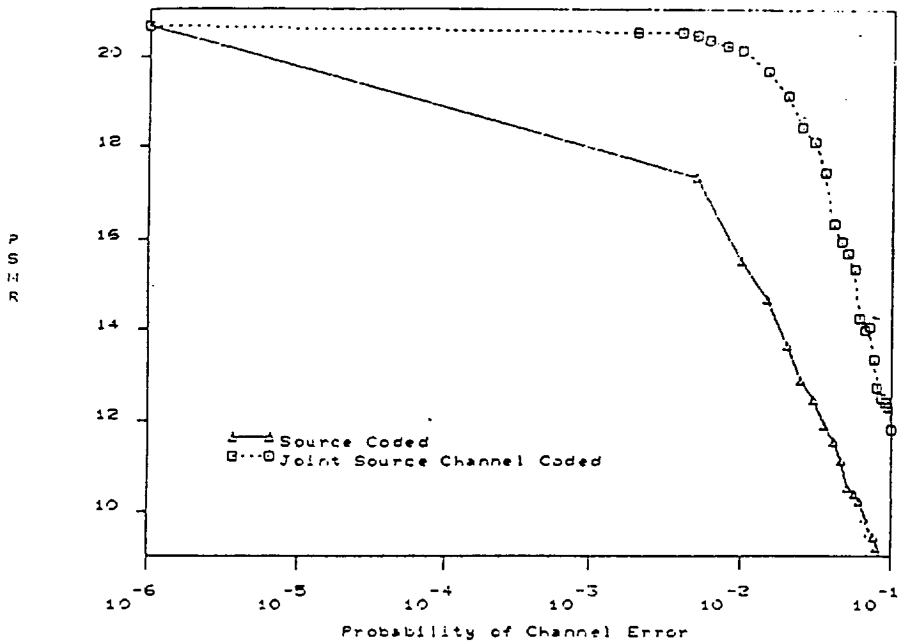

Figure 6. Performance Results of Joint Source/Channel Coding Scheme

Appendix 2- Item 3

IMPLEMENTATION ISSUES IN MAP JOINT SOURCE/CHANNEL CODING

Khalid Sayood* , Jerry D. Gibson@ and Fuling Liu*

*Department of Electrical Engineering
University of Nebraska
Lincoln, NE 68588-0511

@Department of Electrical Engineering
Texas A&M University
College Station, TX 77843

ABSTRACT

One of Shannon's many fundamental contributions was his result that source coding and channel coding can be implemented separately without any loss of optimality. However, the assumption underlying this result may at times be violated in practice. Various joint source/channel coding approaches have been developed for handling such situations. A MAP approach to joint source/channel coding has been proposed which uses a MAP decoder and a modification of the source coder to provide error correction. We present various implementation strategies for this approach and provide results for an image coding application.

I. Introduction

One of Shannon's many fundamental contributions was his result that source coding and channel coding can be treated separately without any loss of performance as compared to an optimum system [1]. The basic design procedure implied by Shannon's theorems consists of designing a source encoder which changes the source sequence into a series of (approximately) independent, equally likely binary digits followed by a channel encoder which accepts binary digits and puts them into a form suitable for reliable transmission over the channel 9 [2]. One aspect of the overall optimum system not addressed by Shannon is any increase in system complexity that results from this separation, and Massey [3] and Ancheta [4] showed that for distortionless transmission of the source under the constraint of linear source and channel coders, a significant reduction in complexity with equivalent performance can be achieved by using a linear joint source/ channel coder. Their scheme also differs from most data compression systems in that the bulk of the system complexity is transferred to the receiver.

The theorem that provides the justification for the separate design of the source coder and the channel coder, often called the Information Transmission Theorem [2], assumes that both the source encoder/decoder pair and the channel encoder/decoder pair are operating in an optimal fashion. Specifically, the source encoder is assumed to present the channel encoder with a sequence suitable for optimal channel coding, and the channel encoder/ decoder pair is assumed to reproduce the source encoder output at the source decoder input with negligible distortion. Unfortunately, there are practical situations where these assumptions are violated--namely, when the source encoder output contains redundancy, which occurs if the source encoder is suboptimal, and when the source decoder input differs from the source encoder output, which is a result of channel errors. These two situations are common occurrences in practical communication systems where source and/or channel models are imperfectly known, complexity is a serious issue, or significant delay is not tolerable. Various approaches have been developed to handle these two situations. These include approaches in which the source and channel coding operations are truly integrated [3-6], approaches that cascade known source coders with known channel coders and allocate the fixed bit rate to the source coder and channel coder to maximize system performance [7-15], approaches in which the source coder and/or receiver is modified to account for the presence of a given noisy channel [16-26], and approaches which use some knowledge of the source and source coder properties to detect channel errors and compensate for their effects [27-35]. The research described in this paper is concerned with the implementation of a joint source/channel coder design which was an extension of the work presented in [31,32]. This approach utilizes structure in the source encoder output by using a MAP decoder to correct errors introduced by the channel.

*Supported by NASA Lewis Research Center (NAG-3-806) and NASA Goddard Space Flight Center (NAG 5-916).

II. Previous Work

Based on the MAP design criteria, a decoder structure was proposed in [32] which takes advantage of redundancy in the channel input sequence to provide error correction. The decoder maximizes the quantity log $P(Y \mid \hat{Y})$ where

$$Y = (y_1, y_2, \ldots, y_L)$$

is the channel input sequence while

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_L)$$

is the channel output sequence. If a Markov model is imposed on the channel input sequence, the path metric can be written as

$$\log P(Y \mid \hat{Y}) = \Sigma \log P(y_i \mid \hat{y}_i, y_{i-1}) \tag{1}$$

and

$$P(y_i = a_j \mid y_{i-1} = a_m, \hat{y}_i = a_n) \tag{2}$$

$$= \frac{P(\hat{y}_i = a_n \mid y_i = a_j) P(y_i = a_j \mid y_{i-1} = a_m)}{\Sigma_l \, P(\hat{y}_i = a_n \mid y_i = a_l) P(y_i = a_l \mid y_{i-1} = a_m)}$$

The proof of the above can be found in [31,32].

Based on analysis of the decoder a parameter called the error correction capability was defined in [35] as

$$I = 1 - H(y_n \mid y_{n-1}) / \log M \tag{3}$$

We noted that a desirable property of a joint source channel coder would be to increase I. The approach proposed for this requires the modification of the source coder. In general, a source coder consists of two operations, data compression and data compaction [36]. The data compression operation usually consists of redundancy removal and involves some loss of information. Examples of data compression schemes are DPCM, transform coding and vector quantization. The data compaction schemes are information preserving. They may result in a variable rate out. Examples include Huffman coding and runlength coding. Generally in discussions of joint source/channel coder design, the data compaction operations are not included. The reason for this is that due to the variable rate output, the data compaction schemes are highly vulnerable to channel noise and, therefore, are not considered for noisy channel applications.

A possible approach to achieving the objective of increasing I is to insert an invertible transoperation between the data compression and data compaction stages. An example of such an operation called the π operator, was presented in [35]. The operation can be described as follows. Let the input to the π operator be selected from the alphabet

$$S = \{s_0, s_1, s_2, \ldots, s_{N-1}\},$$

and let the output alphabet be denoted by

$$A = \{a_0, a_1, a_2, \ldots, a_{N^2-1}\}.$$

Then the input/output mapping is given by

$$x_n = s_i, \; x_{n-1} = s_j \rightarrow y_n = a_{jN+i}.$$

This operator and its effects are described in more detail in [35].

While this approach achieves the objective of increasing the error correcting capability, it also results in a variable rate system. For this situation the branch metric of the form of (2) becomes difficult to implement. We explain these difficulties and propose implementatable approximations of the metrics in the next section. Section V contains simulation results which demonstrate the viability of these approximations. The use of a variable rate coder also complicates the structure of the decoder. In Section IV we present a modified Viterbi decoder which can be used with variable rate codes.

III. Development of the Path Metric

Before we begin our discussion of the path metric for variable rate case we need to summarize the derivation of (2). The derivation consists of two steps. First we show that

$$P(y_i = a_j \mid y_{i-1} = a_m, \hat{y}_i = a_n) \tag{4}$$

$$= \frac{P(\hat{y}_i = a_n \mid y_i = a_j)P(y_i \mid y_{i-1})}{P(y_i = a_n \mid y_{i-1} = a_m)}$$

Then we show that the denominator can be written as

$$P(\hat{y}_i = a_n \mid y_{i-1} = a_m) = \Sigma_l \, P(\hat{y}_i = a_n \mid y_i = a_l)P(y_i = a_l \mid y_{i-1} = a_m) \tag{5}$$

Note first that in this derivation the channel input alphabet and output alphabet are the same. We have assumed hard decision at the output of the channel and for a fixed rate coder this translates into identical alphabets at the input and output of the channel. For the case where we have variable rate codes there is a subtle difference. In fact, there are two different ways in which we can view the output of the channel. The first approach is to assume that there is a Huffman decoder at the output of the channel. The Huffman decoder output alphabet is the same as the joint source/channel (JSC) coder output alphabet. Thus the branch metric as derived in (2) can be used directly. However, now the computation of the individual factors of the branch metric becomes somewhat more involved. Specifically, consider the calculation of $P(\hat{y}_i = a_n \mid y_i = a_j)$, where the channel is assumed to be a binary symmetric channel with known crossover probability p. Let $l(a_j)$ be the number of bits in the binary codeword corresponding to the symbol a_i.

If $l(a_n) = l(a_j)$, as is the case when a fixed rate code is used, then

$$P(\hat{y}_i = a_n \mid y_i = a_j) = p^{d_{nj}}(1 - p)^{n-d_{nj}} \tag{6}$$

where d_{nj} is the Hamming distance between the binary codewords corresponding to a_n and a_j and n is the number of bits in each of the codewords. However, when $l(a_n) \neq l(a_j)$, the calculation may not be as simple. To see this we need to introduce some more notation. Let the codeword corresponding to a_n be represented by

$$a_n = (a_{n_1}, a_{n_2}, \ldots, a_{n_{l(a_n)}})$$

Then, if $l(a_n)$ is less than $l(a_j)$

$$P(\hat{y}_i = a_n \mid y_i = a_j) = \Pi_{k=1}^{l(a_n)} Pr(a_{n_k} \mid a_{j_k}) \tag{7}$$

and the calculation is still relatively straightforward. However, if $l(a_n)$ is greater than $l(a_j)$,

$$P(\hat{y}_i = a_n \mid y_i = a_j) = \Sigma_l \, P(\hat{y}_i = a_n, y_{i+1} = a_l \mid y_i = a_j) \tag{8}$$

or in more familiar terms

$$P(\hat{y}_i = a_n \mid y_i = a_j) = \Sigma_l \, P(\hat{y}_i = a_n \mid y_i = a_j, y_{i+1} = a_l)P(y_{i+1} = a_l \mid y_i = a_n) \tag{9}$$

where we have used the chain rule and the Markov property of JSC coder output. The second factor in the summand is simply the transition probability of the JSC coder output while the first factor can be calculated as

$$P(\hat{y}_i = a_n \mid y_i = a_j, y_{i+1} = a_l)$$

$$= \Pi_{k=1}^{l(a_j)} Pr(a_{n_k} \mid a_{j_k}) \Pi_{k=l(a_j)+1}^{l(a_n)} Pr(a_{n_k} \mid a_{l_{k-l(a_j)}})$$

as long as $l(a_n)$ is less than or equal to $l(a_j) + l(a_l)$. If not we simply repeat the process again to obtain

$$P(\hat{y}_i = a_n \mid y_i = a_j, y_{i+1} = a_l) = \Sigma_h \, P(\hat{y}_i = a_n, y_{i+2} = a_h \mid y_i = a_j, y_{i+1} = a_l) = \Sigma_h \, P(\hat{y}_i = a_n \mid y_i = a_j, y_{i+1} = a_l, y_{i+2} = a_h)P(y_{i+2} = a_h \mid y_{i+1} = a_l) \tag{10}$$

Again $l(a_m)$ should be less than $l(a_j) + l(a_l) + l(a_h)$.

Obviously this process can continue if there is a large variation in the codeword lengths. Therefore, this approach becomes cumbersome for moderately large codebooks.

A somewhat different way of looking at this issue, suggested in a slightly different context by Massey [37], is to block the channel output bit stream into fixed length words where the fixed length is longer than the longest binary codeword in the channel input. Then, the path metric becomes the logarithm of

$$\frac{P(\hat{y}_i = \hat{r} \mid y_i)P(y_i \mid y_{i-1})}{\Sigma_l \, P(\hat{y}_i = \hat{r} \mid y_i = a_l)P(y_i = a_l \mid y_{i-1})} \tag{11}$$

where $\hat{r}$ denotes the word corresponding to a received block of bits. While there are some complications here as well, in the interpretation of $P(\hat{y}_i \mid y_i)$, the main difficulty is a computational one. The simplest implementation of the JSC decoder requires that the path metrics be stored in a lookup table. In the case of identical input and output alphabets of size M, the lookup table size is M^3. However, with this approach, the lookup table size is $M^2 2^{L+1}$ where L is the longest codeword. This exponential increase with even moderate codeword lengths makes this approach impractical at least for a lookup table implementation. An implementation which does not use a lookup table, and instead computes the path metric at each step may still be possible with special purpose dedicated hardware.

Given the difficulties involved with implementation of the exact path metric of the MAP JSC decoder, we have proposed two approximations which provide a high level of error protection while being computationally simple and easy to implement. First consider (4). We approximate the denominator as

$$P(\hat{y}_i = \hat{r} \mid y_{i-1}) \cong P(\hat{y}_i = \hat{r}).$$

and therefore the entire expression as

$$P(y_i = a_j \mid \hat{y}_i = \hat{r}, y_{i-1} = a_m) \qquad (12)$$

$$\cong \frac{P(\hat{y}_i = \hat{r} \mid y_i = a_j) P(y_i = a_j \mid y_{i-1} = a_m)}{P(\hat{y}_i = \hat{r})}$$

where the number of bits in $\hat{r}$ is the number of bits used to represent a_j. The denominator is further approximated by assuming equally likely reception of bits as

$$P(\hat{y}_i = \hat{r}) = (\tfrac{1}{2})^{l(a_j)} \qquad (13)$$

where $l(a_j)$ is the number of bits in a_j and therefore in $\hat{r}$. The computation of the path metric then proceeds as follows: the conditional probability $P(y_i = a_j \mid y_{i-1} = a_{j-1})$ is read from a lookup table and the transition probability is computed by assuming a binary symmetric channel with known crossover probability. This form of the path metric is easy to implement and the simulation results of Section V show the scheme to be highly effective.

An even simpler approximation is to use the Hamming distance between the received bits and the candidate sequence elements as the branch and path metric. Of course the candidate sequence elements are selected from allowed sequence values. (Recall that the π operator, by construction, disallows certain sequences.) We present results using this metric in Section V. This approximation causes a drop in performance from about a half dB in the low noise region to about 1.5 to 2 dB in the high noise region. Given the simplicity of implementation for this scheme, this may very well be an acceptable cost.

Once the path metric has been obtained, the decoder structure needs to be elucidated. We do so in the next section.

IV. Decoder Structure

The form of the path metric in (1) is a familiar one and several decoder structures exist which maximize (or minimize) additive path metrics of this form. One of the most popular ones is the Viterbi decoder structure. Recall that the Viterbi decoder limits the total number of candidate paths (solutions) to some finite number M where M is the number of different values a solution can take at any given time increment. This is done by using a trellis structure that only includes allowed paths or transitions. For the problem considered here M would be the size of the output alphabet of the π operator. In most applications where the Viterbi decoder is used, the codewords are of fixed length and therefore the candidate paths are of the same length. This is not true in the current case. However, this problem can be resolved rather simply by associating a pointer with each candidate path. The pointer counts the number of bits used to form the path it is associated with.

To see how this works consider the following example. Let the input alphabet to the π operator be of size two; $S = (s_0, s_1)$. Suppose the input sequence to the π operator is

$$s_0\ s_0\ s_0\ s_1\ s_0\ s_0$$

then the output of the π operator will be

$$a_0\ a_0\ a_1\ a_2\ a_0$$

If the Huffman code for the π operator output is

$$a_0{:}0,\ a_1{:}10,\ a_2{:}110,\ a_3{:}111$$

then the transmitted binary sequence will be

0 0 1 0 1 1 0 0

Suppose there is an error in the fourth bit and the received sequence is

0 0 1 1 1 1 0 0

The decoder operation is shown in Figure 1, where the metric being used is the Hamming distance. The branches are labelled with a pair of numbers. The first number is the accumulated number of bits used by the path that includes that branch while the second number is the Hamming distance between the received bits and the candidate solution. The receiver assumes a starting value of a_0. In the first step there are two possibilities, that the transmitted word was a_0 or a_1. If we assume the transmitted word was a_0 we use up one bit and the Hamming distance is zero. If a_1 is assumed then we use two bits and the Hamming distance is one. Therefore, the lower branch (to a_0) is labelled 1,0 while the branch to a_1 is labelled 2,1. This procedure is continued with conflicts being resolved by picking the path with the lower Hamming distance. The procedure is shown in Figure 1.

V. Simulation Results

The techniques presented in this paper were applied to an image coding scheme. The data compression scheme was a DPCM system with a fixed four level nonuniform Max quantizer and a one-tap predictor. The data compaction scheme is a sixteen-level Huffman coder. The average rate for this system was 2.3 bits per pixel. End of line resynchronization is assumed for the receiver. A block diagram of the system is shown in Figure 2.

The performance with both metrics is shown in Figure 3 and Figure 4. Both figures plot the same results where Figure 3 emphasizes the performance in the low noise region and Figure 4 emphasizes performance at high channel error rates. The curves are labeled "Approx 1," "Approx 2," and "No Protection." The curve labeled Approx 1 is the performance curve for the system which uses the metric approximation of (12) and (13). The curve labeled Approx 2 is the system which uses the second approximation, i.e., the Hamming distance between the received bits and the candidate sequence elements. The curve labeled "No Protection" is the system without the joint source/channel coding scheme. Both metric approximations provide a high degree of protection for low to moderate channel error rates. At high channel error rates, while both the systems provide substantial performance improvements over the unprotected system, the system with the Hamming distance metric provides lower performance than the system with the approximation of (12) and (13). However, as mentioned before, this might be a small cost to pay for the simplicity of implementation.

REFERENCES

[1] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-432, July 1948.

[2] T. Berger, *Rate Distortion Theory*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[3] J.L. Massey, "Joint Source and Channel Coding," in *Communication Systems and Random Process Theory*, J.K. Skwirzynski, ed., Sijthoff and Nordhoff: Netherlands, pp. 279-293, 1978.

[4] T.C. Ancheta, Jr., "Bounds and Techniques for Linear Source Coding," Ph.D. dissertation, Dept. of Elec. Eng., Univ. of Notre Dame, Aug. 1977.

[5] J.G. Dunham and R.M. Gray, "Joint source and channel trellis encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 516-519, July 1981.

[6] E. Ayanoglu and R.M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 855-865, Nov. 1987.

[7] J.W. Modestino and D.G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. COM-27, pp. 1644-1659, Nov. 1979.

[8] J.W. Modestino, D.G. Daut and A.L. Vickers, "Combined source-channel coding of images using the block cosine transform," *IEEE Trans. Commun.*, vol. COM-29, pp. 1262-1274, Sept. 1981.

[9] J.W. Modestino, V. Bhaskaran, and J.B. Anderson, "Tree encoding of images in the presence of channel errors," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 677-697.

[10] D. Comstock and J.D. Gibson, "Hamming coding of DCT compressed images over noisy channels," *IEEE Trans. Commun.*, vol. COM-32, pp. 856-861, July 1984.

[11] C.C. Moore and J.D. Gibson, "Self-orthogonal convolutional coding for the DPCM-AQB speech encoder," *IEEE Trans. Commun.*, vol. COM-32, pp. 980-982, Aug. 1984.

[12] R.C. Reininger, "Backward adaptive lattice and transversal predictors in ADPCM," *IEEE Trans. Commun.*, vol. COM-33, pp. 74-82, Jan. 1985.

[13] D.J. Goodman and C.E. Sundberg, "Combined source and channel coding for variable bit rate speech transmission," *Bell Syst. Tech. J.*, vol. 62, pp. 2017-2036, Sept. 1983.

[14] D.J. Goodman and C.E. Sundberg, "Transmission errors and forward error correction in embedded differential pulse code modulation," *Bell Syst. Tech. J.*, vol. 62, pp. 2735-2764, Nov. 1983.

[15] J.D. Gibson and R.C. Reininger, "Forward and Backward Prediction in ADPCM Over Nonideal Channels," *Conf. Rec.*, Int. Conf. Commun., June 8-12, 1980, Seattle, WA, pp. 42.5.1-42.5.5.

[16] A.J. Kurtenbach and P.A. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 291-302, April 1969.

[17] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827-838, Nov. 1987.

[18] V. Vaishampayan and N. Farvardin, "Optimal block cosine transform image coding for noisy channels," submitted to *IEEE Trans. Commun.*

[19] H. Kumazawa, M. Kashahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electronics and Engineering in Japan*, vol. 67-B, no. 4, pp. 39-47, 1984.

[20] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," Tech. Res. Rpt., SRC TR 88-15, Univ. of Maryland, College Park, MD.

[21] V. Vaishampayan and N. Farvardin, "The Joint Design of Block Source Coders and Modulation Signal Sets," Twenty-Second Annual Conf. of Information Sciences and Systems, March 16-18, 1988, Princeton Univ., Princeton, NJ.

[22] K-Y Chang and R.W. Donaldson, "Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels," *IEEE Trans. Commun.*, vol. COM-20, pp. 338-350, June 1972.

[23] N. Rydbeck and C.W. Sundberg, "Analysis of digital errors in nonlinear PCM systems," *IEEE Trans. Commun.*, vol. COM-24, pp. 59-65, Jan. 1976.

[24] J.R.B. DeMarca and N.S. Jayant, "An Algorithm for Assigning Binary Indices to the Codevectors of a Multi-dimensional Quantizer," *Proc.* IEEE Int. Commun. Conf., Seattle, WA, pp. 1128-1132, June 1987.

[25] J.-H. Chen, G. Davidson, A. Gersho and K. Zeger, "Speech Coding for the Mobile Satellite Experiment," *Proc.*, IEEE Int. Comm. Conf., Seattle, WA, pp. 756-763, June 1987.

[26] K.A. Zeger and A. Gersho, "Zero redundancy channel coding in vector quantization," IEEE Electronics Letters, vol. 23, pp. 654-655, June 1987.

[27] R. Steele, D.J. Goodman, and C.A. McGonegal, "Partial correction of transmission errors in DPCM without recourse to error correction coding," *Elec. Lett.*, vol. 13, pp. 351-353, June 1977.

[28] R. Steele, D.J. Goodman, and C.A. McGonegal, "A difference detection and correction scheme for combatting DPCM transmission errors," *IEEE Trans. Commun.*, vol. COM-27, pp. 252-255, Jan. 1979.

[29] K.N. Ngan and R. Steele, "Enhancement of PCM and DPCM images corrupted by transmission errors," *IEEE Trans. Commun.*, vol. COM-30, pp. 257-269, Jan. 1982.

[30] R.C. Reininger and J.D. Gibson, "Soft decision demodulation and transform coding of images," *IEEE Trans. Commun.*, vol. COM-31, pp. 572-577, April 1983.

[31] K. Sayood and J.C. Borkenhagen, "Utilization of Correlation in Low Rate DPCM Systems for Channel Error Protection," *Proceedings*, IEEE International Conference on Communications, June 1986, pp. 1888-1892.

[32] K. Sayood and J.C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," submitted to *IEEE Transactions on Communications*.

[33] M.E. Hellman, "On using natural redundancy for error detection," *IEEE Trans. Commun.*, vol. COM-22, pp. 1690-1693, Oct. 1974.

[34] M.E. Hellman, "Convolutional source encoding," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 651-656, Nov. 1975.

[35] K. Sayood and J.D. Gibson, "Maximum a posteriori joint source/channel coding," Twenty-Second Annual Conf. on Information Sciences and Systems, March 16-18, Princeton Univ., Princeton, NJ.

[36] R.E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, 1987.

[37] J.L. Massey, "Variable Length Codes and the Fano Metric," *IEEE Trans. Inform. Theory*, Jan. 1972.

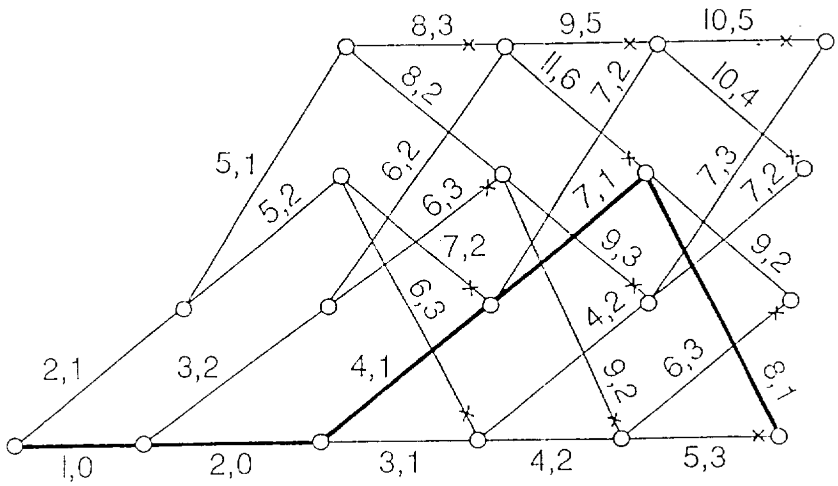

Figure 1. Decoding procedure for variable length codes

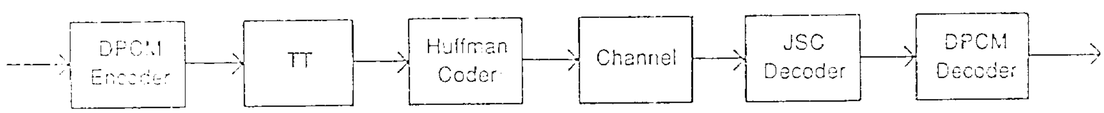

Figure 2. Proposed system

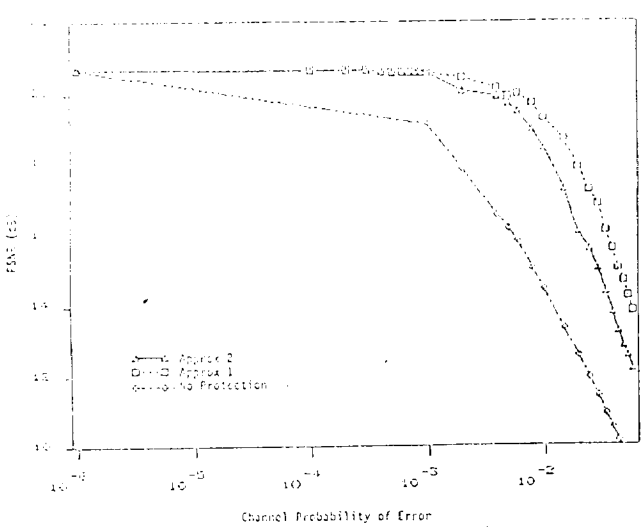

Figure 3. Comparison of performance with different decoder metrics

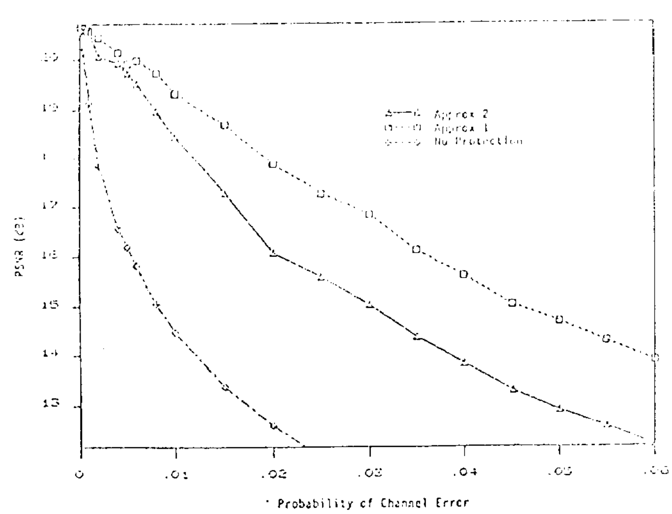

Figure 4. Comparison of performance with different decoder metrics

Appendix 2- Item 4