Then from (22)

$$P_e(\hat{R}_1) \geq 1 - \frac{1}{M_1} \sum_Y \sum_{X_2} Q_2(X_2) \cdot \left\{ \sum_{X_1} M_1 Q_1(X_1) P(Y|X_1, X_2)^{1/\beta_1} \right\}^{\beta_1}$$
$$= 1 - (M_1)^{\beta_1 - 1} \sum_Y \sum_{X_2} Q_2(X_2) \cdot \left\{ \sum_{X_1} Q_1(X_1) P(Y|X_1, X_2)^{1/\beta_1} \right\}^{\beta_1}$$
$$\geq 1 - (M_1)^{\beta_1 - 1} \max_{Q_1, Q_2} \sum_Y \sum_{X_2} Q_2(X_2) \cdot \left\{ \sum_{X_1} Q_1(X_1) P(Y|X_1, X_2)^{1/\beta_1} \right\}^{\beta_1}. \quad (24)$$

Putting $\beta_1 = 1 + \rho_1$, we have

$$P_e(\hat{R}_1) \geq 1 - (M_1)^{\rho_1} \max_{Q_1, Q_2} \sum_Y \sum_{X_2} Q_2(X_2) \cdot \left\{ \sum_{X_1} Q_1(X_1) P(Y|X_1, X_2)^{1/(1+\rho_1)} \right\}^{1+\rho_1}. \quad (25)$$

Lemma 3: For any discrete memoryless channel, we have

$$\max_{\substack{Q_1(X_1)\\ Q_2(X_2)}} \sum_{Y \in \mathcal{Y}_N} \sum_{X_2 \in \mathcal{X}_{2N}} Q_2(X_2) \cdot \left\{ \sum_{X_1 \in \mathcal{X}_{1N}} Q_1(X_1) P(Y|X_1, X_2)^{1/(1+\rho_1)} \right\}^{1+\rho_1}$$
$$= \left[\max_{\substack{Q_1(x_1)\\ Q_2(x_2)}} \left(\sum_{y \in B} \sum_{x_2 \in A_2} Q_2(x_2) \cdot \left\{ \sum_{x_1 \in A_1} Q_1(x_1) P(y|x_1, x_2)^{1/(1+\rho_1)} \right\}^{1+\rho_1} \right) \right]^N,$$
$$0 > \rho_1 > -1. \quad (26)$$

Proof (Omitted): From (25) and (26) we have

$$P_e(\hat{R}_1) \geq 1 - (M_1)^{\rho_1} \left\{ \max_{Q_1, Q_2} \left(\sum_i \sum_l Q_2(l) \cdot \left[\sum_k Q_1(k) P(i|k, l)^{1/(1+\rho_1)} \right]^{1+\rho_1} \right)^N \right\}, \quad 0 > \rho_1 > -1. \quad (27)$$

Let $M_1 = \lceil e^{NR_1} \rceil$; then inequality (27) can be put in the form

$$P_e(\hat{R}_1) \geq 1 - \exp\left\{ -N\left[-\rho_1 \hat{R}_1 + \min_{Q_1, Q_2} E_1(\rho_1, Q) \right] \right\}, \quad 0 > \rho_1 > -1. \quad (28)$$

Thus, similarly to Case 1, we have established the following theorem.

Theorem 3: For any discrete memoryless channel, if $\hat{R}_1 > \max_{Q_1, Q_2} I(Q, P)$, the probability of decoding error $P_e(\hat{R}_1)$ approaches unity as N tends to infinity.

Case 3: With side information X_1 available at the decoder with probability $Q_1(X_1)$, the output Y is decoded into the integer k if

$$P(Y|X_1, X_{2k}) > P(Y|X_1, X_{2k'}), \qquad 1 \leq k' \leq M_2, \ k' \neq k. \quad (29)$$

In analogy to Case 2, we deduce that

$$P_e(\hat{R}_2) \geq 1 - \exp\left\{ -N\left[-\rho_2 \hat{R}_2 + \min_{Q_1, Q_2} E_2(\rho_2, Q) \right] \right\}, \quad 0 > \rho_2 > -1, \quad (30)$$

which approaches unity as N tends to infinity for $\hat{R}_2 > \max_{Q_1, Q_2} I_2(Q, P)$. Combining all three cases yields Theorem 1.

Acknowledgment

The author wishes to acknowledge a reviewer for his helpful comments, especially for (11)–(13).

References

[1] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channel with correlated sources," *Bell. Syst. Tech. J*, vol. 52, pp. 1037–1076, 1973.

[2] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, 1972.

[3] E. C. van der Meulen, "Random coding theorem for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, 1975.

[4] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 357–359, 1973.

[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

The Capacity of the Physically Degraded Gaussian Broadcast Channel with Feedback

ABBAS EL GAMAL

Abstract—**Bounds on the output entropy of the additive white Gaussian noise (AWGN) channel with feedback are used to prove that the capacity of the degraded additive white Gaussian noise (DAWGN) broadcast channel is not increased by feedback.**

I. Introduction

The degraded broadcast channel is perhaps the best understood multiuser channel. A coding theorem and a converse were established both for the discrete memoryless case and for the additive white Gaussian noise (AWGN), case with an average power constraint [6]. In [1] it was shown that feedback cannot increase the capacity of the discrete memoryless degraded broadcast channel. We will show that the result in [1] is also true for the physically degraded additive white Gaussian noise broadcast channel. We first modify Shannon's entropy inequalities [2] to get upper and lower bounds on the entropy of the output of the AWGN channel with feedback. The new bounds combined with [6] and [1] are then used to prove a weak converse for the

Manuscript received May 31, 1977; revised March 31, 1980. This paper was supported in part by the National Science Foundation under Grant ENG 76-03684.

The author is with the Department of Electrical Engineering, Stanford University, Durand Building, ISL, Stanford, CA 94306.

0018-9448/81/0700-0508$00.75 ©1981 IEEE

degraded additive white Gaussian noise (DAWGN) broadcast channel with feedback.

II. Bounds on the Output Entropy

Stam [3] and Blachman [4] have proved Shannon's statement [2] that if $\boldsymbol{X}$ and $\boldsymbol{Z}$ are two independent random vectors of length n with differentiable n-dimensional density functions and if $\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{Z}$, then

$$e^{2H(\boldsymbol{Y})/n} \geq e^{2H(\boldsymbol{X})/n} + e^{2H(\boldsymbol{Z})/n}, \tag{1}$$

where $H(\boldsymbol{X})$ is the differential entropy, i.e.,

$$H(\boldsymbol{X}) \triangleq -\int f_{\boldsymbol{X}}(\boldsymbol{x}) \ln f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{2}$$

where $f_{\boldsymbol{X}}(\boldsymbol{x})$ is the n-dimensional density function of $\boldsymbol{X}$. If we let $\boldsymbol{X}$ be the input of an AWGN channel and $\boldsymbol{Z}$ the noise vector, then (2) can be applied to get

$$e^{2H(\boldsymbol{Y})/n} \geq e^{2H(\boldsymbol{X})/n} + 2\pi e N, \tag{3}$$

where N is the variance of the noise. In the presence of feedback, each component of the $\boldsymbol{X}$ vector can depend causally on the noise; thus, in general, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are no longer independent. However, looking more closely at Blachman's proof of inequality (2), we can make the following modification.

Lemma 1 (Lower Bound): Let $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, NI)$ and $\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{Z}$, where for all $i=1,2,\cdots,n$, X_i and Z_i are conditionally independent given $Y^{i-1} \triangleq Y_1, Y_2, \cdots, Y_{i-1}$, Z_i and Y^{i-1} are independent, and X_i has differentiable density function conditional on every y^{i-1}. Then

$$e^{2H(\boldsymbol{Y})/n} \geq \exp\left[\frac{2}{n}\sum_{i=1}^{n} H\left(X_i | Y^{i-1}\right)\right] + 2\pi e N. \tag{4}$$

The proof of this lemma is given in the Appendix and follows Blachman [4]. Next, we give an upper bound on the output entropy $H(\boldsymbol{Y})$ of the AWGN channel with feedback.

Lemma 2 (Upper Bound): If

$$E\sum_{i=1}^{n} X_i^2 \leq nP, \tag{5}$$

then

$$H(\boldsymbol{Y}) \leq \frac{n}{2}\ln 2\pi e(P+N). \tag{6}$$

To prove Lemma 2 we need the following useful inequality.

Hadamard [5]*:* If $A \triangleq [a_{ij}]$ is any $n \times n$ positive definite matrix, then

$$|A| \leq \prod_{i=1}^{n} a_{ii}, \tag{7}$$

where A denotes the determinant of A. It is also well-known that [2]

$$H(\boldsymbol{Y}) \leq \frac{n}{2}\ln 2\pi e |K_{\boldsymbol{Y}}|^{1/n} \tag{8}$$

where $K_{\boldsymbol{Y}}$ is the covariance matrix of $\boldsymbol{Y}$. Hence, using (7)

$$|K_{\boldsymbol{Y}}| \leq \prod_{i=1}^{n}\left(EX_i^2 + 2EX_iZ_i + N\right). \tag{9}$$

But $EX_iZ_i = 0$ for all $i=1,2,\cdots,n$. Substituting in (8) we obtain

$$H(\boldsymbol{Y}) \leq \sum_{i=1}^{n} 1/2 \ln 2\pi e\left(EX_i^2 + N\right)$$
$$\leq \frac{n}{2}\ln 2\pi e(P+N),$$

where the last inequality follows from the concavity of the logarithm function and the constraint (5). Lemma 2 is proved.

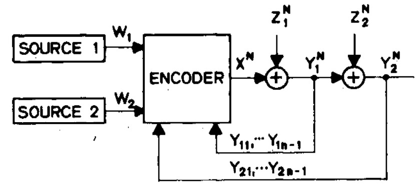

Fig. 1. Physically degraded AWGN broadcast channel.

III. A Converse for the Physically Degraded AWGN Broadcast Channel with Feedback

The model of the physically degraded AWGN broadcast channel with feedback is depicted in Fig. 1.

At the nth transmission the encoder maps the pair $\{W_1, W_2\}$ where $W_1 \in \{1,\cdots,M_1\}$, $W_2 \in \{1,\cdots,M_2\}$, and the past outputs $(Y_{11},\cdots,Y_{1n-1},Y_{21},\cdots,Y_{2n-1})$ into a real number X_n. Thus

$$X_n = f_n(W_1, W_2, Y_{11},\cdots,Y_{1n-1},Y_{21},\cdots,Y_{2n-1}),$$
$$Y_{1n} = X_n + Z_{1n}, \tag{10}$$

and

$$Y_{2n} = X_n + Z_{1n} + Z_{2n}, \qquad n=1,2,\cdots,N,$$

where $(Z_{11},\cdots,Z_{1N})$ is a sequence of independent identically distributed (i.i.d.) normal random variables (RV) with mean zero and variance N_1, and $(Z_{21},\cdots,Z_{2N})$ is an independent sequence of i.i.d. normal RV with mean zero and variance $(N_2 - N_1)$ where $N_2 > N_1$. The average power constraint on the input sequence $(x_1,\cdots,x_N)$ is given by

$$\frac{1}{N}\sum_{n=1}^{N} x_n^2 \leq P. \tag{11}$$

Bergmans [6] established that the capacity region of the AWGN broadcast channel (which includes the degraded class as a special case) is the set of all (R_1, R_2) such that

$$R_1 \leq 1/2\ln\left(1+\frac{\alpha P}{N_1}\right) \triangleq C_1(\alpha) \tag{12}$$

and

$$R_2 \leq 1/2\ln\left(1+\frac{\bar{\alpha}P}{N_2+\alpha P}\right) \triangleq C_2(\alpha), \qquad \alpha \in [0,1],$$

and $\bar{\alpha} = 1-\alpha$.

Following a similar approach we shall prove that in the physically degraded case the capacity region remains unchanged when noiseless feedback is added.

Theorem 3: No point (R_1, R_2) such that for some α,

$$R_i \geq C_i(\alpha), \qquad i=1,2, \tag{13a}$$
$$R_i = C_i(\alpha) + \delta, \qquad \text{some } i, \delta > 0, \tag{13b}$$

is achievable.

Proof (by contradiction): Assume for some α and some $\delta > 0$ that $(R_1, R_2) = (C_1(\alpha), C_2(\alpha)+\delta)$ is achievable, i.e., there exists a sequence of codes with decoding error probabilities $p_{e_1}^n, p_{e_2}^n \to 0$. Applying Fano's inequality, we have

$$H(W_i|\boldsymbol{Y}_i) \leq p_e^n \log M_i + 1 \triangleq n\lambda_{in}, \qquad i=1,2, \tag{14}$$

where

$$p_e^n = \max\left\{p_{e_1}^n, p_{e_2}^n\right\}.$$

Next we upper bound R_1 and R_2:

$$H(W_2) \triangleq nR_2 = I(W_2;\boldsymbol{Y}_2) + H(W_2|\boldsymbol{Y}_2)$$
$$\leq I(W_2;\boldsymbol{Y}_2) + n\lambda_{2n}$$
$$= H(\boldsymbol{Y}_2) - H(\boldsymbol{Y}_2|W_2) + n\lambda_{2n}.$$

Applying Lemma 2 to bound $H(\mathbf{Y}_2)$ we get

$$nR_2 \le \frac{n}{2}\ln 2\pi e(P+N_2) - H(\mathbf{Y}_2|W_2) + n\lambda_{2n}.$$

Therefore

$$H(\mathbf{Y}_2|W_2) \le \frac{n}{2}\ln 2\pi e(\alpha P+N_2) + n\lambda_{2n} - n\delta. \tag{15}$$

Also

$$\begin{aligned} H(W_1|W_2) &\triangleq nR_1 = I(W_1;\mathbf{Y}_1|W_2) + H(W_1|W_2,\mathbf{Y}_1) \\ &\le I(W_1;\mathbf{Y}_1|W_2) + n\lambda_{1n} \\ &\le I(W_1;\mathbf{Y}_1\mathbf{Y}_2|W_2) + n\lambda_{1n} \\ &= \sum_{j=1}^{n} I\left(W_1;Y_{1j}Y_{2j}|W_2Y_1^{j-1}Y_2^{j-1}\right) + n\lambda_{1n} \\ &= \sum_{j=1}^{n} I\left(W_1;Y_{1j}|W_2Y_1^{j-1}Y_2^{j-1}\right) \\ &\quad + \sum_{j=1}^{n} I\left(W_1;Y_{2j}|W_2Y_1^{j-1}Y_2^{j-1}Y_{1j}\right) + n\lambda_{1n} \\ &\le \sum_{j=1}^{n} H\left(Y_{1j}|W_2Y_1^{j-1}Y_2^{j-1}\right) - H\left(Y_{1j}|X_j\right) + n\lambda_{1n} \\ &\quad + \sum_{j=1}^{n} I\left(X_j;Y_{2j}|W_2Y_1^{j-1}Y_2^{j-1}Y_{1j}\right) \\ &= \sum_{j=1}^{n} H\left(Y_{1j}|W_2Y_1^{j-1}Y_2^{j-1}\right) - \frac{n}{2}\ln 2\pi eN_1 + n\lambda_{1n}, \end{aligned} \tag{16}$$

where for $1 \le j \le n$, $I(X_j;Y_{2j}|W_2Y_1^{j-1}Y_2^{j-1}Y_{1j})$ is equal to zero by the Markovity of the channel (see [1]). Hence

$$nR_1 \le \sum_{j=1}^{n} H\left(Y_{1j}|W_2Y_2^{j-1}\right) - \frac{n}{2}\ln 2\pi eN_1 + n\lambda_{1n}.$$

From hypothesis (13) we get

$$\sum_{j=1}^{n} H\left(Y_{1j}|W_2Y_2^{j-1}\right) \ge \frac{n}{2}\ln 2\pi e(N_1+\alpha P) - n\lambda_{1n}. \tag{17}$$

Now, we wish to use Lemma 1 to obtain a lower bound on $H(\mathbf{Y}_2|W_2)$. First, from the definition of the channel, and for all $1 \le j \le n$,

$$Y_{1j} = X_j + Z_{1j}.$$

Conditioning on any past sequence y_2^{j-1} and any message w_2, the density function $p(y_{1j}|y_2^{j-1},w_2)$ clearly *exists* and is *differentiable*. (It is the convolution of $p(z_{1j})$ and the conditional probability of X_j given (y_2^{j-1},w_2).) Therefore we can apply Lemma 1 to the sum $\mathbf{Y}_2 = \mathbf{Y}_1 + \mathbf{Z}_2$ to obtain

$$e^{2/n H(\mathbf{Y}_2|W_2)} \ge e^{2/n\sum_{j=1}^{n} H(Y_{1j}|W_2Y_2^{j-1})} + 2\pi e(N_2-N_1). \tag{18}$$

Combining (17) and (18), we obtain the lower bound

$$H(\mathbf{Y}_2|W_2) \ge \frac{n}{2}\ln 2\pi e\left[(N_1+\alpha P)e^{-2\lambda_{1n}} + (N_2-N_1)\right]. \tag{19}$$

Finally, combining (15) and (19) and letting $\lambda_{1n},\lambda_{2n} \to 0$ we get a contradiction; the same contradiction holds if (13b) is true for $i=1$, thus proving the theorem.

Remark: Wyner and Ziv [7] used a similar technique to prove a converse for the binary symmetric channel (BSC), replacing (2) by

$$H(\mathbf{X}) \ge n\nu \Rightarrow H(\mathbf{Y}) \ge h\left(p_0 * h^{-1}(\nu)\right), \tag{20}$$

where p_0 is the BSC parameter and $p_1 * p_2 = p_1(1-p_2) + p_2(1-p_1)$. Adding feedback to the BSC, (20) can be replaced by

$$\sum_{j=1}^{n} H\left(X_j|Y^{j-1}\right) \ge n\nu \Rightarrow h(\mathbf{Y}) \ge h\left(p_0 * h^{-1}(\nu)\right), \tag{21}$$

and a similar converse can be proved for the BSC with feedback. However, the result in [1] includes the degraded BSC broadcast channel with feedback as a special case. Generalizations of (21) to other discrete memoryless channels as in [8], [9] should be straightforward.

Remark: It has been recently shown [10] that feedback *can* increase the capacity of the nonphysically degraded AWGN broadcast channel.

Acknowledgment

The problem was suggested to the author by Professor Thomas Cover. The author is indebted to the referee for many valuable comments.

Appendix
Proof of Lemma 1

The claim is true for $n=1$ (see [3]). Assume it is true for $n=m$. Since X_{m+1} and Z_{m+1} are independent given Y^m, then

$$e^{2H(Y_{m+1}|Y^m=y^m)} \ge e^{2H(X_{m+1}|Y^m=y^m)} + 2\pi eN. \tag{22}$$

Taking the logarithm of both sides of (23), we get

$$2H(Y_{m+1}|Y^m=y^m) \ge \ln\left(e^{2H(X_{m+1}|Y^m=y^m)} + 2\pi eN\right).$$

Since $\ln(e^x+b)$ is convex downward in x if $b \ge 0$, averaging over y^m using Jensen's inequality gives

$$2H(Y_{m+1}|Y^m) \ge \ln\left(e^{2H(X_{m+1}|Y^m)} + 2\pi eN\right). \tag{23}$$

Now

$$H(Y^{m+1}) = H(Y^m) + H(Y_{m+1}|Y^m).$$

Hence

$$\frac{2}{m+1}H(Y^{m+1}) = \frac{m}{m+1}\left(\frac{2}{m}H(Y^m)\right) + \frac{1}{m+1}2H(Y_{m+1}|Y^m).$$

By the induction hypothesis and by (24)

$$\frac{2}{m+1}H(Y^{m+1}) \ge \frac{m}{m+1}\ln\left(e^{2/m\sum_{j=1}^{m}H(X_j|Y^{j-1})} + 2\pi eN\right) + \frac{1}{m+1}\ln\left(e^{2H(X_{m+1}|Y^m)} + 2\pi eN\right).$$

Again using Jensen's inequality, we get

$$\frac{2}{m+1}H(Y^{m+1}) \ge \ln\left(e^{2/m+1\sum_{j=1}^{m+1}H(X_j|Y^{j-1})} + 2\pi eN\right),$$

which proves Lemma 1.

References

[1] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 379–381, May 1978.
[2] C. E. Shannon, "A math. theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, Oct. 1948.
[3] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. Contr.*, vol. 2, pp. 101–112, 1959.
[4] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, Apr. 1965.
[5] R. Bellman, *Introduction to Matrix Analysis*, 2nd Edition. New York: McGraw-Hill, 1970.
[6] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, p. 279, Mar. 1974.
[7] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part 1," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, Nov. 1973.

[8] H. Witsenhausen, "Entropy inequalities for discrete channels," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 5, Sept. 1974.
[9] R. Ahlswede and J. Körner, "On the connection between the entropies of input and output distributions of discrete memoryless channels," *Braso Conf. on Probability Theory*, 1974.
[10] L. H. Ozarow, "Coding and capacity for additive white Gaussian noise multi-user channels with feedback," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1979.

Incremental Tree Coding of Speech

ANDREW C. GORIS AND JERRY D. GIBSON

***Abstract*—Tree coding of speech has been investigated by several workers. Virtually all of these investigations have involved incremental tree coding in that no matter how deep the tree is searched, only a single path map symbol is released at a time. As noted by Gray, even if a good long-term fit is found, the first step in the fit *may* be a poor one, thus yielding large sample distortions. Hence, it is important to stay on a path long enough to achieve the promised long-term distortion value. The relative frequency of path switching for the single symbol release rule is investigated for the (M, L) and truncated Viterbi tree search algorithms, various search depths, and different code generators. In addition, two multiple symbol release rules are investigated. One rule releases a fixed number of path symbols at a time, while the other rule releases a variable number of path symbols, the exact number depending on how many symbols are required for the average sample distortion to be less than or equal to the L-depth path average distortion. Speech sources are considered exclusively.**

I. Introduction

Tree source coding with a fidelity criterion requires the selection of a suitable distortion measure, code generator, and tree search algorithm. Additionally, and perhaps just as important as any of the above components, the system designer must specify a rule for releasing path map digits to the channel. Virtually all of the tree coders that have been investigated search the available tree to some depth, say L, but release only a single path map symbol at a time. Gray [1] calls this technique "incremental" encoding. As explained by Bodie [2], the logic behind this approach is that the best paths tend to stem from a single node not too far back in the tree. However, Gray [1] has pointed out the possibility that the first step in any good long-term fit *may* be a poor one, thus producing large sample distortions. Therefore, it is important to stay on a path long enough to achieve the fidelity promised by the L-depth search.

Two possible alternatives to the single symbol release rule, both motivated by [1], are to release a fixed number of path symbols, say $l \leq L$, of the best path in each L-depth search, or to release a variable number of path symbols, the exact number depending on how many symbols are required for the l-symbol distortion to be "close" to the best L-depth fit. These two approaches are studied in this paper via system simulations for speech sources and a differential pulse code modulation (DPCM) code tree with fixed prediction and adaptive quantization. The distortion measure is the unweighted mean squared error, and both the (M, L) algorithm [2], [3] and a truncated Viterbi algorithm are employed for tree searching.

Manuscript received March 27, 1980; revised August 21, 1980. This work was supported in part by the National Science Foundation under Grant ENG 77-28527 and was presented at the 1979 International Conference on Communications, Boston, MA, June 10–14. The parts of this work performed by A. C. Goris were submitted in partial fulfillment of the requirements for the M. S. degree in electrical engineering at Texas A & M.

A. C. Goris is with the Desktop Computer Division, Hewlett-Packard Company, Fort Collins, CO 80525.
J. D. Gibson is with the Department of Electrical Engineering, Texas A & M University, College Station, TX 77843.

II. Tree Coder Description

The code tree to be searched is generated by the DPCM structure in Fig. 1. In this figure

$$p(k) = s(k) - \hat{s}(k|k-1) \tag{1}$$

is the prediction error, which is quantized to yield the path map symbols

$$q(k) = Q[p(k)]. \tag{2}$$

The possible coder outputs are given by

$$\begin{aligned}\hat{s}(k) &= \hat{s}(k|k-1) + q(k)\\ &= P[\hat{S}(k-1)] + q(k),\end{aligned} \tag{3}$$

where $\hat{S}(k-1) = [\hat{s}(k-1)\hat{s}(k-2)\cdots\hat{s}(k-N)]^T$ is an N-vector of past coder outputs. For the present work, the prediction operator P is assumed to be linear of order N, hence

$$\hat{s}(k) = \sum_{i=1}^{N} a_i \hat{s}(k-i) + q(k), \tag{4}$$

where the $\{a_i, i = 1, 2, \cdots, N\}$ are called *predictor coefficients* or tap weights. These predictor coefficients are computed using one of two different methods. For one method, the coefficients are calculated based on McDonald's average speech data [4]. Experimental studies on the speech data used in this research indicated that predictors of order three or greater based on McDonald's average data produce lower signal-to-noise ratios (see Section III) than when $N = 2$.[1] As a result, a second-order predictor given by

$$a_1 = 1.515 \text{ and } a_2 = -0.752, \tag{5}$$

henceforth called the McDonald or generalized predictor, was selected for in-depth study.

For the second method, the predictor coefficients are obtained from autocorrelation terms computed over each complete utterance. Only fourth-order predictors are considered, and these coefficients, henceforth called matched predictors, are given in Table I. The four utterances used are described in the Appendix. While such an approach could never be used in an actual data compression system, frequent forward-adaptive updating of the coefficients can be employed [5]. The primary goal here is to determine the effects of generalized and matched code generators on the path switching problem.

As noted by a reviewer, experimental results [3], [5] indicate that cascading a smoother with the autocorrelation-matched code generator generally provides a subjective improvement in output speech quality, and that smoothing has a rate distortion theoretic basis as well [10, pp. 238–239]. For the present work, a smoother was not included since smoother design is somewhat ad hoc [3], and since it was felt that the smoother would not significantly alter path switching probabilities. This is only intuition, however, and the reader should keep in mind that a smoother is not used here.

The quantization indicated in (2) is performed by a four-level nonuniform quantizer with the step point and output level proportionality constants taken from Max [6] for a Gaussianly distributed input. The step size parameter adapts at each time instant k according to

$$\Delta(k) = \min\{1.6\Delta(k-1), 2000\} \tag{6a}$$

for an outer level at time $k-1$ or

$$\Delta(k) = \max\{0.8\Delta(k-1), 8\} \tag{6b}$$

[1]This is because fixed predictors computed from one set of data may be poorly matched to another set of speech data [8], [9].

0018-9448/81/0700-0511$00.75 ©1981 IEEE