TABLE I
VALUES OF PARAMETERS THAT SPECIFY THE POINT C_1 IN FIG. 2

s	q_0	y_0	y_0/x_0
.6	.13	.13	.54
.65	.02	.03	.14
.7	1.6×10^{-3}	3.0×10^{-3}	.017
.75	6.6×10^{-5}	1.5×10^{-4}	1.1×10^{-3}
.8	3.9×10^{-7}	1.1×10^{-6}	9.4×10^{-6}
.85	3.2×10^{-11}	1.1×10^{-10}	1.3×10^{-9}
.90	2.7×10^{-20}	1.1×10^{-19}	2.0×10^{-18}

curve $A'B_2'$ in the probability plane. For $s>0.5$, we can see by numerical calculation that the curve $A'B_1'$ is very close to the straight line $\overline{A'B_1'}$. Also, the boundary curve AB_1 of G_O, although convex outward, is well approximated by the straight line $\overline{AB_1}$.

Now let us examine condition (9) in order to compare G_J with G_O. Let us define

$$\begin{aligned} f(q_1,q_2) &= \xi_1-\xi_2 \\ &= h(sq_1+\bar{s}q_2)-q_1h(\bar{s}\bar{q}_2)-\bar{q}_1h(\bar{s}q_2) \\ &\quad -q_2h(\bar{s}\bar{q}_1)-\bar{q}_2(\bar{s}q_1)+(q_1*q_2)h(s). \end{aligned}$$

For $s=0.5, f(q_1,q_2)$ is nonpositive for all q_1,q_2. For $q_1=q_2=1/2$, $f(1/2,1/2)=1-2h(\bar{s}/2)+h(s)/2$ is an increasing function of s and becomes 0 for $s=s_0=0.5754$. For this value of s, condition (9) is satisfied only at A', B_1', and B_2', $\xi_1-\xi_2=f(q_1,q_2)$ is negative on the curve $A'B_2'$ (except at the end points), and $\eta_1-\eta_2=f(q_2,q_1)$ is negative on the curve $A'B_1'$. Thus the boundaries of G_J and G_O coincide only at A, B_1, and B_2 for $s=s_0$. Since the boundary AB_1 of G_O is well approximated by the straight line $\overline{AB_1}$, G_J will be a fairly good approximation to G even in this case.

When s exceeds s_0, the region in the probability plane for which $\xi_1<\xi_2$ moves towards the q_2 axis. This region and the region for which $\eta_1<\eta_2$ are shaded in Fig. 3. The points C_1' and C_2' in Fig. 3 are on the boundaries of the above regions, and the points C_1 and C_2 in Fig. 2 are generated by these probabilities, respectively. Thus the boundaries of G_J and G_O coincide with each other, and therefore with the boundary of G, except along B_1C_1 and B_2C_2. Therefore G_J will be a good approximation to G for $s>s_0$. If we connect B_1 and C_1 as well as B_2 and C_2 by a straight line, the resulting region will be a fairly good approximation to G even if we do not calculate G_J accurately.

We can further show that the segments B_1C_1 and B_2C_2 of the boundary curve of Fig. 2 become very short for large s. Let the q_2 coordinate of C_1' in Fig. 3 be q_0, and let the R_1 and R_2 coordinates of C_1 in Fig. 2 be x_0 and y_0. The values of q_0, y_0 (in bits) and y_0/x_0 are listed in Table I. For $s\geqslant 0.7$, q_0 and y_0 are calculated from the following relation obtained by expansion:

$$\log_2 q_0=\log_2(1/\bar{s})+1-F(s,q_m)/\bar{s}$$

$$\begin{aligned} F(s,q_m)=\Big[&s\log_2\{(1-\bar{s}q_m)/\bar{s}q_2\}-h(\bar{s}\bar{q}_m)+h(\bar{s}q_m) \\ &+q_m\bar{s}\log_2(s/\bar{s})+(1-2q_m)h(s)\Big]/\bar{q}_m \end{aligned}$$

$$y_0=q_1\big[\bar{q}_m\bar{s}\{1+\log_2(1/q_1)\}+(q_m-\bar{q}_ms)\log_2(1/s)\big].$$

The table shows that the B_1C_1 and B_2C_2 segments of the boundary diminish very rapidly for increasing s, and G_J is practically coincident with the capacity region G for s larger than 0.7.

We have shown that, for a particular discrete two-user channel, G_J is a good approximation to G for strong interference and is practically coincident with G for very strong interference.

REFERENCES

[1] R. Ahlswede, "Multi-way communication channels," *Second Int. Symp. Information Theory*, B. N. Petror and F. Csáki, Eds. Budapest: Akadémiai Kiadó, 1973, pp. 23–52.
[2] H. Sato, "Two-user communication channels," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 3, pp. 295–304, May 1977.
[3] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 5, pp. 569–570, Sept. 1975.
[4] ——, "On the capacity of multiple-terminal communication networks," Ph.D. dissertation, Dep. Elec. Eng., Stanford University, Stanford, CA, Aug. 1975.
[5] P. P. Bergmans, "The Gaussian interference channel," presented at the 1976 IEEE Int. Symp. Information Theory, Ronneby, Sweden, July 1976.
[6] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 1, pp. 1–37, Jan. 1977.
[7] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Annals of Probability*, vol. 2, no. 5, pp. 805–814, 1974.
[8] H. Liao, "Multiple access channels," Ph.D. Dissertation, Dep. Elec. Eng., University of Hawaii, Honolulu, HI, 1972.

The Feedback Capacity of Degraded Broadcast Channels

ABBAS EL GAMAL

Abstract—**The fact that the capacity region of the discrete memoryless physically degraded broadcast channel is not increased by feedback is established.**

I. INTRODUCTION

The capacity region of the discrete memoryless degraded broadcast channel (Cover [1]) was established in [2]–[4] and [6]. Bergmans [2] exhibited an achievable rate region. A converse for the binary symmetric broadcast channel was established by Wyner and Ziv [3]. Gallager [4] then proved a converse for the general discrete memoryless degraded broadcast channel. An alternative proof of the converse was given by Ahlswede [6]. Using methods similar to those in [4], it will be shown that the capacity region is unchanged by feedback when the degradation is physical.

II. PRELIMINARIES AND DEFINITIONS

The model under investigation is shown in Fig. 1.

There are two sources, the first producing an integer $W_1\in\mathfrak{M}_1=\{1,\cdots,M_1\}$, and the second an integer $W_2\in\mathfrak{M}_2=\{1,\cdots,M_2\}$. At the nth transmission the encoder maps the pair $\{W_1,W_2\}$ and the past outputs $\{Z_1,\cdots,Z_{n-1}\}$ and $\{Y_1,\cdots,Y_{n-1}\}$ into X_n. Thus

$$X_n=f_n(W_1,W_2,Z_1,\cdots,Z_{n-1},Y_1,\cdots,Y_{n-1}),\qquad n=1,2,\cdots,N. \tag{1}$$

The channel consists of three finite alphabets $x\in\mathfrak{X}=\{1,\cdots,I\}, y\in\mathfrak{Y}=\{1,\cdots,J\}, z\in\mathfrak{Z}=\{1,\cdots,K\}$, and two transition matrices $\{P_1(y|x)\}$ and $\{P_2(z|y)\}$. By the discrete memorylessness of the channel, for any N,

$$p(y,z|x)=\prod_{i=1}^{N}p_1(y_i|x_i)p_2(z_i|y_i),$$

Manuscript received December 13, 1976; revised August 31, 1977. This work was supported in part by the National Science Foundation under Grant ENG 76-03684.

The author is with Stanford University, Stanford, CA 94305.

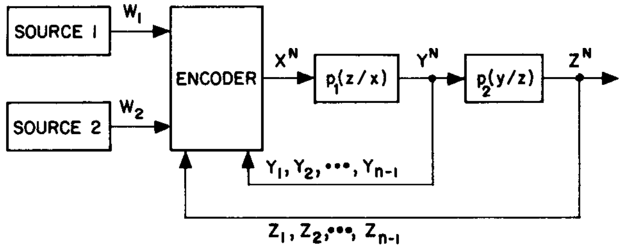

Fig. 1.

for $\boldsymbol{x}\in\mathfrak{X}^N$, $\boldsymbol{y}\in\mathcal{Y}^N$, $\boldsymbol{z}\in\mathfrak{Z}^N$.

Define an (M_1,M_2,N,λ_N) code for the channel to be a set of functions $\{f_n\}$ together with the associated set of codewords generated as in (1), and two decoding maps

$$g_1:\mathcal{Y}^N\to\mathfrak{M}_1, \qquad g_2:\mathfrak{Z}^N\to\mathfrak{M}_2,$$

such that

$$\max\left\{p_{e,1}^N,p_{e,2}^N\right\}\triangleq\lambda_N,$$

where, assuming uniform distribution on $\mathfrak{M}_1\times\mathfrak{M}_2$

$$p_{e,1}^N=\sum_{(w_1,w_2)\in\mathfrak{M}_1\times\mathfrak{M}_2}\frac{1}{M_1M_2}p\{g_1(\boldsymbol{Y})\neq w_1|(w_1,w_2)\text{ sent}\},$$

$$p_{e,2}^N=\sum_{(w_1,w_2)\in\mathfrak{M}_1\times\mathfrak{M}_2}\frac{1}{M_1M_2}P\{g_2(\boldsymbol{Z})\neq w_2|(w_1,w_2)\text{ sent}\},$$

are the average probabilities of error.

A rate pair (R_1,R_2) is said to be *achievable* if there exists a sequence of $(2^{nR_1},2^{nR_2},n,\lambda_n)$ codes with $\lambda_n\to 0$. The *capacity region* is defined to be the closure of the set of all achievable (R_1,R_2) pairs.

As established by [2]–[4], and [6], the capacity region for the channel without feedback is given by the following.

Theorem 1: Let U,X,Y,Z be a joint ensemble such that X,Y,Z are as before, the number of sample points U can take on is $\min\{I,J,K\}$, and the probability mass function on U,X,Y,Z is of the form

$$p(u,x,y,z)=Q_1(u)Q_2(x|u)p_1(y|x)p_2(z|y).$$

Then the set of all pairs $\{(I(X;\,Y|U),I(U;\,Z))\}$ is the capacity region.

An equivalent characterization is $R^*=\{(R_1,R_2):R_2+\lambda R_1\leqslant C(\lambda),\ \forall\lambda>0\}$, where $C(\lambda)=\max_{Q_1(u)Q_2(x|u)}\{I(U;Z)+\lambda I\cdot(X;\,Y|U)\}$. It will be shown that $C(\lambda)$ is unchanged when feedback is added.

III. Feedback Converse

Theorem 2: (converse): If for some $\lambda>0$, $\epsilon>0$ we have

$$R_2+\lambda R_1\geqslant C(\lambda)+\epsilon, \tag{2}$$

then there exists an $\delta>0$ such that

$$\max\left\{p_{e,1}^N,p_{e,2}^N\right\}\geqslant\delta, \qquad \text{for all } N.$$

Proof: Given any (M_1,M_2,N) code, the probability mass function on the joint ensemble $W_1,W_2,\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}$ is of the form

$$p(w_1,w_2,\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$$
$$=\frac{1}{M_1M_2}\prod_{n=1}^N q_n(x_n|w_1,w_2,z_1,\cdots,z_{n-1},y_1,\cdots,y_{n-1})$$
$$\cdot\prod_{n=1}^N P_1(y_n|x_n)\prod_{n=1}^N P_2(z_n|y_n) \tag{3}$$

where $\{q_n(x_n|w_1,w_2,z_1,\cdots,z_{n-1},y_1,\cdots,y_{n-1})\}$, $1\leqslant n\leqslant N$, is the set of probability transition functions generated by the code. It follows that

i) $$I(W_2;\boldsymbol{Z})=H(W_2)-H(W_2|\boldsymbol{Z})=NR_2-H(W_2|\boldsymbol{Z}),$$

ii) $$I(W_1;\,\boldsymbol{Y}|W_2)=H(W_1|W_2)-H(W_1|\boldsymbol{Y},W_2)\geqslant NR_1-H(W_1|\boldsymbol{Y}),$$

where R_i is the rate of source i, $i=1,2$.

Thus

$$N(R_2+\lambda R_1)\leqslant[I(W_2;\boldsymbol{Z})+\lambda I(W_1;\,\boldsymbol{Y}|W_2)]+[H(W_2|\boldsymbol{Z})+\lambda H(W_1|\boldsymbol{Y})]. \tag{4}$$

The following lemma relates the first bracketed quantity to $C(\lambda)$.

Lemma 3: For all $\lambda\geqslant 0$,

$$I(W_2;\boldsymbol{Z})+\lambda I(W_1;\,\boldsymbol{Y}|W_2)\leqslant\sum_{n=1}^N[I(U_n;\,Z_n)+\lambda I(X_n;\,Y_n|U_n)]\leqslant NC(\lambda),$$

where

$$U_n\triangleq(W_2,Y_1,\cdots,Y_{n-1},Z_1,\cdots,Z_{n-1}). \tag{5}$$

Proof: A similar lemma was proved in [4] for the channel without feedback. We follow steps parallel to those in [4] in order to show the necessity of introducing a different definition for U_n, as well as adding an additional identity that demonstrates the intuitive fact that, even if the receiver Y were given the sequence $\boldsymbol{Z}$, the achievable rate region would remain unchanged. Now consider

$$\begin{aligned}I(W_2;\boldsymbol{Z})&=\sum_{n=1}^N I(W_2;\,Z_n|Z_1,\cdots,Z_{n-1})\\&=\sum_{n=1}^N H(Z_n|Z_1,\cdots,Z_{n-1})-H(Z_n|W_2,Z_1,\cdots,Z_{n-1})\\&\leqslant\sum_{n=1}^N H(Z_n)-H(Z_n|W_2,Z_1,\cdots,Z_{n-1},Y_1,\cdots,Y_{n-1})\\&=\sum_{n=1}^N I(Z_n;\,U_n).\end{aligned}$$

The change in (5) from [4] in which $U_n=(W_2,Y_1,\cdots,Y_{n-1})$ is necessary since, with feedback, Z_n and $Z_1,\cdots,Z_{n-1}$ are not necessarily independent given $Y_1,\cdots,Y_{n-1}$.

Next we show that

$$I(W_1;\,\boldsymbol{Y}|W_2)\leqslant\sum_{n=1}^N I(X_n;\,Y_n|U_n).$$

From the definition of the channel it is clear that

$$\begin{aligned}&I(W_1;\,\boldsymbol{Y}|W_2)\\&\quad=I(W_1;\,\boldsymbol{Y},\boldsymbol{Z}|W_2)\\&\quad=\sum_{n=1}^N I(W_1;\,Y_n,Z_n|W_2,Y_1,\cdots,Y_{n-1},Z_1,\cdots,Z_{n-1})\\&\quad=\sum_{n=1}^N I(W_1;\,Y_n,Z_n|U_n).\end{aligned}$$

Applying the data processing inequality we get

$$\begin{aligned}\sum_{n=1}^N I(W_1;\,Y_n,Z_n|U_n)&\leqslant\sum_{n=1}^N I(W_1,X_n;\,Y_n,Z_n|U_n)\\&=\sum_{n=1}^N H(Y_n,Z_n|U_n)-H(Y_n,Z_n|U_n,X_n,W_1)\end{aligned}$$

But by the discrete memoryless assumption (Y_n, Z_n) and (U_n, W_1) are conditionally independent given X_n. Thus

$$\sum_{n=1}^{N} I(W_1, X_n;\, Y_n, Z_n | U_n) = \sum_{n=1}^{N} I(X_n;\, Y_n, Z_n | U_n).$$

Since

$$I(X_n;\, Y_n, Z_n | U_n) = I(X_n;\, Y_n | U_n) + I(X_n;\, Z_n | Y_n, U_n),$$

it remains to show that $I(X_n;\, Z_n | Y_n, U_n) = 0$. But this is true because $U_n \rightarrow X_n \rightarrow Y_n \rightarrow Z_n$ form a Markov chain in this order. Hence

$$H(Z_n | Y_n, U_n) = H(Z_n | Y_n, U_n, X_n) = H(Z_n | Y_n).$$

This establishes the required result, that

$$I(W_1;\, Y | W_2) \leqslant \sum_{n=1}^{N} I(X_n;\, Y_n | U_n),$$

and Lemma (3) is proved.

Now, combining Lemma 3 and (4), and applying Fano's inequality, we obtain

$$(R_2 + \lambda R_1) \leqslant C(\lambda) + \left[R_2 P_{e,2}^N + \frac{1}{N} h(P_{e,2}^N) + \lambda R_1 P_{e,1}^N + \lambda \frac{1}{N} h(P_{e,1}^N) \right]. \quad (6)$$

Now taking the limit in (6) as N tends to infinity, and assuming that (R_1, R_2) is achievable, we obtain

$$(R_2 + \lambda R_1) \leqslant C(\lambda).$$

But this contradicts Assumption 2). Therefore there must exist $\delta > 0$ such that

$$\max\{P_{e,1}^N, P_{e,2}^N\} \geqslant \delta, \qquad \text{for all } N,$$

and the proof of Theorem 2 is completed.

IV. Discussion

It has been shown that feedback cannot increase the capacity of the physically degraded broadcast channel. This is consistent with Shannon's result on the discrete memoryless channel with feedback [5]. We believe that a similar result can be proved for any discrete memoryless broadcast channel. However, the correlation between the outputs of the channel seem to cause some difficulties in generalizing the proof even to the general degraded case.

It is also important to point out that, in the case of multiple access channels, feedback can increase the channel capacity [7].

Additional Comment: While this paper was being refereed, it has come to the author's attention that W. J. Leighton and H. H. Tan of Princeton University were independently able to prove the result in this paper. Their proof is essentially along the same lines.

Acknowledgment

The author would like to thank Prof. Thomas Cover for discussions during the preparation of this paper.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.
[2] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.
[3] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain sequences and applications, Part II," *IEEE Trans. Inform. Theory*, vol. II-19, pp. 772–777, Nov. 1973.
[4] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.
[5] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, IT-2 pp. 8–19, Sept. 1956.
[6] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 6, Nov. 1975.
[7] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 100–102, Jan. 1975.

Optimal Linear Coding Schemes for Feedback Communication with Noisy Side Information

HIDEAKI SAKAI, TAKASHI SOEDA, AND HIDEKATSU TOKUMARU

***Abstract*—Optimal linear coding schemes for feedback communication with noisy side information are derived. The performance of two schemes with side information are compared with the information-theoretic bounds. The side information at the transmitter significantly reduces the mean-squared error. The corresponding continuous-time system is briefly mentioned.**

I. Introduction and Summary

This correspondence considers the problem of transmitting amplitude-continuous information X taken from a distribution $N(0, \sigma_X^2)$ once every N channel uses over a time-discrete amplitude-continuous additive white Gaussian noise channel. At the nth time instant, the signal S_n having average energy not more than E_0 is sent over the channel and is corrupted by zero-mean white Gaussian noise N_n having variance σ_N^2. Also there is a noiseless feedback link from the receiver to the transmitter through which a convenient feedback signal f_n can be transmitted with one time unit delay. This problem has been well studied in the literature, and some simple coding schemes actually attain the rate-distortion bounds (RDB) both for the discrete-time case [1] and the continuous-time case [2].

In addition, the communication schemes considered here are provided with noisy side information $Y + W_n$ where Y is the side information correlated with X and $\{W_n\}$ is a sequence of independent identically distributed $N(0, \sigma_W^2)$ random variables. We assume that all random quantities are mutually uncorrelated except X and Y. When the side information was available only at the receiver, the performance of a suboptimal coding scheme was calculated in [3]. In this note, we derive the optimal linear coding schemes for the following two assumptions about the side information, by using the theory of innovation processes.

Case 1: Only the receiver can use the side information. This is the same situation as in [3].

Case 2: Both the transmitter and the receiver can use the side information.

When the side information is noise-free we can compare the performance of the optimal linear schemes for the above cases with the RDB's. In this special situation, it follows from the result of Wyner and Ziv [4] that the RDB's are the same for both cases. However, it is shown below that the optimal linear scheme for Case 2 can attain the ideal performance, whereas the one for Case 1 cannot. In other words, the noise-free side information at the transmitter contributes significantly to the reduction of the mean-squared error by causal linear coding, whereas it is no use for reducing the error by noncausal block coding. However there is no difference between Case 1 and Case 2 in the continuous-time version of the problem, for which the optimality of the linear scheme can be established without knowledge of the corresponding rate-distortion functions.

Manuscript received December 30, 1976; revised June 28, 1977.
H. Sakai and T. Soeda are with the Department of Information Sciences and Systems Engineering, Faculty of Engineering, University of Tokushima, Tokushima, 770, Japan.
H. Tokumaru is with the Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, 606, Japan.