Note that like P_2, the P in Theorem 7 is effective for infinite C, only if C is an effective set. Note also that in the set used to show that $\alpha(c,P) \leq |C| - 1$ is the best possible, although $|C| - 1$ accesses are required by P to find each c of size $|c| = |C| - 1$, it is not the $|C| - 1$ addresses in $D(c)$ that are accessed, in general. Thus P_3 finds c^3 in two accesses, but the two accesses are never both in $D(c^3)$.

VII. Acknowledgment

I would like to express my appreciation to Cecil H. Green whose gift of a chair, to the Department of Electrical Engineering of M.I.T. for use in rotation by faculty members exploring new directions of research, made concentrated effort on this and related topics possible.

References

[1] N. Abramson, *Information Theory and Coding*. New York: McGraw-Hill, 1963, esp. ch. 3.
[2] G. J. Chaiten, "A theory of program size formally identical to information theory," to be published in *J. Ass. Comput. Mach.*
[3] P. Elias, "Efficient storage and retrieval by content and address of static files," *J. Ass. Comput. Mach.*, vol. 21, pp. 246–260, 1974.
[4] ——, "Minimum times and memories needed to compute the values of a function," *J. Comput. Syst. Sci.*, vol. 8, pp. 196–212, 1974.
[5] ——, "Universal codeword sets and representations of the integers," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 194–203, Mar. 1975.
[6] P. Elias and R. A. Flower, "The complexity of some simple retrieval problems," to be published in *J. Ass. Comput. Mach.*, July 1975.
[7] R. A. Flower, "Computer updating of a data structure," Res. Lab. Electron., M.I.T., Cambridge, Quart. Progress Rep., vol. 110, pp. 147–154, July 1973.
[8] ——, "An analysis of optimal retrieval systems with updates," Res. Lab. Electron., M.I.T., Tech. Rep. 488.
[9] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, esp. ch. 3.
[10] L. G. Kraft, "A device for quantizing, grouping and coding amplitude modulated pulses," M.S. thesis, Dep. Elec. Eng., M.I.T., Cambridge, 1949.
[11] M. Minsky and S. Papert, *Perceptrons*. Cambridge, Mass.: M.I.T., 1969.
[12] R. L. Rivest, "Analysis of associative retrieval algorithms," Ph.D. dissertation, Comput. Sci. Dep., Stanford Univ., Stanford, Calif., 1974.
[13] ——, "On hash-coding algorithms for partial-match retrieval," in *Proc. 1974 Switching and Automata Theory Conf.*, 1974, pp. 95–103.
[14] T. Welch, "Bounds on information retrieval efficiency in static file structures," M.I.T., Cambridge, Mass., MAC TR-88, Proj. MAC, 1971.

An Achievable Rate Region for the Broadcast Channel

THOMAS M. COVER, FELLOW, IEEE

***Abstract*—Let $p(y_1, y_2 \mid x)$ denote a discrete memoryless channel with a single source X and two independent receivers Y_1 and Y_2. We exhibit an achievable region of rates (R_{11},R_{12},R_{22}) at which independent information can be sent, respectively, to receiver 1, to both receivers 1 and 2, and to receiver 2. The achievability of the region is shown by using a version of the asymptotic equipartition property involving many simultaneous "typicality" constraints. These results immediately generalize to yield an achievable rate region for the m-sender n-receiver channel in terms of standard mutual information quantities.**

I. Introduction

WE SHALL define a two-receiver memoryless *broadcast channel*, denoted by $(X, p(y_1, y_2 \mid x), Y_1 \times Y_2)$ or by $p(y_1, y_2 \mid x)$, to consist of three finite sets X, Y_1, Y_2 and a collection of probability distributions $p(\cdot, \cdot \mid x)$ on $Y_1 \times Y_2$, one for each $x \in X$. The interpretation is that x is an input to the channel and y_1 and y_2 are the respective outputs at receiver terminals 1 and 2 as shown in Fig. 1. The problem is to communicate simultaneously with receivers 1 and 2 as efficiently as possible.

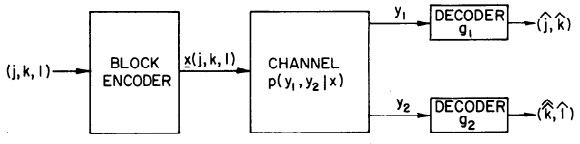

Fig. 1. Broadcast channel.

This problem was posed in [1], and a sequence of contributions to the solution has appeared in [2]–[9]. The achievable region $\mathscr{R}$ in this paper coincides with the capacity region in the special case of the degraded broadcast channel [3]. The results here are closely related to those of [8], [9]. The independent contribution [8] contains a statement of the achievability of the region $\mathscr{R}$ treated in this paper and outlines a proof based on results in [9].

The nth extension for a broadcast channel is the broadcast channel

$$(X^n, p(y_1, y_2 \mid x), Y_1^n \times Y_2^n) \tag{1}$$

Manuscript received July 29, 1974; revised February 13, 1975. This work was supported in part by the National Science Foundation under Grant GK-34363 and in part by the Advanced Research Projects Agency under Contract DAHC-15-73-C-0187 with Stanford Research Institute.

The author is with the Departments of Electrical Engineering and Statistics, Stanford University, Stanford, Calif. 94305.

where $p(y_1, y_2 \mid x) = \prod_{j=1}^{n} p(y_{1j}, y_{2j} \mid x_j)$, for $x \in X^n$, $y_1 \in Y_1^n$, $y_2 \in Y_2^n$.

An $((M_{11},M_{22},M_{12}),n)$ code for a broadcast channel consists of three sets of integers

$$\begin{aligned} \mathcal{M}_{11} &= \{1,2,\cdots,M_{11}\} \\ \mathcal{M}_{12} &= \{1,2,\cdots,M_{12}\} \\ \mathcal{M}_{22} &= \{1,2,\cdots,M_{22}\} \end{aligned} \tag{2}$$

an encoding function

$$x: \mathcal{M}_{11} \times \mathcal{M}_{12} \times \mathcal{M}_{22} \to X^n \tag{3}$$

and two decoding functions

$$\begin{aligned} g_1&: Y_1^n \to \mathcal{M}_{11} \times \mathcal{M}_{12}; g_1(y_1) = (\hat{j},\hat{k}) \\ g_2&: Y_2^n \to \mathcal{M}_{12} \times \mathcal{M}_{22}; g_2(y_2) = (\hat{k},\hat{l}). \end{aligned} \tag{4}$$

The set $\{x(j,k,l) \mid (j,k,l) \in \mathcal{M}_{11} \times \mathcal{M}_{12} \times \mathcal{M}_{22}\}$ is called the set of codewords. As illustrated in Fig. 1, we think of integers j and l as being arbitrarily chosen by the transmitter to be sent to receivers 1 and 2, respectively. The integer k is also chosen by the transmitter and is intended to be received by both receivers. Thus k is the "common" part of the message, and j and l are the "independent" parts of the message.

If the message (j,k,l) is sent, let

$$\lambda(j,k,l) = \Pr\{g_1(Y_1) \neq (j,k) \text{ or } g_2(Y_2) \neq (k,l)\} \tag{5}$$

denote the probability of error, where we note that Y_1, Y_2 are the only chance variables in the preceding expression. We define the *average probability of error* of the code as

$$P(e) = P_n(e) = \frac{1}{M} \sum_{j,k,l} \lambda(j,k,l) \tag{6}$$

where

$$M = M_{11}M_{22}M_{12}. \tag{7}$$

This probability of error is calculated under a special distribution, namely, the uniform distribution over the codewords.

We shall define the *rate* (R_{11},R_{12},R_{22}) of an $((M_{11},M_{12},M_{22}),n)$ code by

$$\begin{aligned} R_{11} &= \frac{1}{n} \log M_{11} \\ R_{12} &= \frac{1}{n} \log M_{12} \\ R_{22} &= \frac{1}{n} \log M_{22} \end{aligned} \tag{8}$$

all defined in bits per transmission. Thus R_{ii} is the rate of transmission of independent information to receiver i, for $i = 1,2$, and R_{12} is the portion of the information common to both receivers.

Definition: The rate (R_{11},R_{12},R_{22}) is said to be *achievable* by a broadcast channel if, for any $\varepsilon > 0$ and for all n sufficiently large, there exists an $((M_{11},M_{12},M_{22}),n)$ code with

$$\begin{aligned} M_{11} &\geq 2^{nR_{11}} \\ M_{12} &\geq 2^{nR_{12}} \\ M_{22} &\geq 2^{nR_{22}} \end{aligned} \tag{9}$$

such that $P_n(e) < \varepsilon$.

Comment: Note that the total number $M = M_{11}M_{12}M_{22}$ of codewords for a code satisfying (9) must exceed $2^{n(R_{11}+R_{12}+R_{22})}$.

Definition: The *capacity region* $\mathcal{R}^*$ for a broadcast channel is the set of all achievable rates (R_{11},R_{12},R_{22}).

The goal of this paper is to determine an achievable region $\mathcal{R}$ and thus bound $\mathcal{R}^*$ below by $\mathcal{R} \subseteq \mathcal{R}^*$.

Comment: The extension of the definition of the broadcast channel from two receivers to k receivers is notationally cumbersome but straightforward given the following comment. The index sets M_{11},M_{12},M_{22} are replaced by $2^k - 1$ index sets $I(\theta)$, $\theta \in \{0,1\}^k$, $\theta \neq 0$, with the interpretation that the integer $i(\theta)$ selected in index set $I(\theta) = \{1,2,\cdots,M(\theta)\}$ is intended (by the proper code selection) to be received correctly by every receiver j for which $\theta_j = 1$ in $\theta = (\theta_1,\theta_2,\cdots,\theta_k)$. Then, for example, the rate of transmission over the nth extension of a broadcast channel to the ith receiver will be given by

$$R_i = \frac{1}{n} \log \prod_{\substack{\theta \in \{0,1\}^k \\ \theta_i = 1}} M(\theta) = \frac{1}{n} \sum_{\theta_i = 1} \log M(\theta). \tag{10}$$

In the two-receiver broadcast channel, the corresponding sets in the new notation are $M_{12} = I(1,1)$, $M_{11} = I(1,0)$, $M_{22} = I(0,1)$.

II. Achievable Region $\mathcal{R}$

In this section we shall exhibit an achievable $\mathcal{R}$. We shall prove the achievability of $\mathcal{R}$ in Section IV. First we shall define three auxiliary random variables U,R,V taking values in finite sets $\mathcal{U},\mathcal{R},\mathcal{V}$. Let $x: \mathcal{U} \times \mathcal{R} \times \mathcal{V} \to \mathcal{X}$ denote an arbitrary mapping of the auxiliary random variables into the input alphabet. The picture we have in mind is given in Fig. 2.

For each assignment of probability distributions $p(u),p(r),p(v)$ and mapping function $x(\cdot)$, we associate the joint probability distribution function

$$p(u,r,v,y_1,y_2) = p(u)p(r)p(v)p(y_1, y_2 \mid x(u,r,v)). \tag{11}$$

Mutual information quantities like

$$I(U,R;Y_1) = \sum_{u,r,y_1} p(u,r,y_1) \log \frac{p(u,r,y_1)}{p(u,r)p(y_1)} \tag{12}$$

are defined in the usual way.

Define

$$\begin{aligned} \boldsymbol{I} &= (I_1,I_2,\cdots,I_6) \\ &= (I(U;R,Y_1),I(V;R,Y_2),I(R;U,Y_1), \\ &\quad I(R;V,Y_2),I(U,R;Y_1),I(V,R;Y_2)). \end{aligned} \tag{13}$$

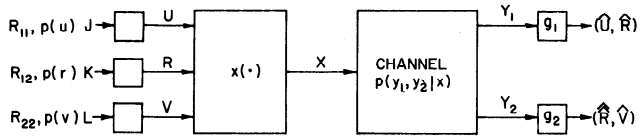

Fig. 2. Auxiliary random variables.

Let $\mathscr{I}$ denote the set of all $I \in R^6$ generated by all assignments of $p(u),p(r),p(v),x(\cdot)$. Let $C_0(\mathscr{I})$ denote the convex hull of $\mathscr{I}$. Let $R(I)$ denote the set of all $(R_{11},R_{12},R_{22}) \in R^3$ satisfying the six inequalities

$$\begin{aligned} R(I): \qquad R_{11} &< I_1 \\ R_{22} &< I_2 \\ R_{12} &< I_3 \\ R_{12} &< I_4 \\ R_{11} + R_{12} &< I_5 \\ R_{12} + R_{22} &< I_6. \end{aligned} \tag{14}$$

Theorem 1: The region

$$\mathscr{R} = \bigcup_{I \in C_0(\mathscr{I})} R(I) \tag{15}$$

is achievable.

We can express the capacity region in another form. Observe that an arbitrary point I on the boundary of $C_0(\mathscr{I})$ can always be expressed as the convex combination of no more than six (extreme) points of $\mathscr{I}$. Let

$$I = \sum_{i=1}^{6} p(q^{(i)}) I_{q^{(i)}}, \qquad I_{q^{(i)}} \in \mathscr{I} \tag{16}$$

be the desired convex combination, where $q^{(i)} = (p(u \mid q^{(i)}), p(r \mid q^{(i)}),p(v \mid q^{(i)}),x(\cdot \mid q^{(i)}))$ is an element in the set of all assignments $(p,x(\cdot))$ and $I_{q^{(i)}}$ is the vector of mutual informations induced by this assignment. Let Q denote a random variable with $\Pr\{Q = q^{(i)}\} = p(q^{(i)}),p(q^{(i)}) \geq 0$, $i = 1,2,\cdots,6, \sum p(q^{(i)}) = 1$. It then follows from inspection of the definition of mutual information that, for example,

$$\sum_{i=1}^{6} p(q^{(i)}) I_{q^{(i)}}(U;R,Y_1) = I(U;R,Y_1 \mid Q). \tag{17}$$

Thus $\mathscr{R}$ can be expressed as follows.

Theorem 1': $\mathscr{R}$ is the union of all $(R_{11},R_{12},R_{22}) \in R^3$ satisfying the inequalities

$$\begin{aligned} R_{11} &< I(U;\, R,\, Y_1 \mid Q) \\ R_{22} &< I(V;\, R,\, Y_2 \mid Q) \\ R_{12} &< I(R;\, U,\, Y_1 \mid Q) \\ R_{12} &< I(R;\, V,\, Y_2 \mid Q) \\ R_{11} + R_{12} &< I(U,\, R;\, Y_1 \mid Q) \\ R_{12} + R_{22} &< I(V,\, R;\, Y_2 \mid Q) \end{aligned} \tag{18}$$

where the union is over all random variables Q such that Q takes on no more than six values in the set of all assignments $(p,x(\cdot))$. This is the same region as in Theorem 1.

The following conjecture is a result more of wishful thinking than of soundly based intuition. If true, this conjecture would substantially reduce the computation necessary to compute $\mathscr{R}$.

Conjecture: The set of achievable rates in Theorem 1 is unchanged if the alphabet sizes of the auxiliary random variables are restricted to $\mathscr{U} = \mathscr{R} = \mathscr{V} = \{0, 1, 2,\cdots, m-1\}$, $m = |\mathscr{X}|$, and if the mapping function $x(\cdot)$ is given by

$$x = u + r + v (\text{mod } |\mathscr{X}|). \tag{19}$$

III. Jointly Typical Sequences

Before proceeding with the proof, we shall define a simultaneous asymptotic equipartition property. This will allow us to decode messages at the receiver by simply checking to see which of the possible input messages is jointly "typical" with the received output. If there is one and only one such candidate, we shall declare this to be the message.

Let $\{X^{(1)},X^{(2)},\cdots,X^{(k)}\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(x^{(1)},x^{(2)},\cdots,x^{(k)})$. Let S denote an ordered subset of these random variables and consider n independent copies of S. Thus,

$$\Pr\{S = s\} = \prod_{i=1}^{n} \Pr\{S_i = s_i\}. \tag{20}$$

For example, if $S = (X^{(j)},X^{(k)})$, then

$$\begin{aligned} \Pr\{S = s\} &= \Pr\{(X^{(j)},X^{(k)}) = (x^{(j)},x^{(k)})\} \\ &= \prod_{i=1}^{n} p(x_i^{(j)},x_i^{(k)}). \end{aligned} \tag{21}$$

We know by the law of large numbers that for a given subset S of random variables

$$-\frac{1}{n} \log p(S_1,S_2,\cdots,S_n) = -\sum_{i=1}^{n} \log p(S_i) \to H(S) \tag{22}$$

with probability one. This convergence takes place simultaneously with probability one for all 2^k subsets

$$S \subseteq \{X^{(1)},X^{(2)},\cdots,X^{(k)}\}. \tag{23}$$

Definition: The set A_ε of *jointly ε-typical* n-sequences $(x^{(1)},x^{(2)},\cdots,x^{(k)})$ is defined by

$$\begin{aligned} &A_\varepsilon(X^{(1)},X^{(2)},\cdots,X^{(k)}) \\ &= \Big\{(x^{(1)},x^{(2)},\cdots,x^{(k)}) \in (\mathscr{X}^{(1)})^n \\ &\quad \times (\mathscr{X}^{(2)})^n \times \cdots (\mathscr{X}^{(k)})^n : \left| -\frac{1}{n} \log p(s) - H(S) \right| \leq \varepsilon, \\ &\quad \forall S \subseteq \{X^{(1)},X^{(2)},\cdots,X^{(k)}\} \Big\} \end{aligned} \tag{24}$$

where s denotes the ordered set of sequences in $\{x^{(1)},\cdots,x^{(k)}\}$ corresponding to S. Let $A_\varepsilon(S)$ denote the restriction of A_ε to the coordinates corresponding to S.

Thus, for example,

$$\begin{aligned}A_\varepsilon(X^{(1)},X^{(2)}) &= \Big\{(x^{(1)},x^{(2)}): \Big|-\frac{1}{n}\log p(x^{(1)},x^{(2)}) - H(X^{(1)},X^{(2)})\Big| < \varepsilon,\\ &\quad \Big|-\frac{1}{n}\log p(x^{(1)}) - H(X^{(1)})\Big| < \varepsilon,\\ &\quad \Big|-\frac{1}{n}\log p(x^{(2)}) - H(X^{(2)})\Big| < \varepsilon\Big\}. \end{aligned} \tag{25}$$

The following is a version of the asymptotic equipartition property involving simultaneous constraints. (Compare with Forney [10].)

Lemma 1: For any $\varepsilon > 0$, there exists an integer n such that $A_\varepsilon(S)$ satisfies

$$\begin{aligned}&\text{i) } \Pr\{A_\varepsilon(S)\} \geq 1-\varepsilon, \qquad \forall S \subseteq \{X^{(1)},\cdots,X^{(k)}\}\\ &\text{ii) } \quad s \in A_\varepsilon(S) \Rightarrow \Big|-\frac{1}{n}\log p(s) - H(S)\Big| < \varepsilon\\ &\text{iii) } \quad (1-\varepsilon)2^{n(H(S)-\varepsilon)} \leq |A_\varepsilon(S)| \leq 2^{n(H(S)+\varepsilon)}.\end{aligned} \tag{26}$$

Proof: Part i) follows from the convergence of the random variables in the definition of $A_\varepsilon(S)$. Part ii) follows immediately from the definition of $A_\varepsilon(S)$ in (24). Part iii) follows from

$$1 \geq \sum_{s\in A_\varepsilon(S)} p(s) \overset{(ii)}{\geq} \sum_{A_\varepsilon(S)} 2^{-n(H(S)+\varepsilon)} = |A_\varepsilon(S)|2^{-n(H(S)+\varepsilon)}, \tag{27}$$

and

$$(1-\varepsilon) \leq \sum_{s\in A_\varepsilon(S)} p(s) \overset{(ii)}{\leq} |A_\varepsilon(S)|2^{-n(H(S)-\varepsilon)}. \tag{28}$$

Corollary: If $(w,z) \in A_\varepsilon(W,Z)$, then

$$2^{-n(H(W|Z)+2\varepsilon)} \leq p(w \mid z) \leq 2^{-n(H(W|Z)-2\varepsilon)}. \tag{29}$$

Proof: $(w,z) \in A_\varepsilon(W,Z)$ implies

$$2^{-n(H(W,Z)+\varepsilon)} \leq p(w,z) \leq 2^{-n(H(W,Z)-\varepsilon)} \tag{30}$$

and

$$2^{-n(H(Z)+\varepsilon)} \leq p(z) \leq 2^{-n(H(Z)-\varepsilon)}. \tag{31}$$

The corollary follows from $p(w \mid z) = p(w,z)/p(z)$.

We shall need to know the probability that conditionally independent sequences are jointly typical. Let the discrete random variables W,Z,Q have joint distribution $p(w,z,q)$. Let W',Z' be conditionally independent given Q, with the marginals

$$\begin{aligned}p(w \mid q) &= \sum_z p(w,z,q)/p(q)\\ p(z \mid q) &= \sum_w p(w,z,q)/p(q).\end{aligned} \tag{32}$$

The unconditional version of the following lemma has been observed and proved by Forney [10] as crucial in giving the natural proof of Shannon's second theorem. This lemma has also been independently used by the author on source compression for dependent ergodic sources [11].

Lemma 2: Let

$$(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{Q}) \sim \prod_{i=1}^n p(w_i,z_i,q_i)$$

and

$$(\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) \sim \prod_{i=1}^n p(w_i \mid q_i)p(z_i \mid q_i)p(q_i)$$

as in the preceding. Then, for n such that $\Pr\{A_\varepsilon\} \geq 1-\varepsilon$,

$$\begin{aligned}(1-\varepsilon)2^{-n(I(W;Z|Q)+7\varepsilon)} &\leq \Pr\{(\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) \in A_\varepsilon(W,Z,Q)\}\\ &\leq 2^{-n(I(W;Z|Q)-7\varepsilon)}.\end{aligned} \tag{33}$$

Proof: From the corollary, $(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q}) \in A_\varepsilon(W,Z,Q)$ implies

$$\begin{aligned}2^{-n(H(W|Q)+H(Z|Q)+H(Q)+6\varepsilon)} &\leq p(\boldsymbol{w} \mid \boldsymbol{q})\,p(\boldsymbol{z} \mid \boldsymbol{q})\,p(\boldsymbol{q})\\ &\leq 2^{-n(H(W|Q)+H(Z|Q)+H(Q)-6\varepsilon)}\end{aligned} \tag{34}$$

and

$$2^{-n(H(W,Z,Q)+\varepsilon)} \leq p(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q}) \leq 2^{-n(H(W,Z,Q)-\varepsilon)}. \tag{35}$$

Observe that

$$\begin{aligned}H(W \mid Q) + H(Z \mid Q) + H(Q) - H(W,Z,Q)\\ = H(W \mid Q) + H(Z \mid Q) - H(W, Z \mid Q)\\ = I(W\,;\,Z \mid Q).\end{aligned} \tag{36}$$

Consequently,

$$\begin{aligned}p(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q})2^{-n(I(W;Z|Q)+7\varepsilon)} &\leq p(\boldsymbol{w} \mid \boldsymbol{q})\,p(\boldsymbol{z} \mid \boldsymbol{q})\,p(\boldsymbol{q})\\ &\leq p(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q})2^{-n(I(W;Z|Q)-7\varepsilon)}.\end{aligned} \tag{37}$$

Now summing the terms in the preceding equation over A_ε,

$$\begin{aligned}(1-\varepsilon)2^{-n(I(W;Z|Q)+7\varepsilon)} &\leq \sum_{A_\varepsilon} 2^{-n(I(W;Z|Q)+7\varepsilon)}p(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q})\\ &\leq \sum_{A_\varepsilon} p(\boldsymbol{w} \mid \boldsymbol{q})\,p(\boldsymbol{z} \mid \boldsymbol{q})\,p(\boldsymbol{q})\\ &= \Pr\{(\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) \in A_\varepsilon(W,Z,Q)\}\\ &\leq 2^{-n(I(W;Z|Q)-7\varepsilon)} \sum_{A_\varepsilon} p(\boldsymbol{w},\boldsymbol{z},\boldsymbol{q})\\ &\leq 2^{-n(I(W;Z|Q)-7\varepsilon)}.\end{aligned} \tag{38}$$

Q.E.D.

IV. Achievability of $\mathscr{R}$

For any $I \in C_0(\mathscr{I})$, we shall show how to achieve any rate (R_{11},R_{12},R_{22}) satisfying the six inequalities in (18). Consider a given assignment $p(u \mid q)$, $p(r \mid q)$, $p(v \mid q)$, $x(\cdot \mid q)$, $p(q)$, $q \in \{q^{(1)},q^{(2)},\cdots,q^{(6)}\}$ and the associated $\boldsymbol{I} \in C_0(\mathscr{I})$ given in (17).

Random Encoding for Theorem 1'

First, generate a sequence of n independent identically distributed (i.i.d.) random variables $\boldsymbol{Q} = (Q_1,Q_2,\cdots,Q_n)$. Here $\boldsymbol{Q} = \boldsymbol{q}$ plays the role of a time-sharing parameter that at each time k informs the transmitter and both receivers that the mode of operation is $Q_k = q_k$, where q_k is one of the six modes in $\{q^{(1)},q^{(2)},\cdots,q^{(6)}\}$. Conditioned on $\boldsymbol{Q} = \boldsymbol{q}$, generate $2^{nR_{11}}$ random n-sequences of random variables drawn according to $p(\boldsymbol{u} \mid \boldsymbol{q})$, $2^{nR_{12}}$ random n-sequences of random variables drawn according to $p(\boldsymbol{r} \mid \boldsymbol{q})$, and $2^{nR_{22}}$ random n-sequences of random variables drawn according to $p(\boldsymbol{v} \mid \boldsymbol{q})$. Index the strings by $j = 1,2,\cdots,2^{nR_{11}}$, $k = 1,2,\cdots,2^{nR_{12}}$, and $l = 1,2,\cdots,2^{nR_{22}}$, respectively. Thus, for example, the jth n-sequence (word) $\boldsymbol{u}(j)$ has probability

$$\Pr\{\boldsymbol{U}(j) = \boldsymbol{u}(j) \mid \boldsymbol{Q} = \boldsymbol{q}\} = \prod_{i=1}^{n} p(u_i(j) \mid q_i). \quad (39)$$

Also, $\boldsymbol{U}(j),\boldsymbol{R}(k),\boldsymbol{V}(l),\forall j,k,l$, are conditionally independent given $\boldsymbol{Q}$.

To each (j,k,l) there corresponds a triple of n-sequences $(\boldsymbol{u}(j),\boldsymbol{r}(k),\boldsymbol{v}(l))$, and the codeword

$$\boldsymbol{x}(j,k,l) = (x_1(j,k,l),x_2(j,k,l),\cdots,x_n(j,k,l)) \quad (40)$$

where

$$x_m(j,k,l) = x(u_m(j), r_m(k), v_m(l) \mid q_m). \quad (41)$$

The code book consists of the M n-sequences $\boldsymbol{x}(j,k,l)$, $(j,k,l) \in \mathcal{M}_{11} \times \mathcal{M}_{12} \times \mathcal{M}_{22}$.

Decoding Rule

Both receivers know $\boldsymbol{q}$. If $\boldsymbol{y}_1$ is received, declare $(\hat{j},\hat{k}) = (j,k)$ was sent if there is one and only one pair $(j,k) \in \mathcal{M}_{11} \times \mathcal{M}_{12}$ such that $(\boldsymbol{u}(j),\boldsymbol{r}(k),\boldsymbol{y}_1,\boldsymbol{q}) \in A_\varepsilon(\boldsymbol{U},\boldsymbol{R},Y_1,\boldsymbol{Q})$, i.e., if there is only one input pair (j,k) that is jointly typical with the output. If $\boldsymbol{y}_2$ is received, declare $(\hat{\hat{k}},\hat{l}) = (k,l)$ was sent if there is one and only one pair (k,l) such that $(\boldsymbol{r}(k),\boldsymbol{v}(l),\boldsymbol{y}_2,\boldsymbol{q}) \in A_\varepsilon(\boldsymbol{R},\boldsymbol{V},Y_2,\boldsymbol{Q})$.

Proof of Achievability of $\mathscr{R}$: Let J,K,L be independent random variables drawn according to uniform distributions on $\mathcal{M}_{11},\mathcal{M}_{12},\mathcal{M}_{22}$, respectively. Let the code be chosen randomly according to the encoding description. Then the probability of error (over J,K,L and the random code) is given by

$$\bar{P}(e) = \Pr\{(\hat{J},\hat{K}) \neq (J,K) \text{ or } (\hat{\hat{K}},\hat{L}) \neq (K,L)\}. \quad (42)$$

By the symmetry induced by the random coding, we see that each transmitted message (j,k,l) yields the same probability of error. Thus setting $(j,k,l) = (1,1,1)$, we have precisely

$$\bar{P}(e) = \Pr\{(\hat{J},\hat{K}) \neq (1,1) \text{ or } (\hat{\hat{K}},\hat{L}) \neq (1,1)\}. \quad (43)$$

Let $E(j,k,1)$ denote the event $(\boldsymbol{U}(j),\boldsymbol{R}(k),\boldsymbol{Y}_1,\boldsymbol{Q}) \in A_\varepsilon(\boldsymbol{U},\boldsymbol{R},Y_1,\boldsymbol{Q})$. This is the event that $(\boldsymbol{U}(j),\boldsymbol{R}(k),\boldsymbol{Y}_1,\boldsymbol{Q})$ is jointly ε-typical (as calculated under the conditional measure for Y_1 given $(J,K,L) = (1,1,1)$), thus implying that (j,k) are candidates for the decoded message $g(\boldsymbol{Y}_1)$. Similarly, let $E(k,l,2)$ denote the event $(\boldsymbol{R}(k),\boldsymbol{V}(l),\boldsymbol{Y}_2,\boldsymbol{Q}) \in A_\varepsilon(\boldsymbol{R},\boldsymbol{V},Y_2,\boldsymbol{Q})$, where A_ε is defined under the conditional measure for Y_2, given $(J,K,L) = (1,1,1)$.

Now

$$\begin{aligned}\bar{P}(e) &= \Pr\left\{E^c(1,1,1) \cup E^c(1,1,2) \bigcup_{(j,k)\neq(1,1)} E(j,k,1) \bigcup_{(k,l)\neq(1,1)} E(k,l,2)\right\} \\ &\leq \Pr\{E^c(1,1,1)\} + \Pr\{E^c(1,1,2)\} + \sum_{k\neq 1} \Pr\{E(1,k,1)\} \\ &\quad + \sum_{j\neq 1} \Pr\{E(j,1,1)\} + \sum_{\substack{j\neq 1\\ k\neq 1}} \Pr\{E(j,k,1)\} \\ &\quad + \sum_{(k,l)\neq(1,1)} \Pr\{E(k,l,2)\}. \end{aligned} \quad (44)$$

The first two terms correspond to the event that the correct codeword does not fall in the decoding set. The last terms correspond to the event that some incorrect codeword falls in the decoding set.

Choosing n sufficiently large that

$$\Pr\{A_\varepsilon(\boldsymbol{U},\boldsymbol{R},\boldsymbol{V},Y_1,Y_2)\} \geq 1 - \varepsilon \quad (45)$$

we see from Lemma 1 that

$$\Pr\{E^c(1,1,1)\} < \varepsilon, \qquad \Pr\{E^c(1,1,2)\} < \varepsilon. \quad (46)$$

Consider the event $E(j,1,1)$. We observe, for $j \neq 1$, $k = 1$, that $\boldsymbol{U}(j)$ is conditionally independent of $(\boldsymbol{R}(1),\boldsymbol{Y}_1)$ given $\boldsymbol{Q}$. Also, the distribution of $\boldsymbol{U}(j)$ given $\boldsymbol{Q}$ is the same as that of $\boldsymbol{U}(1)$ given $\boldsymbol{Q}$. Thus Lemma 2 applies with the substitution $\boldsymbol{U}(j) = \boldsymbol{W}'$ and $(\boldsymbol{R}(1),\boldsymbol{Y}_1) = \boldsymbol{Z}'$ in (33). That is, for $j \neq 1$,

$$\begin{aligned}\Pr\{E(j,1,1)\} &= \Pr\{(\boldsymbol{U}(j),\boldsymbol{R}(1),\boldsymbol{Y}_1,\boldsymbol{Q}) \in A_\varepsilon(\boldsymbol{U},\boldsymbol{R},Y_1,\boldsymbol{Q})\} \\ &\leq 2^{-n(I(U;R,Y_1|Q)-7\varepsilon)}. \end{aligned} \quad (47)$$

Consequently,

$$\sum_{j=1}^{M_{11}} \Pr\{E(j,1,1)\} \leq 2^{-n(I(U;R,Y_1|Q)-R_{11}-7\varepsilon)}. \quad (48)$$

Therefore, this term can be made less than ε for

$$R_{11} \leq I(U;R,Y_1 \mid Q) - 7\varepsilon - \frac{1}{n}\log\frac{1}{\varepsilon}. \quad (49)$$

Thus this term $\to 0$ as $n \to \infty$ if

$$R_{11} < I(U;\, R,\, Y_1 \mid Q). \quad (50)$$

This is the first condition in Theorem 1'.

Similarly, applying Lemma 2 for the terms

$$E(1,k,1) \text{ with } (\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) = (\boldsymbol{R}(k),(\boldsymbol{U}(1),\boldsymbol{Y}_1),\boldsymbol{Q})$$

$$E(j,k,1) \text{ with } (\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) = ((\boldsymbol{U}(j),\boldsymbol{R}(k)),\boldsymbol{Y}_1,\boldsymbol{Q})$$

$$E(k,1,2) \text{ with } (\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) = (\boldsymbol{R}(k),(\boldsymbol{V}(1),\boldsymbol{Y}_2),\boldsymbol{Q})$$

$$E(1,l,2) \text{ with } (\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) = (\boldsymbol{V}(l),(\boldsymbol{R}(1),\boldsymbol{Y}_2),\boldsymbol{Q})$$

and

$$E(k,l,2) \text{ with } (\boldsymbol{W}',\boldsymbol{Z}',\boldsymbol{Q}) = ((\boldsymbol{R}(k),\boldsymbol{V}(l)),\boldsymbol{Y}_2,\boldsymbol{Q}) \quad (51)$$

for $j \neq 1$, $k \neq 1$, $l \neq 1$, we find (first letting $n \to \infty$, then $\varepsilon \to 0$) that $\bar{P}_n(e) \to 0$ whenever the conditions of Theorem 1′ are satisfied.

Finally, if $\bar{P}_n(e) < \varepsilon$, there must exist at least one $((M_{11},M_{22},M_{12}),n)$ code with $P_n(e) < \varepsilon$. Thus $\bar{P}_n(e) \to 0$ implies that there exists a sequence of $((M_{11},M_{22},M_{12}),n)$ codes with $P_n(e) \to 0$, for any $(R_{11},R_{12},R_{22}) \in \mathcal{R}$. Q.E.D.

V. Concluding Remarks

It is clear from the information-theoretic form of the rate region in Theorem 1′ that extension to multiple receivers will lead to a capacity region of the same form. However, it is not yet known whether $\mathcal{R}$ is the true capacity region.

Acknowledgment

I would like to thank Dirk Hughes-Hartogs for his frequent interactions on this problem. I have also benefited from discussions with Carroll Keilers, Martin Hellman, Aydano Carleial, and Aaron Wyner.

References

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972. Reprinted in *Key Papers in the Development of Information Theory*, D. Slepian, Ed. New York: IEEE Press, 1974, pp. 437–449.
[2] P. P. Bergmans, Ph.D. dissertation, Stanford Univ., Stanford, Calif., 1972.
[3] ——, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.
[4] A. Wyner and J. Ziv., "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 769–772, Nov. 1973.
[5] A. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 772–777, Nov. 1973.
[6] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-20, pp. 279–280, Mar. 1974.
[7] R. G. Gallager, "Coding for degraded broadcast channels," to appear in *Probl. Peredach. Inform.*
[8] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, Mar. 1975.
[9] M. Ulrey, "A coding theorem for a channel with several senders and receivers," submitted to *Inform. Contr.*; also, Ph.D. dissertation, Ohio State Univ., Columbus, 1973.
[10] G. D. Forney, "Information theory," unpublished class notes, 1972.
[11] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-21, pp. 226–228, Mar. 1975.

Decoding for Channels with Insertions, Deletions, and Substitutions with Applications to Speech Recognition

LALIT R. BAHL, MEMBER, IEEE, AND FREDERICK JELINEK, FELLOW, IEEE

***Abstract*—A model for channels in which an input sequence can produce output sequences of varying length is described. An efficient computational procedure for calculating Pr $\{Y \mid X\}$ is devised, where $X = x_1,x_2,\cdots,x_M$ and $Y = y_1,y_2,\cdots,y_N$ are the input and output of the channel. A stack decoding algorithm for decoding on such channels is presented. The appropriate likelihood function is derived. Channels with memory are considered. Some applications to speech and character recognition are discussed.**

Manuscript received April 4, 1974; revised January 2, 1975.
L. R. Bahl is with the Computer Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., and also the Department of Electrical Engineering and Computer Science, Columbia University, New York, N.Y.
F. Jelinek is with the Computer Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y. 10598.

I. Introduction

SOME OF THE phenomena encountered in the recognition of speech by machine can be modeled as the transmission of information through a noisy channel that allows substitution, insertion, and deletion errors. Most speech recognition processes include a segmentation stage followed by a classification stage. In the segmentation stage, the acoustic waveform is segmented into units corresponding to a single sound. The segments are of varying lengths depending on the duration of each sound. If the segmentation process is not perfect, some sounds may be completely missed, which produces deletion errors. The segmenter may also insert extra segment boundaries, which produces