VIII. PROCESSING AND TRANSMISSION OF INFORMATION

Prof. P. Elias
Prof. R. M. Fano
Prof. D. A. Huffman
Dr. R. A. Silverman
W. G. Breckenridge

E. Ferretti
K. Joannou
J. B. O'Loughlin
L. S. Onysckevych

A. J. Osborne
E. T. Schoen
F. F. Tung
S. H. Unger
M. L. Wetherell

A. LOGICAL DEVICE FOR SMOOTHING OF BINARY SIGNALS

It has been shown (1) that the synthesis of an arbitrary sequential switching circuit results in a network of logical elements arranged with feedback paths. The feedback network can be shown to behave properly if, at a certain location within each feedback loop, there is an element having smoothing action on the binary signals which come into it. Such a smoothing element gives a change in its binary output signal only if a change in the input signal persists over a time longer than some fixed critical time. A smoothing element may be contrasted with an element having only delay properties, for the delay element would reproduce exactly, at its output, all input changes in signal value.

Figure VIII-1 contrasts the effect on a binary signal of a delay element and of a smoothing element. It is of theoretical interest to determine the relationship between smoothing and delay. The circuit of Fig. VIII-2 demonstrates that a smoothing element can be constructed of a delay element, and of a "majority" element whose output signal is reintroduced as one of its inputs around a feedback loop having arbitrarily short delay time.

The importance of the circuit derived is that smoothing elements are thus shown to be no longer necessary as building blocks of sequential circuits. Instead they may be replaced by delay elements from which other circuits having the terminal properties of a smoothing element may then be constructed. In addition, delay elements are much easier to represent algebraically than are smoothing elements.
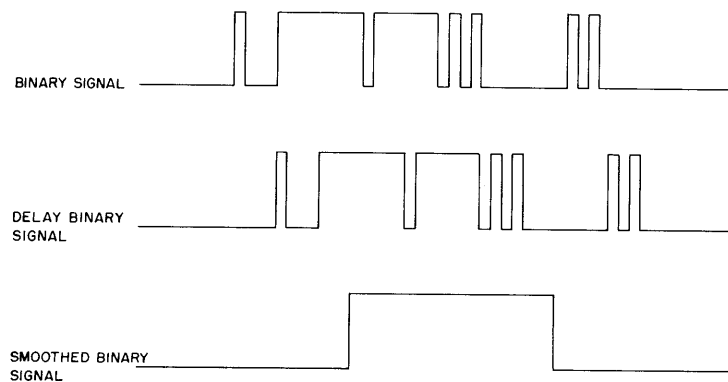
D. A. Huffman

Fig. VIII-1. The effects of delay and smoothing upon an arbitrary binary signal

References

1. D. A. Huffman, The synthesis of sequential switching circuits, Technical Report 274, Research Laboratory of Electronics, M.I.T., Jan. 10, 1954.
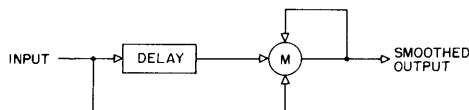

Fig. VIII-2. Representation of a smoothing element. The majority circuit, M, has a zero (unity) output if, and only if, two or three of the three inputs have a zero (unity) value

B. SOME REMARKS ON SCATTERING FROM EDDIES

In a paper with the same title to be published in the Proceedings of the I.R.E., the author gives a simple demonstration of the gaussian character of electromagnetic radiation scattered from refractive index fluctuations associated with pressure fluctuations in a turbulent atmosphere. The argument makes explicit use of the Central Limit Theorem of probability theory applied to dependent summands, unlike an earlier derivation (1).

It appears that coupling between eddies of nearly like sizes should appreciably enhance the average scattered power calculated by Villars and Weisskopf (2, 3). Since their calculations lead to values of the average scattered power which are considerably smaller than the experimentally observed values, the difficulty of the problem of determining the range of eddy coupling is especially cogent.

R. A. Silverman

References

1. R. A. Silverman and M. Balser, Statistics of electromagnetic radiation scattered by a turbulent medium, Phys. Rev. 96, 560 (1954).

2. F. Villars and V. F. Weisskopf, The scattering of electromagnetic waves by turbulent atmosphere fluctuations, Phys. Rev. 94, 232 (1954).

3. F. Villars and V. F. Weisskopf, On the scattering of radio waves by turbulent fluctuations in the E-layer of the ionosphere, Proc. IRE, forthcoming issue.

C. A DISCUSSION OF RATES FOR RANDOMLY VARYING CHANNELS, INCLUDING AN UNSOLVED CASE

Consider a channel the instantaneous state of which depends on an index which is a random variable. (Communication under fading conditions is a physical realization of such a channel.) Thus, for example, the binary channel with transition probabilities

$$\left\|\begin{matrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{matrix}\right\|$$

is generalized to

$$\left\|\begin{matrix} \alpha(s) & 1-\alpha(s) \\ \beta(s) & 1-\beta(s) \end{matrix}\right\|$$

where the transition probabilities are now functions of the random variable s, the "state of nature." We shall assume that the noise has no memory on the transmitted symbols, so that we should be able to calculate quantities on a per-symbol basis.

We shall confine our attention to the case of "uncorrelated fading," i.e., the successive states of nature are independent. A natural first problem is that of finding the rate of transmission of information in each of the four situations schematically represented as follows:

Case I.
$$\begin{matrix} S \\ \downarrow \\ T \rightarrow C \rightarrow R \end{matrix}$$

Case II.
$$\begin{matrix} S \\ \downarrow\searrow \\ T \rightarrow C \rightarrow R \end{matrix}$$

Case III.
$$\begin{matrix} \nearrow S \\ \downarrow \\ T \rightarrow C \rightarrow R \end{matrix}$$

Case IV.
$$\begin{matrix} \nearrow S \searrow \\ \downarrow \\ T \rightarrow C \rightarrow R \end{matrix}$$

The symbol T stands for transmitter, R for receiver, C for channel, and S for "state of nature." Case I represents the situation in which neither the transmitter nor the receiver knows nature's state (the parameter s), so that S is effectively just more noise in addition to that already included in C. In Case II, the receiver, but not the transmitter, knows nature's state (which conditions the channel transition probabilities). In Case III the transmitter, but not the receiver, knows nature's state, and in Case IV both the transmitter and the receiver know nature's state. Expressions for the rate (and capacity) in Case I and in Case IV are almost self-evident (see below). What is surprising is that, although an expression for the rate in Case II is easy to find, an expression for the rate in Case III has yet to be established.

Let us denote by $p(s, x, y)$ the probability that nature chooses the state s, that x is transmitted, and that y is received. We assume that $p(s, x, y)$ can be factored as

$$p(s, x, y) = p(s)\, p(x/s)\, p(y/s, x)$$

i.e., nature chooses the state s without reference to any action at the transmitter, x is chosen at the transmitter with a probability which may depend on nature's choice of

state if known, and y is received with a probability which is a function both of nature's state and the transmitted symbol. The trivariate distribution $p(s, x, y)$ can be summed to give various bivariate and conditional distributions. For example

$$p(y) = \sum_{SX} p(s, x, y)$$

$$p(x/y) = \sum_{S} p(s, x, y) \bigg/ \sum_{SX} p(s, x, y)$$

By the notation $< f(s, x, y) >$ is meant the average

$$< f(s, x, y) > = \sum_{SXY} p(s, x, y) \, f(s, x, y)$$

where the sum is over all triples of nature's states s, transmitted symbols x, and received symbols y. In terms of this notation, we can justify the following expressions for the rates in Cases I, II, and IV.

$$\text{Case I.} \qquad R_I = \left\langle \log \frac{p(x/y)}{p(x)} \right\rangle$$

$$\text{Case II.} \qquad R_{II} = \left\langle \log \frac{p(x/sy)}{p(x)} \right\rangle$$

$$\text{Case IV.} \qquad R_{IV} = \left\langle \log \frac{p(x/sy)}{p(x/s)} \right\rangle$$

(All logarithms are to the base two.) The corresponding capacities are obtained by maximizing R_I and R_{II} with respect to the a priori distribution $p(x)$, and R_{IV} with respect to the conditional a priori distribution $p(x/s)$. Before discussing the difficulties of Case III, we shall give some examples which illustrate the above theory. Detailed calculations are omitted.

Example I. Binary Symmetric Channel.

The channel

$$\left\| \begin{matrix} q & p \\ p & q \end{matrix} \right\|$$

is written

$$\left\| \begin{matrix} q & p \\ p & q \end{matrix} \right\| = q \left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| + p \left\| \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right\| \tag{1}$$

i.e., zeros are received as zeros and ones are received as ones with probability q (this state is designated as state α), and zeros are received as ones and ones are received as zeros with probability p (this state is designated as state β). $H(x)$ denotes the entropy $-\sum_X p(x)\log p(x)$; p_o is the probability with which a zero is transmitted; the corresponding probability for the state α is $p_{o\alpha}$, for the state β is $p_{o\beta}$; $\underset{\alpha}{\text{Max}}$ denotes the maximum of the expression that follows with respect to α. For this example, we find that

$$R_I = H(p + (q-p)p_o) - H(p)$$

with the corresponding capacity

$$C_I = \underset{p_o}{\text{Max}}\ R_I = 1 - H(p)$$

attained for $p_o = \frac{1}{2}$, a familiar result. When the receiver but not the transmitter knows nature's state, we have

$$R_{II} = H(p_o)$$

the capacity in this case is

$$C_{II} = \underset{p_o}{\text{Max}}\ R_{II} = 1$$

attained for $p_o = \frac{1}{2}$. In this case the state information removes all the equivocation. See Shannon's "correction channel" argument (1). When both the transmitter and the receiver know nature's state, we have

$$R_{IV} = q\ H(p_{o\alpha}) + p\ H(p_{o\beta})$$

The corresponding capacity is

$$C_{IV} = \underset{p_{o\alpha},p_{o\beta}}{\text{Max}}\ R_{IV} = 1$$

attained for $p_{o\alpha} = p_{o\beta} = \frac{1}{2}$, i.e., both channel α and channel β are symmetric, and attain capacity for an a priori probability of zeros equal to $1/2$.

Example II. Binary Asymmetric Channel.

The channel

$$\begin{Vmatrix} q & p \\ 0 & 1 \end{Vmatrix}$$

is written

$$\begin{Vmatrix} q & p \\ 0 & 1 \end{Vmatrix} = q \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} + p \begin{Vmatrix} 0 & 1 \\ 0 & 1 \end{Vmatrix}$$

i.e., in state α, which has probability q, zeros and ones are received without error, while in state β, which has probability p, both zeros and ones are received as ones. (Channel α is noiseless, but channel β is "completely noisy.") For this example we find that

$$R_I = H(qp_0) - p_0 H(p)$$

The corresponding capacity is

$$C_I = \underset{p_0}{\text{Max}} \, R_I = \log(1 + 2^{-H(q)/q})$$

attained for

$$p_0 = \left[q(1 + 2^{H(q)/q}) \right]^{-1}$$

If the receiver but not the transmitter knows nature's state, we have

$$R_{II} = qH(p_0)$$

with capacity

$$C_{II} = \underset{p_0}{\text{Max}} \, R_{II} = q > C_I$$

attained for $p_0 = \frac{1}{2}$. This last result means that since the channel is noiseless with probability q and the receiver knows when it is noiseless, q bits can be sent through the channel if zeros and ones are selected at the transmitter with probabilities appropriate to a noiseless channel. In Case IV we have

$$R_{IV} = qH(p_{0\alpha})$$

with capacity

$$C_{IV} = \underset{p_{0\alpha}}{\text{Max}} \, R_{IV} = q$$

attained for $p_{0\alpha} = \frac{1}{2}$. Note that the expression for R_{IV} does not contain $p_{0\beta}$, so that $p_{0\beta}$ can have any value between 0 and 1 without changing the rate. This is reasonable, since, whenever the channel is in state β, ones are received regardless of what is sent; the only possible advantage of a particular choice of $p_{0\beta}$ would be to help the receiver guess the state of the channel, which it presumably knows already.

Example III. Ternary Channel.
The channel

$$\begin{Vmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{Vmatrix}$$

is written

$$\left\| \begin{matrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{matrix} \right\| = \frac{1}{3} \left\| \begin{matrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{matrix} \right\| + \frac{1}{3} \left\| \begin{matrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{matrix} \right\| + \frac{1}{3} \left\| \begin{matrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{matrix} \right\|$$

i.e., one of the transmitted symbols is always noisefree, while the other two are indistinguishable. In state $\alpha(\beta, \gamma)$, which has probability $1/3$, the symbol corresponding to the top (middle, bottom) row of the channel matrix is noisefree.

For this example, we find that

$$R_I = \frac{4}{3} - \frac{1}{6}\Big[(1 + 3p_0)\log(1 + 3p_0) + (1 + 3p_1)\log(1 + 3p_1)$$

$$+ (4 - 3p_0 - 3p_1)\log(4 - 3p_0 - 3p_1)\Big]$$

The probability of the symbol corresponding to the top row of the channel matrix is designated by p_0, that of the symbol corresponding to the middle row by p_1, that of the symbol corresponding to the bottom row by $p_2 = 1 - p_0 - p_1$. The corresponding capacity is

$$C_I = \max_{p_0, p_1} R_I = \frac{1}{3}$$

attained for $p_0 = p_1 = p_2 = \frac{1}{3}$, as demanded by the symmetry of the average channel. If the receiver but not the transmitter knows nature's state, we have

$$R_{II} = \frac{H(p_0)}{3} + \frac{H(p_1)}{3} + \frac{H(p_0 + p_1)}{3}$$

with capacity

$$C_{II} = \max_{p_0, p_1} R_{II} = H\left(\frac{1}{3}\right) > C_I$$

attained for $p_0 = p_1 = p_2 = \frac{1}{3}$. When both the transmitter and the receiver know nature's state, we have

$$R_{IV} = \frac{H(p_{0\alpha})}{3} + \frac{H(p_{1\beta})}{3} + \frac{H(p_{2\gamma})}{3}$$

The corresponding capacity is

$$C_{IV} = \max_{p_{0\alpha}, p_{1\beta}, p_{2\gamma}} R_{IV} = 1$$

attained for

$$p_{0\alpha} = p_{1\beta} = p_{2\gamma} = \frac{1}{2}$$

$$p_{1\alpha} + p_{2\beta} = p_{0\beta} + p_{2\beta} = p_{0\gamma} + p_{1\gamma} = \frac{1}{2}$$

This last result means that when both transmitter and receiver can track nature's state, the transmitter should use the ternary channel as a noiseless binary channel by using the noisefree symbol of the moment as one symbol and the noisy pair of the moment as the other symbol. Of course, the noisefree symbol should be sent with probability 1/2, and the noisy pair with total probability 1/2, with the latter probability distributed in any way between the noisy symbols.

The reader will note the inequalities

$$R_I \leqslant R_{II} \leqslant R_{IV}$$

$$C_I \leqslant C_{II} \leqslant C_{IV}$$

which are immediate consequences of the fact that the entropy cannot increase when averaged (2).

We proceed now to a discussion of the problem of determining the rate in Case III, when the transmitter but not the receiver knows nature's state. We have not yet suc-ceeded in finding an expression for the rate (much less the capacity) in this case. The difficulty lies in the fact that when the transmitted symbols depend on nature's state, the received symbols contain information not only about possible transmitted messages but also about nature's state, and it is not clear how to separate these two components. In Case IV this difficulty does not arise, since in this case the receiver learns of nature's state directly through its private noiseless channel.

Case III is further complicated by the fact that maximizing with respect to the con-ditional a priori distribution $p(x/s)$ is not the only way the transmitter can enhance its rate. There seems also to be the possibility of maximizing the rate with respect to a set of noiseless transducers. This can be seen by considering Example I above, the symmetric binary channel. If the transmitter is driven by a message source with $p_o = \frac{1}{2}$, and if the transmitter knows nature's momentary state, it can produce error-free information at the receiver at the rate of one bit per symbol by simply reversing the designation of zeros and ones whenever the channel is in state β, without the receiver needing to know nature's momentary state.

Any expression for the rate R_{III} in Case III must satisfy two limiting conditions: 1. If the transmitter ignores nature's state, so that $p(x/s) = p(x)$, then R_{III} must reduce to the expression given above for R_I. 2. If the transmitter takes a unique action for every given state of nature, i.e., if for every s, $p(x/s)$ is zero for all x but a given $x(s)$, then R_{III} must vanish, since in this case all the information conveyed by the transmitted symbols is about nature's state and none is about pos-sible messages.

An expression for R_{III} which satisfies these two limiting properties is

$$R_{III} \stackrel{?}{=} \left\langle \log \frac{p(x/y)}{\sum\limits_{S} p(x/s)\, p(s/y)} \right\rangle$$

Unfortunately, not only has no adequate justification for this expression been given, but an apparent counterexample seems to be available.

Consider Example I (the symmetric binary channel) for the special values

$$q = \frac{2}{3}, \quad P_{0\alpha} = \frac{1}{2}, \quad P_{0\beta} = 1 \tag{2}$$

Then it is easily verified that 1/3 of the received digits are zeros and that 2/3 are ones. The received zeros always come from the zeros of some message; half of the received ones come from the ones of some message, the other half result from the fact that the channel is in state β, where only zeros (which are received as ones) can be transmitted. Thus the received text is made up of runs of ones interspersed with runs of zeros. To uniquely specify the transmitted sequence corresponding to a given received sequence, we need know only how many ones in each run of ones are "meaningful," i.e., correspond to the transmission of a one in the state α, and how many are "spurious," i.e., correspond to the compulsory transmission of a zero in state β. There is no need to locate the "meaningful" ones in the runs.

The probability of a run of r ones followed by a zero is

$$p(r) = \left(\frac{2}{3}\right)^{r} \left(\frac{1}{3}\right)$$

The probability of m "meaningful" ones in a run of r ones (followed by a zero) is

$$\binom{r}{m}\left(\frac{1}{2}\right)^{m}\left(\frac{1}{2}\right)^{r-m} = \binom{r}{m}\left(\frac{1}{2}\right)^{r} \tag{3}$$

The average entropy per symbol required to specify the number of "meaningful" ones in the runs of ones of the received text is thus

$$\sum_{r=0}^{\infty} \frac{p(r)\,H(r)}{(r+1)} \tag{4}$$

where $H(r)$ is the entropy of the binomial distribution in Eq. 3. A calculation shows that the sum in Eq. 4 is 0.298.

If the usual notions of additivity of information apply to Case III, we should have

(message information in received text)

+ (correction entropy) = (entropy of corrected text)

or, in particular,

$$R_{III} + 0.298 = 0.667$$

i.e.,

$$R_{III} = 0.369$$

However, if we substitute the values in Eq. 2 in the formula

$$R_{III} = \left\langle \log \frac{p(x/y)}{\sum_S p(x/s)p(s/y)} \right\rangle$$

we obtain

$$R_{III} = \frac{2}{3} + \left(\frac{1}{3}\right) \log\left(\frac{2}{3}\right) = 0.471$$

This discrepancy in the two values of R_{III} constitutes the counterexample.

The adjective "apparent" which appears in the phrase "apparent counterexample" reflects the author's uncertainty as to the applicability in Case III of the usual notions of additivity of information and "correction channels." Moreover, there is the possibility that some other correction scheme might require less entropy, although it is hard to see how this could be.

Thus the problem of deriving an expression for the rate (and the capacity) in the situation in which the transmitter knows more about the noise than the receiver is as yet unsolved, and the author hopefully awaits suggestions (or a solution!) from some reader of this note.

R. A. Silverman

References

1. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication (University of Illinois Press, Urbana, 1949), p. 36.
2. Ibid., p. 21.