Communication theory

1 — Information is finite

by D. A. Bell

University of Hull

A generation ago one might have said that *language* was one of the main features distinguishing man from the animals. But now it is known that most animals, from chimpanzees to bees, have systematic methods of communication by sounds and gestures; and the unfortunate person who is deaf and dumb (and therefore would a few centuries ago have been regarded as stupid) can communicate by "deaf and dumb language." All of this goes to show that communication can be effected by various means; and the superiority of human speech lies in its speed and flexibility which enable it to convey a very wide range of messages, including abstract ideas.

The introduction of the word *idea* is a cue to point out that the communication or information theory of engineers is not concerned with "ideas": it handles only "messages." This might sound like a severe limitation but in fact it is not, since any set of words, for example, can be regarded as a message; and the "set of words" might be the Bible, the collected works of Shakespeare, or the works of your favourite science-fiction author. By choosing a set of words we have made the number of possible messages finite in the mathematical sense though inconceivably large: there are some 35,000 words in an English dictionary so the number of different sets of, say, 100,000 words is rather more than 10 to the power of 400,000. If I assume that every reader has a copy of the Concise Oxford Dictionary (5th edition, 1964) I can represent any word by a code of the form $n_1\ a\ n_2$ where n_1 is the page number, a is L or R for left-hand or right-hand column and n_2 is the serial number of the word in the column. The opening words of this article would then be represented by: 1L2 509R4 26L13 544R7 767L6.

This is very clumsy and time-consuming as it means looking up every word in the dictionary (though I am sure one would soon get to recognise the codes for common words, like 1L2 and 544R7) but it has several noteworthy features:

(1) It reminds us that communication requires that sender and receiver agree on the code to be used, even if only on a common language.

(2) It is more precise than words. 767L6 in the dictionary reads "might^2. See MAY ^{1}", thus distinguishing it from "might^1" meaning great strength.

(3) It illustrates the point that words may be represented by all sorts of different symbols during the process of communication.

(4) From the sample given above it would appear that the typical length of a code group is 5 characters, which compares with 5 letters for an average English word. But 4 of the 5 characters are now numerals in the scale 0-9 and the fifth has only two values, L or R. So there is some economy.

It also makes it clear that we are talking about the kind of communication which consists in selecting in turn particular signals from a known set of signals or code; and the kind of information which can be communicated in this way is called selective information. Now most of the information we handle is of this kind: the current price of gold; which of the national contestants became Miss World; which premium bond drew a prize; which airline has just had a plane crash; what are the frequencies and times of BBC stereo broadcasts. These are all questions which can be answered by drawing a particular number or name from the range of numbers and names which was known to exist, and less specific or more complex information can be communicated by a more or less lengthy series of words selected from the dictionary. New ideas, on the other hand, cannot always be specified definitively by existing words or groups of words and may have to be assimilated gradually from the context in which new words or phrases are used. If I look in the dictionary for "meaning" I am referred to "significant" and vice versa. But under "bread" I find "Flour moistened, kneaded and baked, usually with leaven". Thus a concrete object can be broken down into its components or alternatively it can be described in terms of shape, colour, texture etc.; but an abstract idea like "meaning" can only be learned through experience of the way in which the word is used. It is also a prime principle of communication theory that one should not communicate information which was already known; this means that the amount of information transmitted is measured by the *increase* in amount of information possessed by the recipient. The method of measuring the amount of information will come later.

Communication is never absolutely certain. The hi-fi enthusiast may ask for "perfect" reproduction, but the engineer knows that at least there will be Johnson noise in the circuits, with power kTB* in bandwidth B. So the engineer must ask "How good is good enough?" Ask him for 60, 70 . . . . dB signal-to-noise ratio and he will tell you whether it is possible and how much it will cost; but ask him for perfection and he will either shake his head or decide for himself what standard the customer will accept as perfect. But if we are communicating only selections from a finite set of signals, it is obvious that the s/n ratio required is just enough to prevent one signal being mistaken for another. This idea is usually illustrated by the analogy of representing the several signals by points in space. (It has to be multi-dimensional space with a large number of dimensions.) These points have to be far enough apart that when the co-ordinates of one of the points are given then in spite of the noise in the system a seeker armed with the co-ordinates will arrive within reach of the desired point and of no other. The sort of practical problems to be solved by communication theory are therefore as follows.

(i) Given a set of messages (of known number) from which selections are to be communicated through a channel of given bandwidth and s/n ratio, what are the best shapes of signal to use to represent the messages?

(ii) With the conditions in (i), what will be the reliability of communication, or

*k = Boltzmann's constant and T = circuit temperature.

how should the conditions be altered to achieve some specified standard of reliability?
(iii) How does speed of communication tie in with everything else?

Ignoring derivations and proofs, we can answer questions (ii) and (iii) by quoting Shannon's key formula

$$C \leqslant W \log (1 + P/N) \qquad (1)$$

which is part of the following theorem: By a sufficiently complicated method of encoding it is possible to communicate information at any rate up to C through a channel of bandwidth W and ratio P/N of signal power to noise power with negligible risk of error. This is the channel capacity theorem. Note that this evades question (i) by postulating "a sufficiently complicated system of encoding." The hypothetical system of coding which allows the equality sign to be used in formula (i) is called "ideal coding." Much effort has been devoted to the search for coding methods which approach this ideal. Another point is that where we have loosely said "with negligible risk of error" one should ask "negligible in comparison with what?" To be precise, Shannon showed that the risk of error may be made as small as we wish by making the signals long enough in time. There are therefore advantages in putting the formula in symmetrical form

$$I \leqslant T\,W \log (1 + P/N) \qquad (2)$$

where I is the amount of information transmitted in time T.

T can be measured in seconds, W in hertz and P/N is a ratio (e.g. of watts); but we have not yet any measure of I. Now any information can be communicated, between two people using the same code book, by a sufficient number of yes/no questions. This was noted by Francis Bacon in 1623 when he devised a code in which each letter of the alphabet was represented by five binary symbols and said that "And here, by the way, we gain no small advantage, as this contrivance shows a method of expressing and signifying one's mind to any distance by objects that are either visible or audible – provided only the objects are but capable of two differences, as bells, speaking trumpets, fireworks, cannon etc."

A simple example is that about 16 binary decisions should suffice to locate any word in the Concise Oxford Dictionary if I start with first or second half, quarters, eighths . . . and finally down to fractions of a page. (I have to say "about" because the number of pages is not a power of 2 and the number of words per page is not uniform; the Dictionary was not designed for this exercise!) It follows that (selective) information can always be expressed as an equivalent number of binary units; and I in (2) is measured in bits or C in (1) in bits per second. But this is not the whole story. If a "sixteen questions" guessing game with the dictionary leads me to the top half of the right-hand column of p.943 I shall think that the word I am seeking is likely to be *pompous* or *pond*, but unlikely to be *pompano* or *pompier*, for example. So the measure of the amount of information which is communicated must take account of the pre-existing probabilities and not merely absolute certainties; and we now take the view that the amount of information communicated is related to the reduction in uncertainty or to the extent to which it allows a reassessment of probabilities at the receiving end of the channel. It can be shown mathematically that the only satisfactory measure of the uncertainty related to a finite group of probabilities is the *entropy*

$$H = -\sum_{i=1}^{N} p_i \log p_i \qquad (3)$$

where the p_i are the individual probabilities in a set of N distinct probabilities. Since probabilities are by definition less than unity, each $\log p_i$ is a negative quantity and H is positive.

Entropy has significance in thermodynamics and statistical mechanics, but the exact relationship between the different applications of entropy need not concern us. It suffices to say that entropy is always associated with ideas of disorder, confusion or indistinguishability of one state of a system from another. It is therefore natural to associate it with uncertainty and use reduction in entropy as a quantitative measure of information.

So far as we are concerned, H in formula (3) is just the weighted mean of all the logarithms of the probabilities, each logarithm being weighted with its own probability of occurrence, and it can be measured in bits. (Readers are probably familiar with the transformation from common logarithms ($\log_{10}$) to natural logarithms ($\log_e$) by multiplying by 2.3. Equally one can work in logarithms to base 2 and if the units in formulae (1), (2) and (3) are bits it must be understood that the logarithms are ($\log_2$). An important property of H is that it has a maximum value of $-\log p$ when all p's are equal and is zero if one probability is unity and all others zero. For if one probability $p_k = 1$, $\log p_k = 0$ and all the other $p_i = 0$; so $\Sigma p \log p = 0$ when one possibility can be selected with certainty.

For a simple application to a communication situation, suppose we are watching a Telex machine which we know is going to print a string of letters. Before a letter is printed there is a probability of 1/26 for each letter of the alphabet and $H = -\log_2 (1/26) = \log_2 26 = 4.7$ bits. If the letter Q is printed, $H = 0$ for this letter; and the information attributed to the communication of one letter is equal to the reduction of entropy of 4.7 bits. But if instead of "a string of letters" the Telex output was known to be English language text, the appearance of Q would be quite improbable but the appearance of E would be probable. This prior knowledge of probabilities constitutes information which we already have at the receiver and thereby reduces the amount of information which has been communicated. This is allowed for by recalculat-

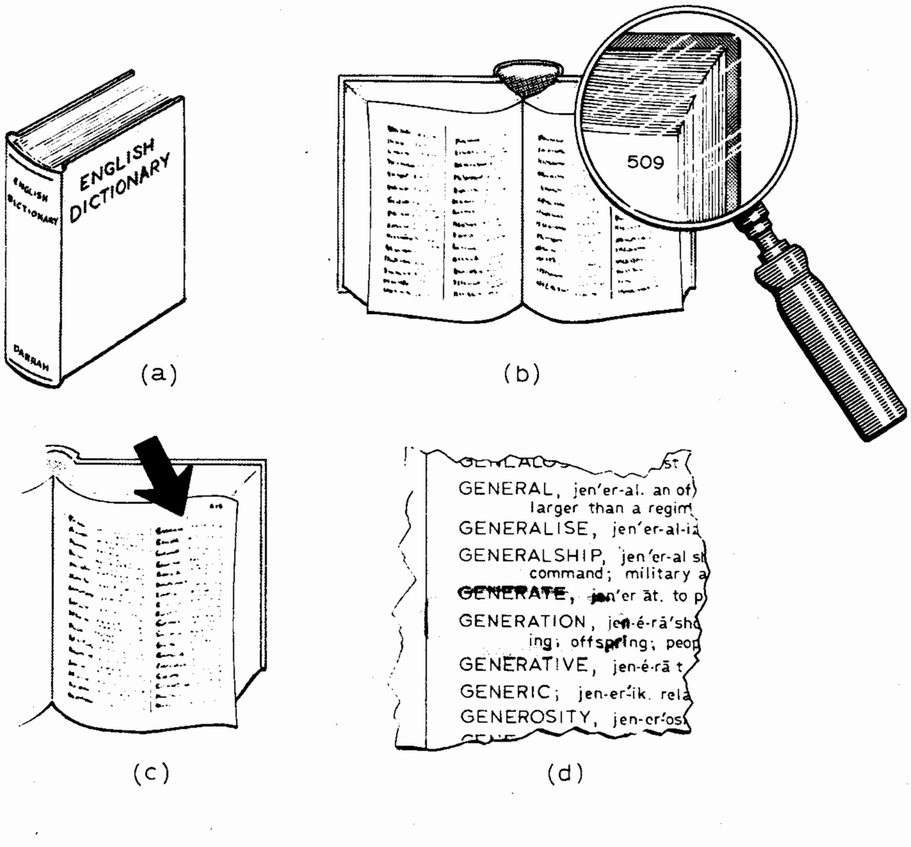


Fig. 1. Identification of a word: (a) the dictionary, (b) the page number, (c) the column, (d) the word.

ing the value of H before the letter was received, putting the English-language weighting for each letter in the formula

$$H = -\sum_{i=1}^{26} p_i \log p_i \qquad (4)$$

This will necessarily be less than the maximum value obtained when all the p's are equal and therefore its reduction to zero will represent less increase in information. (Actually the entropy of the English-language-weighted alphabet of 26 letters is reduced only to 4.3 bits per letter.)

But now let us look at the line engineer's view. Each letter is represented by five units (plus some synchronising pulses) and the receiving equipment must be set up with a threshold which decides between mark and space for each of the five units. Suppose the line is noisy so that there is a 10% chance that any one (but only one) of the units will be incorrectly interpreted. Then 5 letters which differ in one unit from the letter sent will each have a $(1/5) \times 0.1$ chance of being printed and the entropy after receipt of the noisy signal will look like this:

$$H = -\Sigma p \log p = -(5 \times 0.02 \log 0.02 + 0.9 \log 0.9) \qquad (5)$$

In binary units this is 0.701 bits. The information transmitted is the difference between the uncertainty before and the uncertainty after transmission, which in this case with English language is nearly 3.6 bits. So now we are able to measure the amount of information which is communicated even when noise in the channel means that nothing is certain. An important result of applying formula (3) to a binary channel ($N=2$) is that a 50% error rate means zero communication of information. For if when 1 is received the chances are 50 – 50 whether 0 or 1 was transmitted, one might as well toss a coin at the receiver and dispense with the communication channel.

Now we have admitted that there will always be noise in the communication channel. If it is random noise it may have any value of instantaneous amplitude up to infinity, but for just over two-thirds of the time it will not exceed the r.m.s. value. How can we reconcile this presence of occasional noise amplitudes which are many times bigger than the r.m.s. value with the channel capacity theorem?

That there is a real problem is shown by the following very crude and approximate interpretation of formula (2). If the signal-to-noise ratio is good, $1 + P/N \approx P/N$ and the *amplitude* ratio is approximately $\sqrt{(P/N)}$. The logarithm of the square root is half the logarithm of the original quantity, so

$$I \approx 2TW \log [\sqrt{(P/N)}] \qquad (6)$$

Now $2TW$ is the number of independent pulses that can be associated with the time-bandwidth product TW and in the

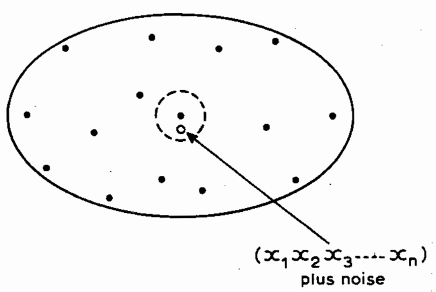


Fig. 2. The distance between signals must be greater than the likely effect of noise.

absence of noise digital information can always be expressed in the form

$$I_D = n \log S \qquad (7)$$

where n is the number of digits and S the number of states or amplitude levels for each digit.

Comparing (6) and (7), the channel capacity theorem seems to be saying that the number of amplitude levels can be spaced at intervals equal to the r.m.s. noise; but the instantaneous noise exceeds the r.m.s. value for about one third of the time, so how can errors then be negligible? The answer is in the first few words of the theorem "By a sufficiently complicated method of coding . . ." A proper derivation of the channel capacity theorem is fairly mathematical, and the further one goes in search of "ideal coding" the more one gets entangled in mathematics; but there are two principles which can be stated non-mathematically:

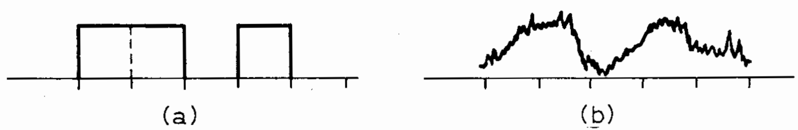
(1) Since the number of messages is finite, one has only to choose a finite set of signals which are sufficiently different from each other that even in the presence of noise one is unlikely to pick the wrong one. (This difference is often called the "distance" between signals.)

(2) A single instantaneous amplitude of noise may have a large value, but it is unlikely that a number of instantaneous values in succession will all have large values, and the more samples you take the nearer their average† will come to what we regard as the r.m.s. value of the noise.

An example of the second principle is

†Strictly speaking this "average" must be the root-mean-square value of the samples and what we normally call the r.m.s. value of the noise is that which we should obtain with an infinite number of samples.

Fig. 3. (a) signal transmitted, 11010; (b) signal received, 11010 or 11011?



that if you listen to the audio output from a high-gain receiver you will hear noise because the ear can respond to pulses lasting only one tenth of a millisecond; but if you connect an a.c. voltmeter with a response time of about a second, it will probably give a perfectly steady reading. This is because it will have averaged the noise over a TW product of about 10,000.

So "ideal coding" requires first that you construct signals with sufficient mutual differences (or distances) and second that you both construct signals which require a large value of TW and wait until the whole of a signal has been received before you try to identify it. Thus in principle ideal coding involves delay; but if W is of the order of kilohertz then T, and hence the delay, need only be of the order of a second to make TW large.

More recently the question has been put, "Supposing I do exceed the channel capacity defined by formula (1), how bad will the system be?" If we regard all differences between the received and transmitted signals as distortion, it is possible to formulate a relationship between the amount of such distortion and the rate of communication. The latter must take account of the fact that information is not received with certainty. For each received symbol one has only a set of probabilities of the various possible transmitted symbols; and in general different symbols may be made to have different probabilities of error. There results a rather complicated mathematical function called the *rate distortion function* which relates the rate of communication which can be achieved to a specified degree of distortion.

All that we have said so far about finite sets of messages seems to apply readily to telegraphy, where digital signals are natural, but what about telephony, television etc. when the signals are basically in continuous analogue form?

The answer is that continuous analogue signals may be reduced to discrete form by the two processes of quantizing in amplitude and sampling in time. No magnitude is ever known with absolute precision so it can always be equated to the nearest of a number of fixed levels if the latter are at close enough intervals. This process of equating to a pre-selected value is known as quantizing, and is no different from expressing a magnitude by a figure taken to a finite number of decimal places. The fineness of quantizing – the number of decimal places in the analogy – is chosen to give the desired accuracy. The other operation which is needed is sampling in time.

It was mentioned in connection with

formulae (6) and (7) that the maximum rate at which independent pulses can be transmitted through a channel is two per unit of time-bandwidth. This is often called the Nyquist rate, since it was stated by Nyquist in relation to telegraphy in 1928* An equivalent statement in very general terms due to Gabor** is that however one may try to construct a minimum signal element it will obey the law

$$\delta f.\delta t \geqslant ½ \qquad (8)$$

where the equivalent extent of the signal in bandwidth and time, δf and δt, is measured by a statistical formula which can be applied however fast or slowly the signal is cut off in frequency and in time. This is mathematically true because the frequency spectrum of a signal is the Fourier transform of its time waveform; but the cut-off points equivalent to this δf and δt do not correspond in any way with 3dB points. Gabor's theorem of the minimum signal is in close analogy with Heisenberg's principle of indeterminacy in physics, which is generally written as $\delta p.\,\delta q \approx h$ where h is Planck's quantum and p and q are a pair of conjugate co-ordinates of a particle such as its momentum and position.

The counterpart of the rule about pulse rate is that any waveform of which the Fourier components can be contained in a bandwidth W and of which the duration is T can be reconstructed unambiguously from $2WT$ suitably chosen samples. This is the sampling theorem. If the waveform corresponds to a low-pass band from 0 to W hertz, then evenly spaced samples at two per cycle of the highest frequency are suitable. (This is the form of the sampling theorem which is most commonly used. Other arrangements of $2TW$ samples are possible, and a different sampling pattern is needed for bandpass signals.) The original waveform is reconstructed if the nth sample of amplitude a_n causes the receiver to generate a unit waveform

$$a_n \frac{\sin \pi\,(2\omega t - n)}{\pi\,(2\omega t - n)}$$

This method of reconstructing the waveform is open to criticism in theory, though in practice it is good enough provided that TW is large. The difficulty is that the waveform $(\sin x)/x$ extends from $x = -\infty$ to $x = +\infty$ so no one of the waveforms used for reconstruction can be completely contained in the time interval T. But the function is small for x outside $\pm\ 4\ \pi$ so the imperfect reconstruction is noticeable only in the neighbourhood of the first and last samples, and this is unimportant if TW is large. Assuming for the moment that formula (1) is of general application, it says that for a given communication rate one can change bandwidth provided one adjusts P/N accordingly, and vice versa. This is a qualitative retrospective justification of systems like f.m. where for a given output the carrier signal-to-noise may be allowed to drop in exchange for the use of a greater bandwidth. The idea of exchanging bandwidth against signal-to-noise was not obvious while we were always thinking of hi-fi transmission of the original sound or other waveform. But it arises naturally from the Shannon approach of communicating signals from a finite and pre-arranged set instead of arbitrary waveforms.

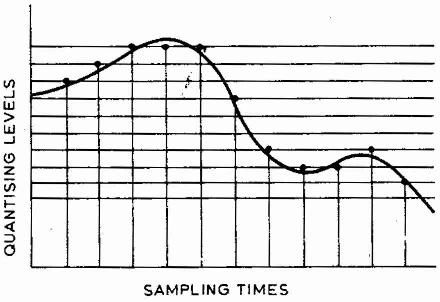


Fig. 4. Digitising a waveform.

Thus we have shown that information is an objectively measurable quantity; and in consequence communication channels can be designed in terms of the communication of information rather than of the faithful transmission of waveforms.

(Next article: redundancy and the exchange rate)

*H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. A.I.E.E.* vol. 47, p.617, 1928.

**D. Gabor, "Theory of Communication," *J.I.E.E., Part III*, vol. 93, p.429, 1946.

Sixty Years Ago

The following extracts from the April 1916 issue of *Wireless World* were drawn from an informative article by Wm. S. Purser entitled The Banjo – A Pastime for Wireless Operators. "One of the popular fallacies regarding the banjo is that one has to have a black face and sing nigger songs . . . . Some talk has been heard in the past of elevating the banjo, and playing classical music upon it . . . . The banjo may be regarded as symbolical of good fellowship . . . . When purchasing an instrument select a British-made *ordinary* banjo and you will have a reliable article which will stand any climate . . . . Do not be misguided by the expression 'Anything will do to learn on'. The banjo should have five strings . . . . Wireless operators and others going on voyages or to out-of-the-way places should purchase strings by the dozen. Having decided on your brand of strings, always get them from the same place . . . . After gut strings have been exposed to the sea air for a long time on the instrument they gradually turn green". Follow that.

Communication theory

2—Redundancy and the exchange rate

by D. A. Bell,

University of Hull

In the April issue we considered the finite nature of selective information, moved on to an entropy measure of information in the presence of uncertainty and cited Shannon's formula for the communication capacity of a noisy channel. But we omitted any mention of **redundancy,** which is perhaps the most difficult concept in communication theory. The dictionary account of "redundant" is

> "Superfluous (freq. of workers in industry), excessive, pleonastic; copious, luxuriant, full."

In a single, more colloquial phrase it means "More than the minimum necessary to do the job." But first one must ask what job is to be done in what conditions. Secondly although one may regard something superfluous or excessive as wasteful, one may sometimes feel inclined to pay a little extra for something which is copious, luxuriant or full; so has redundancy anything positive to offer in communication theory?

In answer to the first question, the job is to communicate information in the presence of noise. If the information is originally in discrete form, such as numbers or written characters, the minimum number of digits which will represent it unambiguously can be obtained from formula (7) in the April issue and if more digits or characters are employed than this the extra ones are redundant. For example, we think of an average English word as being made up of 5 letters from a 26 letter alphabet. Applying the formula we get $5\times \log_2 26 = 23.5$ bits. But it seemed in part 1 that an average English word could also be identified by 4 decimal digits plus one binary symbol, giving $I = 4 \log_2 10 + 1 = 14.3$ bits. Finally we said that by repeated binary division of the dictionary any word could be represented by not more than 16 binary digits. Any form other than the direct use of binary decisions thus requires more bits, in other words it introduces some redundancy.

But before condemning redundancy

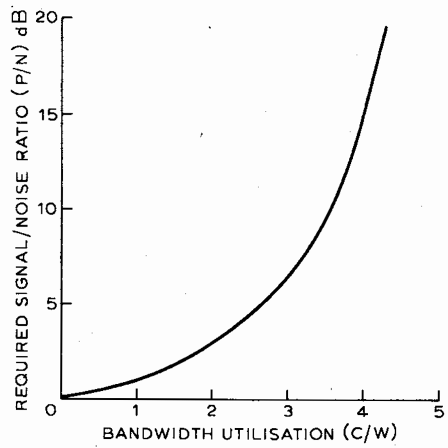


Fig. 1. Required signal/noise ratio, P/N, as a function of bandwidth utilisation: C/W=bits/s per hertz of bandwidth.

as wasteful, consider the possible effects of noise. If due to noise a signal transmitted as 509 R 4 (in our dictionary code April issue) were received as 508 R 4 the end result would be the world "gemmule" instead of "generation", which would make the phrase meaningless. On the other hand the result of changing one letter in "generation" say to "generption" would probably appear as a mis-spelling for which the recipient could guess the appropriate substitution. The redundancy in the spelling of English words has then served a useful purpose in making it possible to correct errors. If we are handling a passage in English language, as distinct from isolated English words, the context also helps. Experiments have shown* that in a passage in English about half the letters can be deleted before it becomes impossible to guess what the words were.

Words have to be pronounceable, which eliminates a large number of combinations. As a simple example, there are $26^3 = 17,576$ different combinations of three letters; but if we require one of the three to be a vowel the number comes down to $5 \times 26^2 = 3,800$ and we should have to include some combinations of four letters to cover as many as 17,576 words. Thus the exclusion of some combinations leads to an increase in average length of words which we recognize as redundancy. If conditions for spoken communication are very bad one may need to add further redundancy by spelling on a basis such as "A for apple, B for baker . . ." Years ago, when telephone lines were not always good, a telephone directory contained a list of such alphabet words headed "Aids to clarity of speech." What was meant, of course, was "aids to overcoming lack of clarity in speech."

So we see that redundancy has value in making it possible to detect or correct errors which may be caused by noise in the communication of information. The redundancy in English words is spread in rather an irregular way, so that it does not give the greatest possible error-protection in return for the extra length of signal involved. Over the last couple of decades a great deal of effort has been put into the design of codes for binary signals to allow the detection or correction of errors automatically. The redundancy of English words is normally utilised by a human operator at the receiving end of the channel, who mentally compares the received "word" with what he "knows to be right." A computer replacing the human operator would have to search the dictionary every time to find the nearest match of what it had received to an English word, and this would be a very slow and cumbersome business. The simplest error-detecting code for mechanised telegraphic use is the Van Duuren code, which represents each letter by a combination of seven binary digits instead of five. With the two extra digits providing redundancy one uses only the 35 combinations of the seven digits (out of a possible 128) which contain 3 mark and 4 space digits. Any single error (and in fact any odd number of errors) will upset the count of 3 mark digits and so will be detected, as will some combinations of multiple errors.

*C. E. Shannon, "Prediction and Entropy of Printed English." *Bell Syst. Tech. Journ.* vol. 30 (1951), p.50.

Among the more complicated and powerful codes are the Hamming codes for correcting single errors, the BCH codes which can be designed for correcting any number of errors and various others which have special properties for special purposes. The common feature of all of them, however, is that they employ more digits than would be necessary if one did not need error-protection. The increased number of digits is usually described as increasing the length of the signal, but this can be taken in two ways. If the bandwidth of the channel, and therefore the digit rate, is left unchanged, the larger number of digits will occupy a longer time; but if the digits are sent at a faster rate, so that the longer string of digits is sent in the same time, the channel will have to be increased in bandwidth. Either way the product TW will have been increased by the redundancy. One has to be careful, since increasing the bandwidth usually increases the noise; and unless the signalling power is increased in proportion, the capacity of the channel will be reduced through the P/N factor. It is simplest to assume that the signalling power is increased with the bandwidth to keep P/N constant. The signalling *energy* (power × time) is then increased in exactly the same ratio whether one allows T or W to increase to accommodate the redundancy.

When dealing with analogue signals there is no obvious datum from which to measure redundancy. One can only say that an increase in TW (or in P/N) for the communication of nominally the same information must represent an increase in redundancy. It is implied that if TW is to be increased then P/N is to be kept constant, or vice versa. Indeed one of the major consequences of Shannon's mathematical theory of communication is that signal-to-noise ratio can be traded against bandwidth, though in practice the terms of trade are rather one-sided. The channel-capacity formula

$$C \leqslant W \log_2(1+P/N) \text{ bits/s} \qquad (1)$$

tells us that for a given communication rate C we can vary W and P/N to any extent we like provided we keep the product of W and log $(1+P/N)$ constant. It is now seen to be untrue that the minimum bandwidth of a channel is fixed by the highest Fourier component at the input to the system: W can be made arbitrarily small provided we are willing to pay the price in increased P/N, but the price many be high. On transforming (1) to give a direct expression for P/N, instead of the logarithm of $1+P/N$, we find

$$P/N \geqslant 2^{C/W}-1 \qquad (9)$$

Now C/W is the communication rate in bits per second divided by the bandwidth in hertz, which is the number of bits per cycle. It could therefore be called the specific communication rate in bits/second per unit bandwidth.

Fig. 1 is a graph of P/N against C/W and shows that if we want to communicate at a rate of more than about two bits per cycle the signal-to-noise ratio required increases rapidly. It might be asked whether we should not plot P/N in decibels, whereupon the graph would approximate to a straight line for large values and the values of P/N in a range up to 13dB do not seem particularly high. In answer to the second point, it must be remembered that this graph represents the maximum performance which could be achieved with ideal coding, and any fairly simple practical system may need as much as 20dB better signal-to-noise ratio. The best reply to the first point is the comment of a radio engineer that decibels improvement in signal-to-noise ratio which are obtained by increasing transmitter power are "gold-plated decibels." An improvement of 10dB or 20dB may not sound impossible from the point of view of a receiver, but increasing the power of a radio transmitter from 1kW to 10kW or 100kW is not to be undertaken lightly. In some cable communication systems it is even more difficult to increase the input signalling power because of such limits as the insulation strength of the cable or the power-handling capacity of repeaters using solid-state devices.

However, Fig. 1 does not tell the whole story. In most systems the noise power increases in proportion to the bandwidth but Fig. 1 relates overall signal-to-noise ratio to channel utilisation in bits per hertz. A different picture results from relating signalling power to bandwidth at constant communication rate, and it is this which is relevant to considering bandwidth expansion or contraction schemes. The modified formula is derived in the Appendix and Fig. 2 is a graph showing change of necessary signalling power with change of bandwidth relative to binary signalling. This is, of course, considering only random channel noise and ignoring any effects such as quantizing noise which may be involved in any coding of signals for change in bandwidth.

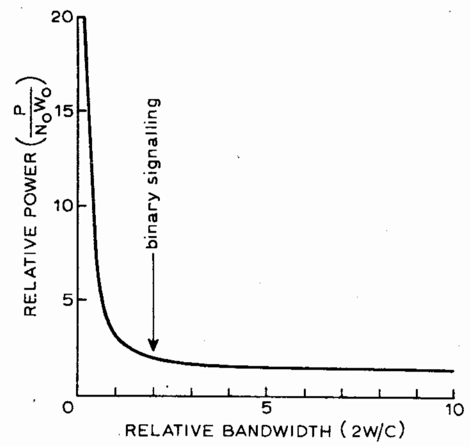


Fig. 2. Actual noise power is usually proportional to bandwidth. N_0 = noise power per hertz; W_0 = bandwidth needed for communicating C bits/s by binary signalling. With bandwidth W and rate C the relative bandwidth is represented by 2W/C.

It is clear from Fig. 2 why systems involving bandwidth expansion (for example f.m.) have been popular but there has been little genuine use of bandwidth compression. There was an early attempt to use f.m. with a very narrow frequency swing, with the mistaken idea that the bandwidth occupied would be limited to the narrow swing, regardless of the modulating frequencies. This was shown mathematically to be incorrect, as can easily be deduced in a non-rigorous way as follows. Any practically realisable† waveform which repeats at regular intervals can be represented by a Fourier series consisting of a fundamental frequency corresponding to the repetition rate and harmonics of that frequency. A wave modulated in frequency by a modulating frequency of n hertz repeats every $1/n$ seconds (neglecting the minor effect of carrier phase if the carrier frequency is not an exact harmonic of n); therefore all Fourier components must be multiples of n and the first sidebands must be distant from the carrier by amount n. The *minimum* bandwidth occupied is thus determined by the modulating frequency, and the actual bandwidth may be more if other sidebands are present in appreciable magnitude. We know, in fact, that the amplitudes of carrier and 1st, 2nd . . . sidebands are proportional to Bessel functions $J_0(x)$, $J_1(x)$, $J_2(x)$ . . . where x is the ratio of frequency swing to modulating frequency. Thus the first idea of using f.m., for reduction of bandwidth, was erroneous.

Then Armstrong introduced wide-band f.m. as a means of improving the received signal-to-noise ratio, but again on a fallacious argument, namely that noise consisted of changes in amplitude and could therefore be rejected by a limiter in a system using phase or frequency modulation. In fact the combination of a noise voltage with a signal voltage will change the phase as well as the amplitude in the resultant and the phase change due to the noise cannot be removed by a limiter. The explanation of the noise advantage of f.m. which is usually given is that its demodulation is a non-linear process; and this accounts correctly for all the facets of f.m. performance. It is, however, legitimate to point out that f.m. provides a form of redundancy, in that modulation by a single tone may produce several significant pairs of

†Functions which cannot be Fourier-analysed are mathematical monstrosities such as those having an infinite number of discontinuities or an unbounded range of variation which could not be realised in practice.

sidebands, whereas a.m. would produce only one pair. Since it also produces an increase in bandwidth for the same rate of communication (however the latter may be defined in the case of speech or music), it represents a move to the right on the graph of Fig. 2 so that the expansion in bandwidth should make it possible to reduce transmitter power.

Another aspect of redundancy, and one much nearer to the everyday use of the word, comes into play when one asks whether all the detail of a message is really necessary. Phoneticians reckon that there are only 38 distinct speech sounds or phonemes in European languages. The codes for the phonemes could be transmitted through a much narrower band than the full range of voice frequencies and a device for transmitting speech on this narrow-band basis is called a vocoder. It has the disadvantage that the reconstructed speech is very flat and impersonal and therefore it has never been put to much use. But it does raise the question whether in speech communication one wants only intelligibility of words or also shades of meaning conveyed by intonation and the details of voice sounds which are individual to each speaker**. This is a case for the customer, not the engineer, to decide what is redundant; and the customer's view is that nothing is redundant in a telephone conversation.

However, there are times when the engineer makes compromises which reduce the transmission of information without the customer being aware of it; and television naturally has the best examples of this sleight of hand. First of all there is interlaced scanning. The eye notices changes in the brightness of large areas as flicker if the changes occur at 25 per second, yet it cannot take in picture details in less than 1/25th second. So if the picture were scanned at 50 times per second half the information would not be taken in by the eye and to that extent would be redundant.

The scanning frequency is also related to the reproduction of motion in the picture, and a closer look at the structure of the video signal provides half the explanation of the miracle of colour television. (I call it a miracle, because if there were no redundancy in the system the transmission of the information needed to construct a colour picture of the same definition as a monochrome picture would need three times the bandwidth – yet colour pictures are being transmitted in the same bandwidth as monochrome.) A picture devoid of motion would produce a periodically repeating waveform consisting of harmonics of the picture frequency, though there is usually so little difference between odd and even lines that only harmonics of the field frequency are significant; and the separation of lines by synchronising pulses means that the harmonics which coincide with harmonics of the line frequency are much stronger. In fact it appears as though the spectrum consisted of line harmonics, each accompanied by its own sidebands spaced at field frequency. As shown in Fig. 3, these sidebands decrease in amplitude with distance from the line harmonic; and although the separation within each group is the field frequency, the groups associated with neighbouring line harmonics interleave, showing that all harmonics of the picture frequency are present. If, as commonly happens, movement affects only a minor part of the picture, the intensity of the spectral components between the lines is small. But if a spectrum consists only of narrow lines most of it is empty and it is very tempting to try to put something in between.

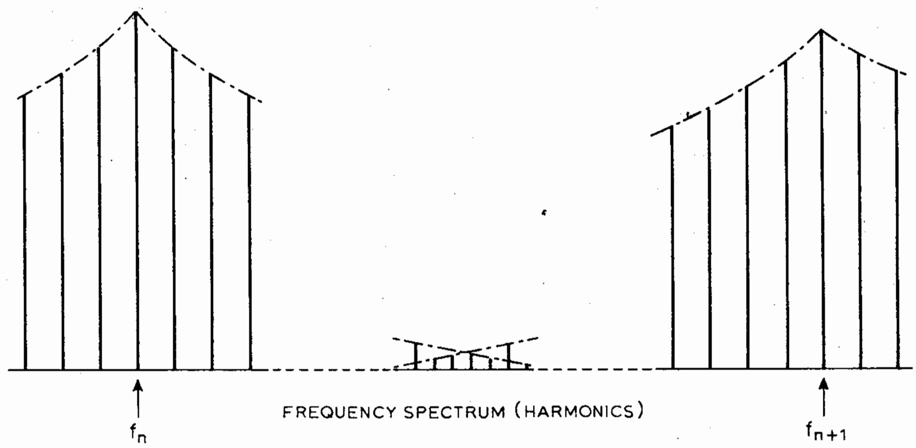


Fig. 3. Part of the spectrum of a typical television signal, showing the n^{th} and $(n+1)^{th}$ harmonics of line frequency. Each of these is accompanied by "sidebands" corresponding to field frequency and these overlap halfway between line harmonics.

The second criticism of the television spectrum, from the point of view of communication theory, is that it is not flat across the bandwidth occupied. Taken on the average for any reasonable waveforms, the amplitudes of the Fourier components are inversely proportional to their harmonic order, so that the video spectrum of a television signal almost always has an approximately inverse-frequency shape. This is obviously inefficient, because it means that if the signal-to-noise ratio is just sufficient at the top end of the band it will be much greater than necessary at the bottom. In the days of monochrome television this seemed to call for a top boost before transmission and a corresponding cut in the receiver, but this was never done. I believe the difficulty is that while theory can say "Taken on the average for any reasonable waveforms . . ." the television engineer must be prepared to handle the exceptional waveforms which produce strong high-frequency components, even though they occupy only a minute fraction of the programme time. Experience with the very modest amount of top boost which is normally employed in f.m. sound broadcasting (pre-emphasis) has shown that it very readily brings the danger of over-modulation.

However, this weakness of the monochrome spectrum has also been exploited in the transmission of colour pictures, since the colour information is placed near the top of the video band and with a subcarrier placed exactly half-way between line harmonics. The use of a subcarrier which is doubly modulated (with I and Q components) is an ineresting variant on the Nyquist-Gabor theorem that one can transmit two independent signal elements per unit of time-bandwidth. If one identifies sine waves of different frequencies by counting cycles over a time T the number of possible different frequencies within a band W is WT. But it is possible to distinguish between sine and cosine waves; so, when using both, the number of distinguishable signals is $2WT$. If you think about superimposing sine and cosine carriers, both independently amplitude-modulated, you will see that it comes to the same thing as modulating a single carrier simultaneously in both amplitude and phase, which is what one first thinks of as double modulation of a single carrier.

The rest of the trick of colour television depends on the engineer deciding that some of the colour information is redundant because the eye would not respond to it. The eye has maximum acuity for changes in brightness but less for changes in colour, so the high frequencies in the colour signal are redundant. The brightness or luminance signal must therefore be transmitted with full bandwidth, exactly as for a monochrome picture; while the colour information is transmitted with reduced bandwidth which confines it to that part of the spectrum which is left comparatively empty in the monochrome signal.

** In some military radiotelephone systems it is considered essential to provide enough bandwidth for a speaker to be immediately recognized as a particular person (e.g. Captain Smith or Lieutenant Brown). – Ed.

There is also a great deal of redundancy in television signals due to correlation between different areas of the picture. This has been known for many years, but it has not been practicable to make use of it until the recent developments in digital techniques which look like revolutionising all forms of communication. But that is the subject for a future article.

To summarise, an absolute measure of redundancy can only be obtained by comparing the communication rate which is actually achieved with the channel capacity which is calculated as a function of T, W and P/N. But more often a change in relative redundancy is thought of as a change in the TW product for a given amount of information, with P/N as an independent variable. It is in this sense that an increase in redundancy can be used to give protection against noise, as in the use of error-correcting codes in digital transmission or wideband f.m. for analogue transmission.

Appendix

If W is changed by a factor x, then N is also changed by a factor x and both can be expressed in terms of their standard or normalised values and x. It is usual to normalise N to noise per unit bandwidth, so that $N = W N_o$. Remembering that we want to work in terms of a constant communication rate it is convenient to take W_o, the normalised value of bandwidth, as that bandwidth which would accommodate C by binary signalling at the Nyquist rate. The noise power in this band, which is $N_o W_o$, is taken as the reference value for signal power. We now re-write formula (9) as

$$\frac{P}{N_o W} \geqslant 2^{C/W} - 1 \qquad \text{(i)}$$

Next put $W = x W_o$ and $W_o = C/2$ so that $C/W = 2/x$.

Formula (i) then becomes

$$\frac{P}{N_o W} \geqslant x(2^{2/x} - 1) \qquad \text{(ii)}$$

In this formula x is the factor of bandwidth expansion or contraction relative to binary signalling and with the equality sign this is plotted in Fig. 2.

(Next article: the digital revolution)

Addendum

Part 1 of "Communication theory" in the April issue should have included a footnote on p.44, middle column, referring to the quotation from Francis Bacon. The footnote should read: This was brought to the author's attention by a letter in *Computer Bulletin*, March 1968, by M. G. Farringdon.

Communication theory

3 — The digital revolution

by D. A. Bell

University of Hull

In speaking of the "digital revolution" I am not thinking of the proliferation of digital computers (though this has its impact on telecommunication through the need for data transmission) nor of the fact that a simple pocket electronic "digital calculator" is now in almost the same price bracket as a good slide rule, its analogue predecessor; I am thinking rather of the impact of digital techniques on television (mostly behind the scenes) and on the use of the Post Office telephone network for the communication of both data and speech, with developments in Viewphone as well. All of these digital developments depend on the flexibility which digital techniques allow in exploiting three key features of Shannon's communication theory, namely *redundancy*, the possibility of an *exchange between signal-to-noise ratio and bandwidth*, and a certain *symmetry between time and frequency* in the calculation of information rate and channel capacity.

The problem of television standards conversion became important in this country as soon as BBC2 was introduced with 625 lines while BBC1 continued to use 405 lines (see *Wireless World* May 1971.) Then the development of international exchange of television programmes, especially since communication via satellite has been available, has drawn attention to the fact that there are also differences between scanning standards used in different parts of the world: in North America they have 525 lines per picture and 60 (interlaced) fields per second, but in Europe we usually have 625 lines with 50 fields per second. (There are also differences in colour transmission, such as between N.T.S.C., PAL and SECAM.) Conversion between these standards requires storage, and storage of information is simple and reliable only when it is in digital form.[1] Since the development of l.s.i. (large scale integration) both semiconductor memories and shift registers have become comparatively cheap, so that the digital solution is also an economic solution. In the simplest case each picture element is represented by several bits (as described below for p.c.m.) and the number of bits is minimised by non-linear quantisation, using small steps for small signals but larger steps at each end of the scale.

There is also a great deal of redundancy in picture signals arising from the fact that neighbouring points are related to each other – they are said to be correlated. This correlation reduces the amount of information in the signal, but it is difficult to use because one must make provision for the occasional picture in which there is not much correlation – e.g. if the camera is panned across a scene. In most pictures, however, the difference between adjacent points in a line is much less than the full swing from black to peak white. It follows that for the same gradation of grey the signal conveying the difference between adjacent points will nearly always require fewer steps than would be needed to convey the absolute value of each point. So this *differential* p.c.m. (or d.p.c.m.) can use fewer bits per sample than direct p.c.m., which is particularly valuable if the picture has to be transmitted through a limited channel. The addition of vision to an ordinary telephone circuit ("Viewphone" in Britain or "Picturephone" in the Bell System, USA.) is a case where (a) economy of bandwidth is of great importance and (b) occasional momentary degradation of the picture may be tolerable. The addition of vision to an ordinary telephone channel may seem an extravagant luxury, but its supporters claim that it is justified by fuel economy if the addition of vision to the communication channel makes it an acceptable alternative to face-to-face meeting and so reduces travel. Bandwidth economy in television was the subject of a great deal of investigation some years ago[2] but in those days only analogue devices were considered for providing the storage from line to line or field to field which is needed to take advantage of the correlation in space between adjacent points or the correlation in time between the same points in successive scans ("picture difference" transmission)[3]. Now Post Office designers of Viewphone suggest digital storage of a line in shift registers[4] so that use may be made of correlations between points which are adjacent vertically, in different lines, as well as that between adjacent points in the same line. A more elaborate system has been proposed for the American Picturephone, utilising correlation between successive scans. This has a buffer store of 67,000 bits (equivalent to about 4,000 words storage for a minicomputer using 16-bit words); and if movement in the picture destroys correlation to such an extent that a scan cannot be compressed into this size, the picture is deliberately degraded in two stages. First the scan is coarsened to use only half as many picture elements per line; and if that does not suffice the second stage is to omit a complete scan and instead re-transmit the previous scan.[5]

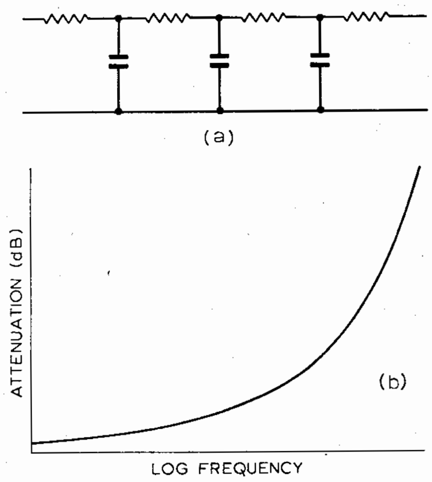


Fig. 1. (a) Lumped circuit approximation to an unloaded line. (b) Frequency response of such a network.

Another digital development in television is the transmission in the field blanking interval of alpha-numeric data. The system is called teletext (Oracle by the IBA and Ceefax by the BBC); and any viewer who has the appropriate adjunct for decoding, storage and character generation (as recently described in this journal) can choose at any time to have the picture on his screen replaced by a page of information consisting of latest news or any one of a hundred or so pages which may be available.

But in more mundane telephone communication, digital methods first came in with the use of p.c.m. on junction lines which are usually in the range of 10 to 30 miles long and employ multi-pair paper-insulated cables – the last kind of cable one would think of for the transmission of a wide frequency band or narrow pulses. These cables have very little inductance of their own so that ideally they would behave as CR circuits (Fig. 1a) with a frequency response which falls continuously for all except the very lowest frequency (Fig. 1(b).

The equivalent circuit of Fig. 1(a) with capacitors at intervals is only an approximation, since in fact the capacitance is uniformly distributed along the cable; but Pupin pointed out a long while ago that if one had inductance distributed along the line as well, or an approximation to this in the form of loading coils at intervals (Fig. 2(a), the cable became a lossy low-pass filter which passed all frequencies up to a definite cut-off but very little beyond. So most telephone cables have these loading coils at intervals so as to maintain a uniform pass-band with a signal/noise ratio better than 50dB which is then divided into telephone channels by frequency-division multiplex. Now a binary signal can tolerate a much lower signal/noise ratio and the signal after regeneration is independent of the signal/noise ratio. Therefore a binary signal can use a much wider frequency range in Fig. 1(b), and does not require the attenuation to be uniform within that range.

In p.c.m. we need to sample the analogue signal twice per cycle, but Nyquist says we can transmit two pulses per second per hertz of bandwidth: so there appears to be no change so far. But if we are to reduce the pulses to a binary scale, i.e. p.c.m., we need several pulses per sample; and in fact eight are used, seven to quantize the signal into 128 levels and one for synchronising and signalling. The bandwidth is increased eightfold; but the characteristic of the cable is such that the useable bandwidth increases more than eight times when one drops the signal/noise ratio from that required for a good telephone channel to that required for binary signalling, so there is a net profit on the exchange. In terms of communication theory, this is an example of trading signal/noise ratio against bandwidth.

Pulse code modulation is now in widespread use on junction lines.* It should be recognised that an important factor contributing to this is the availability of solid-state circuits for handling quantized signals which combine complexity of logic circuitry with small size, low power consumption and reliability. In the days of valve amplifiers the repeaters (line amplifiers) were usually housed in telephone exchanges, or occasionally in separate buildings. Such housing was necessary on account of their bulk, the need for a substantial power supply (with means of maintaining power in case of a mains failure) and the need for access for maintenance. Now, a solid-state regenerative repeater is put in whenever a loading coil is taken out. The loading coils were usually inserted at intervals of about a mile, and housed in cable-jointing chambers below manholes. A small solid-state regenerative repeater can be put there instead, and its small power requirement supplied along the telephone cable from the nearest exchange.

The spacing of repeaters is important. The signal must not be attenuated below the permissible signal/noise ratio

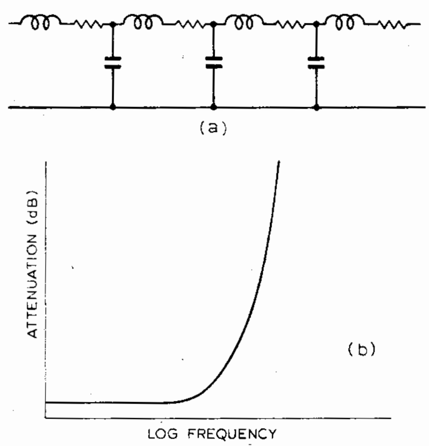


Fig. 2. (a) An inductively loaded line. (b) Bandpass frequency response of the loaded line.

before arriving at the next repeater; but the attenuation increases with frequency so that the maximum frequency which can be used depends on the closeness of spacing of repeaters.

For some thirty years engineers have been seeking an all-electronic telephone exchange. Note that I said "all-electronic" because the present generation of what are generally called "electronic exchanges" are really electronically controlled electromechanical exchanges. You must first understand that a telephone exchange has three functions: (a) to provide a switchable path for speech signals from an incoming to an outgoing line; (b) to receive, act upon and generate signals for controlling the speech path and (c) to interpret the signals and other data so as to provide speech paths efficiently. If we think of the old type of manual switchboard, (a) is provided by jack plugs and cords which can connect selected jack sockets, (b) depends on the senses and limbs of the operator (e.g. seeing a calling subscriber's light, switching ringing tone on a called subscriber's line) and (c) depends on the operator's brain.

In the past generations of automatic exchanges, (b) and what could be done towards (c) were implemented by relay logic, and this has been replaced by computer-like electronic circuits in the current generation of electronically controlled exchanges.[6] But there has been comparatively little success in the search for a solid-state replacement for the metal-to-metal contact in the speech-path switches, though the mechanical form of the contacts has changed from Strowger two-motion switches through cross-bar switches to the sealed reed relays used in the current electronically controlled exchanges.[7] The metal-to-metal contact has virtually infinite ratio of open to closed resistance, perfect linearity and no noise in normal functioning. Engineers became increasingly convinced that these characteristics could not be approached by any solid-state device and therefore solid-state switches could not be employed in an analogue speech path. But the use of digital signals such as p.c.m. makes the characteristics of the switches far less critical, so that solid-state switches would be useable.†

Once an analogue signal has become contaminated with noise, it can never be "cleaned"; and the noise components introduced at successive links in the route are additive. A digital signal, on the other hand, can be regenerated by threshold devices so that it leaves each repeater stage as a perfectly clean on/off signal. The Post Office therefore took the view that a future "System X", which will one day supersede the present generation of electronic-control - plus - reed - relay - switching exchanges, will employ digital transmission on the main trunk circuits. Preliminary trials on trunk transmission between Portsmouth and Guildford using high-speed p.c.m. have recently been initiated by the Post Office.[8] Again there has been an exchange of bandwidth against signal-to-noise ratio, since p.c.m. at 120Mbit/s is sent over a coaxial cable which could only be used up to 12MHz for f.d.m. The first consequence of this change to digital transmission will be that our familiar frequency-division multiplexing will be replaced by time-division multiplexing. Then instead of picking out a frequency channel with filters we shall pick out a time-slot by strobing. Some of the exchange switching can then be replaced by the matching of time-slots between incoming and outgoing lines, but some switching of circuit connections (known as "space switching") will probably be needed as well. But in view of the greater tolerance of digital signals, "cross-points"** could be solid-state.

* Lines 10 to 30 miles long which give direct connection between local exchanges, without using the trunk system.

† See section headed "Electronic exchanges" in "Centenary of the telephone", Wireless World, March 1976, p.92. — Ed.

** See Appendix for explanation of "cross-point".

A few years ago this was unquestioned and there seemed no hope for solid-state cross-points outside digital systems. Recently, however, there has been a revival of interest and a dozen manufacturers in the USA offer private automatic branch exchanges (PABX) using solid-state technology which offers the advantages of small size and little maintenance requirement. Most of them use four-layer devices, analogues to thyristors, as cross-points. But the point is that in this case the speech goes through only one solid-state switching system (or at most two if the subscriber at each end of the line has a solid-state PABX). This contrasts with the number of exchanges which may be involved in a long-distance call over a public network, and of course any noise or distortion arising in successive exchanges is additive.

An immediate advantage expected from digital transmission of speech is that data in binary form can immediately replace the p.c.m. signals. There has been much controversy over the question of whether we need a separate data network in addition to the telephone network. At present digital signals are usually fed into the frequency-division multiplexed analogue telephone system through modems (modulator-demodulators) which convert the on/off binary signal into a more or less complicated system of frequency-shift, phase-shift or multi-level amplitude modulation of an audio tone. It is much simpler to send binary digits straight to line, if the line is tailored to handle digital signals. Since the telephone network provides almost universal coverage, it seems uneconomic to set up another network alongside it just to handle data. So making telephoning digital confers a further advantage if it makes telephone traffic and data traffic fully compatible. (But it must be remembered that the telephone system is at times fully loaded, so that extra trunk lines and switching plant would be needed to enable it to carry a significant load of data traffic as well.)

A development in networks intended solely for data is called "packet switching." This arises because transmitting data has some of the characteristics of sending telegrams: there is no need for communication to be instantaneous and simultaneous with the origination of the information, though one would prefer that any delay in data transmission should be measured in seconds rather than hours. Given adequate buffering, data signals which arise at irregular intervals can be sent at a steady rate (most of the time) over the main transmission channel.

The idea is in fact very old since what we shall now call "a node in a packet-switching network" has as antecedent "a torn-tape relay station in a telegraph network." Telegrams are normally transmitted via punched paper tape and if they had to be sent through a relay station they could be received there on a tape perforator. Originally an operator would read the address on an incoming message and at its end tear off the length of tape and transfer it to an appropriate transmitting machine for its particular destination; but later the address was read by machine and the length of tape directed automatically into the appropriate channel for re-transmission, where it might join a queue of messages for that destination. The equivalent in modern technology is obviously to have at the relay centre a computer with appropriate buffer stores.

The torn-tape telegraph relay is not of great practical interest now because it is usually easier to switch electrical connections, so why has the principle been revived for data transmission? The first answer is that there is a considerable demand for transmitting small units of data, such as banking transactions of individual customers, which contain no more than a few hundred characters. At the present time it may take 15 seconds to set up a long-distance s.t.d. connection so if the data can then be transmitted in a similar or shorter time the whole process is rather inefficient. Secondly there is the problem of concentration of local traffic into trunk lines, which in telephony is achieved at present by frequency multiplexing. In the transmission of telephony by p.c.m. the sampling pulses are generated within the exchange, so that although there are problems of synchronism between exchanges in the proposed digital telephone network the timing is at least all controlled within the system. (Incidentally, one proposed method of matching time slots between different exchanges is called "pulse stuffing" and consists in inserting extra pulses into the signals from one of the channels. This clearly could not be used in data transmission).

So the idea is that data should be divided into units called "packets" of perhaps 1000 bits, each carrying a destination address. These would originate at sources of various types, ranging from a keyboard to a 48kHz line from a computer centre and would all be sent to a local switching centre or node on the special network. The computer here would take them into store, read the address, sort them into strings according to destination and transmit at high speed (of the order of 100 megabits/second) to other nodes in the network. Precautions have to be taken to see that no packet is lost and to see that the packets in a series constituting one message arrive in the right order.

At the present time there is considerable activity in the development of data networks. The first was ARPA, set up by the United States Department of Defense to link a number of computer centres which each contain one or more powerful computers. It has also some access points in Europe. Part of the argument for such a network is that a computer centre may specialise in some particular type of work for which its computer is particularly suitable on account of features such as the balance between fast-access store, backing store, input and output facilities, etc. It is then economic to send different kinds of work to different computers and a communication network serves this purpose. In Europe the emphasis is more on data collection, as implied in the explanation of packet-switching above. The British Post Office has opened an experimental packet-switched service on the route London-Manchester-Glasgow to assess the commercial value of such a system; the European banks are proposing a network called SWIFT for the rapid transfer of information about international banking transactions; there are plans for a European Information Network (EIN); and a EURONET is proposed to link all major scientific research establishments in Western Europe.

It seems a far cry from Shannon to digital television and packet-switched data networks. But the same principles apply throughout:

(1) All information is finite and can therefore be digitised.

(ii) The maximum capacity of any physical channel can be specified in terms of its bandwidth, signal-to-noise ratio and the time for which it is available.

(iii) Within the bounds set by (ii) it is possible to exchange signal-to-noise ratio against the bandwidth-time product.

No doubt we shall see many further applications of these fundamental principles as time goes by.

Appendix: Cross-bar and cross-points

The Strowger system of automatic telephony is based on its two-notion switch. This first moves vertically to one of 10 levels and then horizontally to one of 10 positions in the level so as to select one out of 100 contact pairs. This is a direct mechanisation of the action of the manual exchange operator in

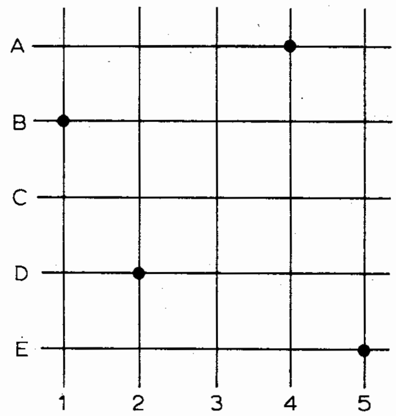


Fig. A. Principle of cross-point switch connecting any of inputs A, B, C, D, E to any of outputs 1, 2, 3, 4, 5.

selecting one out of a square array of jack sockets, but with one important difference. The whole of the Strowger switch remains engaged while the single call is passing through its contact field; but the operator after putting plugs in to make one connection can leave it and attend to other connections until she receives a signal to clear the first connection. The cross-bar switch, which is illustrated in principle in Fig. A comes nearer than Strowger does to simulating the methods of the human operator. The diagram shows input lines A, B, C, D, E crossed by output lines 1, 2, 3, 4, 5. At each point where the lines cross there is a switch, closing of which is indicated by a blob on the diagram. Thus the cross-bar switch can complete more than one circuit at a time: the connections shown are A→ 4, B → 1, D → 2 and E → 5. While the numbers of lines in each direction may be varied, the principle is now universal and the only question is what kind of switch or cross point should be used at each crossing. In the original cross-bar switches the contacts were metal-to-metal and operated by a mechanical system of solenoids, bars and latches; in the current electronically controlled exchanges they are reed switches; and there is an increasing probability of the future use of solid-state devices as cross-points in some applications.

References

1. J. L. E. Baldwin *et.al.*, "DICE: the first intercontinental digital standards converter," *R.Telev. Soc. Journ.* 1974, vol. 15, pp 140-159.
2. D. A. Bell, "Economy of bandwidth in television," *Journ. Brit. Inst. Rad. Engrs.*, 1953, vol. 13, p.447.
3. R. Barthelemey, "Images Differentielles," *Comptes Rendus*, 1949, vol. 229, p.61.
4. J. E. Thompson and G. A. Gerrard, "A differential p.c.m. encoder for Viewphone signals." Conference on Digital Processing of Signals in Communications, 1972, IERE Conference Proceedings No. 23, pp.329-337.
5. D. J. Connor, B. G. Hashell and F. W. Mounts, "A frame-to-frame Picturephone coder for pictures containing differential quantizing noise," *Ibidem*, pp. 363-372.
6. M. T. Hills "Electronic telephone exchanges. Conclusion: Computer-controlled Systems," *Wireless World*, July 1974, p.241.
7. M. T. Hills, "Electronic telephone exchanges. A Review of Techniques," *Wireless World*, June 1974, p.164.
8. "High capacity p.c.m. system," *Wireless World*, February 1975, p.92.