20A-118

LECTURES ON COMMUNICATION THEORY

D. GABOR

TECHNICAL REPORT NO. 238

APRIL 3, 1952

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 238 April 3, 1952

Lectures on Communication Theory

D. Gabor

of the Imperial College of Science and Technology, London

This report presents a series of lectures that were given under the sponsorship of the Research Laboratory of Electronics during the Fall Term, 1951, at Massachusetts Institute of Technology

Abstract

These lectures on selected chapters of communication theory are complementary to the well-known works of American authors on the statistical theory of communication, which is not discussed here at any length. About one-third of the lectures have as their subject the theory of signal analysis or representation, which precedes the statistical theory, both logically and historically. The mathematical theory is followed by a physical theory of signals, in which the fundamental limitations of signal transmission and recognition are discussed in the light of classical and of quantum physics. It is shown that the viewpoints of communication theory represent a useful approach to modern physics, of appreciable heuristic power, showing up the insufficiencies of the classical theory. The final part of the lectures is a report on the present state of speech analysis and speech compression, with suggestions for further research.

I. Introduction

I.1. What is Information?

Communication theory owes its origin to a few theoretically interested engineers who wanted to understand the nature of the goods sold in communication systems. The general answer has, of course, been known for a long time. Communication systems sell information capacity, as power systems sell energy. The way from this general idea to the quantitative definition of the concept of information was a long one, and we are not by any means at its end.

The first step in dispersing the cloud of vagueness which hangs around the concept of information is the realization that information, if it is to be communicable at all, must be of a discrete nature. It must be expressible by the letters of the alphabet, adding to these, if necessary, mathematical symbols. In general, we must make use of an agreed language, in which the process of reducing our chaotic sensations to a finite number of elements has led to some sort of vocabulary. How this vocabulary came to exist, how it is enriched from day to day by new concepts which crystallize in the form of a word, and how it is learned by children – these are dramatically interesting questions, but outside the range of communication theory.

Once we have a vocabulary, communication becomes a process of selection. A selection can always be carried out by simple binary selections, by a series of yeses or noes. For instance, if we want a letter in the 32-letter alphabet, we first answer the question "is it or is it not in the upper half?" By five such questions and answers we have fixed a letter. Writing 1 for a "yes" and 0 for a "no", the letter can be expressed by a symbol such as 01001, where the first digit is the answer to the first question, and so on. This symbol also expresses the order number of the letter (in this example, the number 9) in a binary system.

By the same method we can also communicate quantities. Physical quantities can be measured only with finite accuracy, and if we take as the unit of our measure the smallest interval which we can assert with about 50 percent confidence that the quantity in question is inside it, we can write the result ...010011 $\pm$ 1. Alternatively we can also use a "decimal" (really "binary") point. Any measured number must break off somewhere. It is true that there are numbers, such as $\sqrt{2}$, or π, which do not break off, but in these cases the instruction to obtain them can be communicated in a finite number of words or other symbols. Otherwise, they could never have been defined.

This suggests immediately that the number of digits, that is, the number of yeses and noes by which a certain statement can be described, should be taken as a measure of the information. This step was essentially taken by Hartley in 1928 (1). It may be noted that if n such independent binary selections are carried out, the result is equivalent to one selection of $N = 2^n$ possibilities. Since $n = \log_2 N$, the informative value of a selection from N possibilities appears as the logarithm of N to the base 2. This is unity for one binary selection. The unit is called one "bit" or "binit", short for "binary digit".

The word "possibilities" suggests an extension of this definition. The N selections are all possible, but are they also equally probable? Of course, the answer is that in general they are not. It may be remembered that we are conversing in a certain language. Whether this language is as full of clichés as the lyrics of the music halls, or as full of surprises as the report on a racing day, there will always be certain features which we can predict with more or less certainty from what has gone before. But what we knew before, we evidently cannot count as information.

The extension of the information concept to the case in which we have certain expectations regarding the message has been made by N. Wiener and C. E. Shannon. They chose, for good reasons, the unexpectedness of a message as a measure of its informative value. Consider, for simplicity, a set of 1...i...N possible events which we think may happen, and to which we assign expectation values p_i. In general, the p_i present a thorny problem (for instance, if they represent the probability of the horse i winning a race). They become simple only in the so-called "ergodic" case, in which the next event is picked out of a homogeneous statistical series. In this case we take the past experience as a guide and identify the probabilities p_i with the frequency of the occurrence of the event i in the past, in similar trials.

Assume that we know the probabilities p_i of the events i in such an ergodic series. Using the language of communication theory, let us consider these events as "symbols" delivered by an ergodic "source". By Shannon's definition, the expectation value of the information is

$$H = -\sum_{1}^{N} p_i \log_2 p_i \text{ bits per symbol} \tag{1}$$

an expression which is also called "the entropy of the source". This definition, in order to be acceptable, must satisfy certain postulates. The first is that in the case of equal probabilities it must go over into Hartley's definition; that is, for $p_i = 1/N$ we must have $S = \log_2 N$, which is easily verified. Likewise one can verify also that if the events in question are composite (consisting of two or more independent events), the S are additive. Shannon also shows that in whatever way the events are broken down into component events, with their respective probabilities, the result is the same. A further important property of S is that it becomes a maximum (with the auxiliary condition $\Sigma p_i = 1$) if all p_i are equal (2). The full justification of the definition (Eq. 1) and its importance are revealed by Shannon's fundamental coding theorem: If a noiseless channel has a transmission capacity of C bits per second, codes can be constructed which enable the transmission of a maximum of C/H symbols per second. Thus, with ideal coding, a symbol supplied by a source with entropy H is completely equivalent to H bits, as compared with $\log_2 N$ bits of a source which delivers its N symbols at random and with equal probability.

Equation 1 is identical in form with one of the definitions of physical entropy in statistical mechanics. It may be recalled that H is a quantity which depends on the frequencies with which the symbols are emitted by the source of information, such as a speaker, using the English language, talking of a certain field, having certain habits, and the like. The connection between this expression and the entropy of a physical system is by no means simple, and it will be better to distinguish them, following a suggestion by D. M. MacKay, by calling H the "selective entropy".

I.2. What is the Value of Information?

The quantity H defined in Eq. 1 has proved its great importance for communication engineering in Shannon's general theory of coding. It is, in fact, the only satisfactory definition from the point of view of the communication engineer faced with the problem of utilizing his lines to the best possible advantage. While we shall not go into the theory of coding, since these lectures will be devoted to other features of communication theory, we must warn against mistaking H for a general "measure of information value" and forgetting the qualifications which we have added to it: "Expectation value of the information in the next event in an ergodic series".

The quantity H is a fair measure of what the customer pays for if he buys information capacity, but it would be rash to say that this is what he pays for if he wants to buy information. A moment's consideration will show that what people value in information, in everyday life, contains two new elements, exclusivity and prediction value, in addition to the element of unexpectedness.

A newspaper editor will pay for news if it is a "scoop", that is, exclusive, and if, in addition, it suddenly opens a vista of the future, such as the news of war preparations or the imminent outbreak of a war, which is the key, in the mind of the reader, to a vast field of well-prepared associations (3). (There are, of course, other elements in popular appeal; the field of associations need not extend towards the future, but can merely evoke strong personal experiences in the reader's mind. This, however, need not interest us here.)

A real estate speculator may be prepared to pay well for a piece of "inside information" on the direction in which a town is planning to extend. But let us consider rather a less criminal example – a racing tip. This is simpler and yet contains all the essential elements.

Let there be 1...i...N horses running, whose probability of winning is assessed by popular opinion as p_i. The bookmakers have adjusted the odds roughly in inverse proportion to p_i, so that the expectation value of a dollar returned on any horse is one dollar and the expected gain nil. There would be no betting unless people had their own "hunches". Consider, for simplicity, somebody who has no hunch and is in complete uncertainty. He receives a "tip" which, through rather subjective mental processes, induces him to assess the probabilities of winning as p_i^m (m for "modified") instead of p_i. He now expects for one dollar laid on the i-th horse a gain of

$$(p_i^m / p_i) - 1 \qquad (2)$$

dollars. He can pick out the maximum of this expression, lay his bet on the corresponding horse, and reasonably pay the tipster a commission in proportion to the expected gain. (Though he would be even more reasonable if he paid only after the race.)

This simple example illustrates what information may be worth if it modifies the expectations one step ahead. It also contains the element of exclusivity in the assumption that the odds have not changed, which may remain true so long as the backer does not lay a too heavy bet; then he automatically passes on information.

Let us now generalize this simple case somewhat by considering the case in which we want to look three steps ahead in a probability chain of three sets of events, i, j, and k. Let p_i be the probability of the event i; p_{ij} the probability that after i has happened, it will be followed by j; p_{ijk} the probability that the sequence i, and j will be followed by k. Since we know that one or the other event is bound to happen in each step, we have the relations

$$\Sigma p_i = 1; \quad p_j = \Sigma p_i p_{ij}; \quad \Sigma p_j = 1; \quad p_1 = \Sigma\Sigma p_i p_{ij} p_{ijk}; \quad \Sigma p_k = 1 \qquad (3)$$

The odds are laid according to the *a priori* p_k, and for simplicity we will again assume that they do not change. If, now, we receive a tip on the first step, which modifies the expectations to p_i^m, the expected gain on the event k will be

$$(\Sigma\Sigma p_i^m p_{ij} p_{ijk} / p_k) - 1 \qquad (4)$$

The tip in the first step will be most valuable if it restricts the choice in the third step of the chain and produces a prominent maximum. In general, we cannot assess its value by a single measure, but we have to consider the whole modification which it produces in the probability distribution three steps ahead.

It is evident that any single measure of the "peakedness" of a distribution must to some extent be arbitrary, but if we have to select one, there can be no doubt that the entropy is the best. It has been mentioned previously that the expression

$$H = - \sum_{1}^{N} p_i \log_2 p_i \qquad (1)$$

has a maximum if all p_i are equal, and is zero only in the case of certainty, that is, if one of the p_i is unity and all the rest zero. Adopting this, we have only to compare the modified entropy

$$H^m = -\sum p_k^m \log p_k^m \qquad p_k^m = \sum\sum p_i^m p_{ij}\, p_{ijk} \qquad (5)$$

with its value before the modification by the information, whose value we can then assess as H/H^m.

These examples, which can be easily generalized, are sufficient to illustrate what we mean by the "prediction value of information". It might appear that what we have discussed is the rather unscientific question: "How much do we pay the tipster or the informer?" But the problem has a more serious side. Information is something which enables us to take action, sooner or later. However, there are, often enough, situations so confused that it is impossible or impractical to take action on every bit of information received. In these cases, a criterion which sifts information according to its prediction value and allows us to reserve action for the worth-while cases may be of practical importance. It appears that little work has been done in this direction (4).

The idea that information is something which enables the receiver to take action which is in some way gainful to him is slightly more general than the prediction value, since it also comprises immediate action, and certain action, not based on probabilities. The information in the railway timetable enables me almost certainly to get a train to a destination; the information in a textbook, if it is sufficiently comprehensive, enables the student almost certainly to pass his examinations. An interesting suggestion for a physical measure of the action which a receiver can take on the basis of information comes from Szilárd (5). He measured the information received by a system A from a system B by the reduction of entropy, that is, the reduction of disorder which it enabled A to effect on B. This is of great interest, and we are coming back to it later. At this place a warning may not be untimely. Thermodynamics is the science of the physical interaction of systems which are in almost complete disorder. Forgetting this fact might lead to considerable confusion in the minds of students of electrical engineering. In communication we are usually dealing with systems in only very partial disorder, so that the limits set up by thermodynamics are often of very remote interest only. Szilárd has shown that at the thermodynamic limit a binary selection is worth a reduction of entropy by k log 2, where k is Boltzmann's constant, two-thirds of the heat capacity of, say, a hydrogen atom. To believe that this is what any "yes" or "no" is worth (for instance, the "yes" or "no" which decides a war) would be rather more than absurd!

I.3. Classification of Communication Theory.

The purpose of communication theory is, first, to convey an understanding of communication processes and, second, to suggest improvements in the handling of information. This includes not only communication, but also classification and storing. Taken in this wide sense, communication theory appears to encroach, on the one hand, on such old-established branches of learning as epistemology and mathematical logic; on the other hand, it borders on the techniques of communication. Rather than annex old sciences to this new branch of science, it will be better to restrict the discussion to those chapters which are more or less new, and yet have reached an appreciable degree of development. These can be represented in the following scheme:

Theory of Signals	Signal analysis.	
	Determinate transformations of signals.	Theory of Coding
Statistical Theory of Communication	Random transformations. Signal correlations.	Theory of Coding
	Theory of discrimination.	Cybernetics
	Acting on the basis of information received.	Cybernetics

Signal analysis has as its main purpose the nonredundant description or "representation" of signals, that is, the eliminating of those data which do not convey new information, but which could have been inferred *a priori* from a knowledge of the instruments used. What this type of *a priori* information means may be illustrated by Eddington's famous "Parable of the Fishing Net".

An ichthyologist wants to explore the life of the ocean, and he comes to the conclusion that no sea-creature is less than two inches long. He could have known this, without doing any fishing, merely by noticing that he has used a net with two-inch meshes. Similarly, if a meteorologist wants to investigate lightning and other atmospheric surges with a receiver of 10-kc bandwidth, he need not be surprised if he finds that no surge has a base of less than 100 μsec.

The same information, that is, the same set of new data can appear in an infinity of different forms, either by deliberate encoding or by unwanted but known distortions. This is the subject of the theory of determinate signal transformations.

It is convenient to exclude noise from signal analysis, since this makes for a certain unity of the mathematical methods. Thus, in our classification (which is, of course, a matter of personal opinion) probability has no place in signal analysis.

Statistical Theory of Communication. In all physical communication processes signals suffer unavoidable, unpredictable distortions by what we generally call "noise". This is one of the new elements which enter into the statistical theory. The other is: signals which are not connected by any rigid law (functionally) may be correlated statistically. This means that in the long run there will be less novelty in a message than one would infer from the *a priori* enumeration.

The theory of coding is astride the previous system of division. It has a determinate branch and a statistical branch. The second is mainly the creation of Claude Shannon. Its purpose is to eliminate the redundancies in the message and to create, as nearly as practicable, a code in which every signal is a novelty, so that the channel capacity is fully utilized.

The theory of discrimination, chiefly the work of Norbert Wiener, has as its purpose the picking up of signals from a background, making use of statistical knowledge both on the signal and on the background. It is one of the main objects of interest, but will not be treated in this course.

Cybernetics, also from Wiener, can be loosely defined as the science of purposeful action on the basis of information received. In the previous classification it is shown to overlap the theory of discrimination for two reasons. One is that, of course, information is the basis of action. The other is that part of the action may be itself directed toward an improvement of the discrimination.

Of these branches of communication theory only signal analysis, including noise theory, will be treated in some detail. Special attention will be given to the problem of speech communication, hence the analysis of speech sounds will receive an important part of the space.

In addition, some subjects will be treated which are closely connected with communication theory but which are partly outside the scheme at the beginning of this section. This scheme might give the impression that communication theory is applied mathematics. We talk of applied mathematics in fields in which the physical basis is so secure and well-known that it is hardly worth mentioning. This, however, is by no means the case in all problems of communication theory. Especially when the problem is to decide what is physically possible and what is not, it is necessary to take into very careful consideration the physical limitations; not only those of classical physics but those of quantum theory.

It has been pointed out, first by D. M. MacKay (6), that the points of view of communication theory are useful guides in designing and understanding physical experiments, and the habit is spreading of talking of information theory instead of communication theory when discussing its consequences on scientific method. There was reluctance from some sides (also on the part of the author) to accept this somewhat pretentious name for a new branch of science of which the outlines have just become visible. More recent investigations, however, which will be discussed in the fourth lecture of this course, are going some way towards justifying the claim that the principles of communication theory are powerful heuristic tools when applied to physical problems. They have yet to prove their worth in leading to new discoveries, but there can be little doubt that they are useful guides for understanding modern physics.

Probably the most interesting discoveries may be expected from the application of communication theory to the human nervous system, not only as "terminal equipment", but as a study in itself. In fairness, it must be pointed out that much of the credit will have to go not so much to the theory as to the techniques of electrical communications. The telephone exchange was for a long time a favorite model of neurophysiologists; more recently the modern calculating machines and closed loop systems have provided new models. There is no need to worry much about the distribution of credits, since both the techniques and the theory have been mostly provided by electrical engineers, who are richly rewarded by seeing an astonishing renaissance of their old craft.

II. Signal Analysis.

II.1. The Nature, Scope, and Purpose of Signal Analysis.

The breaking down of quantitative experience into discrete elements will be achieved in two steps. In the first step we will specify the degree of freedom of the communication channel, which means that we will count the number of independent data in a message. In this first step each datum is supposed to have an exact value. In the second step we will determine the number of "yeses" and "noes" contained in each datum, in the presence of noise, and of other sources of uncertainty.

We will first discuss those communication processes which we are accustomed to consider as "one-dimensional": the transmission of a quantity s in time. We do this only in order to establish continuity with habitual methods of thinking; it will be seen later that "dimensionality" is a somewhat arbitrary concept if the result is, as it must be, a finite set of data.

The orthodox method starts with the assumption that the signal s is a function s (t) of the time t. This is a very misleading start. If we take it literally, it means that we have a rule of constructing an exact value s (t) to any instant of time t. Actually we are never in a position to do this, except if we receive a mathematical instruction; but in this case it was not s which was sent, but the instruction. If s is being sent, two difficulties arise. One is the definition of the "instant of time", in such a way that there shall be complete agreement between the transmitter and the receiver. But if there is a waveband W at our disposal, we cannot mark time more exactly than by sending the sharpest signal which will get through W. This, however, as will be shown later, has a time-width of the order $1/W$; hence we cannot physically talk of time elements smaller than $1/W$. The other difficulty is, of course, the exact definition of the value of s. This is outside the scope of signal analysis as we define it and will be dealt with later.

Since signal analysis deals with a finite number of independent data only, it follows that after having specified these, there is nothing else left in the signal, and all physically meaningful mathematical operations can be carried out without any further reference to the signal "s (t)". This was only a sort of crutch to the mind accustomed to analytical thinking, and can now be thrown away. In fact, this is done, for example, in the steady-state analysis of linear circuits, where, after specifying that the class of time functions admitted in the analysis is that of the harmonic functions of time, all further discussion is carried out in terms of their complex amplitudes.

In signal analysis we cannot throw away the crutches so easily, since a very wide class of functions will be admitted, but it may always be remembered that we are dealing with their data only, not with the time functions. What then do the time functions mean?

The answer is that they establish the connection of signal analysis, as a branch of mathematics, with physics. They are instructions to construct apparatus for producing them or handling them. For instance, a harmonic signal may be produced by a pendulum, or by a rotating disk, or by a Thomson-circuit; linear circuits (physical instruments made of copper, iron, insulators, and so forth) are required for transforming it without distortion, that is, without changing the function. If we have the rules for constructing these devices, the concept "time function" becomes a label for these, but has no other physical meaning. Evidently it is difficult to do one's thinking in such epistemologically correct but practically unhandy terms, and the confused concept of time as being "something that exists, but which we cannot measure exactly" cannot be easily discarded by the human mind. There is no harm in it, so long as we remain aware of the limits to which it can actually be measured, that is, in which it has a physical meaning in the given circumstances.

By the fact that it deals with a finite number of data, signal analysis is most closely related to interpolation theory among the branches of mathematics. But while modern interpolation theory insists on full analytical rigor, we will entirely disregard rigor, not only because we want to hand on the results to practical people, but also because rigor is pointless, in view of the intermediate character of signal analysis. In order to make the results physically meaningful, we shall have to add "noise" to the data. Hence, in signal analysis, we need consider only functions which are continuous and differentiable beyond any limit, because any function can be approximated by such a function to any finite limit, however small.

The problems of signal analysis fall naturally into two classes. In the first class, the signal is "dead"; it has happened before the analysis starts, and all its data can be assumed to be known. This is the case, for example, if the message is handed to us in the form of a telegram. In the second class the signal is "alive"; it is known only up to the present instant, as in telephony. Attention will have to be given to both cases, but they will not be dealt with in separate sections.

II.2. Fourier Analysis in Infinite, Semi-Infinite, and Finite Intervals.

Linear signal analysis is the representation of arbitrary signals as series of a set of elementary signals, with suitable coefficients. Nonlinear analysis has not yet been developed to any appreciable extent, though it may become of great importance, for example, in problems of speech communication.

The historically first, and still practically the most important set of elementary functions is that of sine waves, which may be infinite at both ends, or bounded at one end, or at both. It is convenient to start from analysis in terms of infinite waves, considered by Fourier, though this is not the historically first case.

Let s (t) be a function known in the infinite interval. By Fourier's integral theorem we have the two reciprocal relations

$$s(t) = \int_{-\infty}^{\infty} S(f)\, e^{2\pi j f t}\, df \qquad \text{II.1}$$

$$S(f) = \int_{-\infty}^{\infty} s(t)\, e^{-2\pi j f t}\, dt \qquad \text{II.2}$$

By these relations the function of time s (f) has been "imaged" or transformed into another function S of a new variable f, which is called the frequency. s (t) and S (f) are a pair of Fourier transforms. It is useful to call S (f) the "left-hand" transform of s (t), and s (t) the "right-hand" transform of S (f), since S is found at the left, and s at the right, in the standard tables of Campbell and Foster (7). S (f) is also called the "complex spectrum" of s (t). In general, it will be complex even if s (t) is real, but for real signals we have the condition

$$S(f) = S^*(-f) \qquad \text{II.3}$$

The asterisk stands for the complex conjugate. Thus, if the spectrum for positive frequencies is known, the negative frequencies are not needed.

From Eqs. II.1 and II.2 one derives at once

$$\tfrac{1}{2}[s(t) + s(-t)] = \int_{-\infty}^{\infty} S(f) \cos 2\pi ft \, df \qquad \text{II.4}$$

$$\tfrac{1}{2}[S(f) + S(-f)] = \int_{-\infty}^{\infty} s(t) \cos 2\pi ft \, dt \qquad \text{II.5}$$

These are a pair of completely symmetrical transforms. There is no "left-hand, right-hand" difference, but they give only the even part of the signal or of the spectrum. This, however, is completely sufficient if the signal is known only in a semi-infinite interval, for example, for $t < 0$. In this case we are free to imagine it continued in the range of $t > 0$ as an even function, and we can write

$$s(t) = 2 \int_{0}^{\infty} S_c(f) \cos 2\pi ft \, df \qquad \text{II.6}$$

$$S_c(f) = 2 \int_{-\infty}^{0} s(t) \cos 2\pi ft \, dt \qquad \text{II.7}$$

where s (t) and S_c (f) are a pair of cosine Fourier transforms. For a real signal the cosine spectrum is real and an even function of the frequency, hence negative frequencies can be altogether disregarded. One can similarly define a sine spectrum by odd continuation, but this is far less convenient.

Thus if the signal is known up to a time t, we can define the "up-to-date cosine spectrum" as

$$S_c(f,t) = 2 \int_{-\infty}^{t} s(\tau) \cos 2\pi f(t-\tau) \, d\tau \qquad \text{II.8}$$

One can say that complex Fourier analysis regards the signal *sub specie aeternitatis,* when it is all dead, while the up-to-date analysis applies to a live signal whose past is completely unforgotten.

Let us now consider, with R. M. Fano (8), an analysis in which the past is gradually forgotten (see Fig. 1) because it is weighted with an exponential function.

$$e^{-\alpha(t-\tau)} \qquad (\alpha > 0)$$

The up-to-date spectrum of the weighted signal is called by Fano the "short-time spectrum", but we will call it the "perspectivic spectrum", since "short-time spectrum" will be reserved for other applications. The exponentially weighted signal can be easily realized, as Fano has shown, by an RC circuit. Fano's analysis can also be considered as analysis in terms of damped harmonic waves, all damped retrogressively by the same exponential factor.

Let us now consider another special way of "weighting the past", by a function which leaves the signal unaltered in an interval T preceding the present instant and entirely suppresses what has gone before that. In principle, we could proceed exactly as before: introduce into Eq. II.8 for

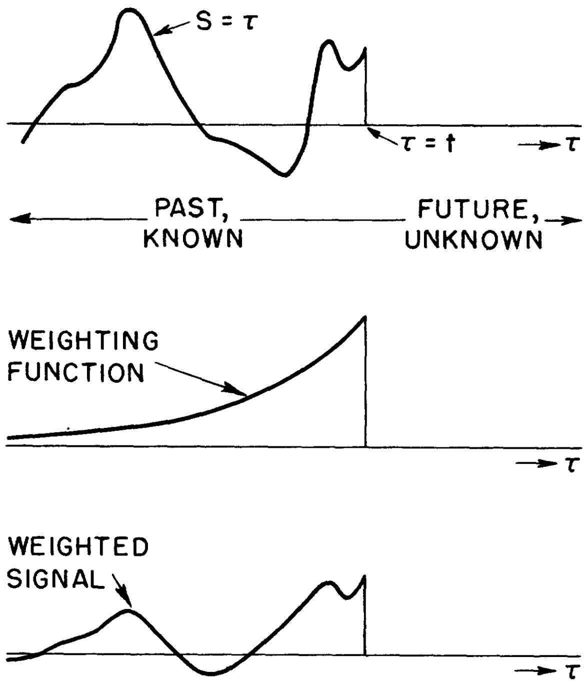

Fig. 1.
Perspectivic weighting of signals.

the up-to-date cosine spectrum this new, discontinuous weighting function. (The reader may carry this out as an exercise.) It may be remembered that the up-to-date spectrum S_c is the Fourier spectrum S of the signal, if it is continued into the future as an even function. In the present case we thus get the spectrum of a signal which is symmetrically repeated around the instant t in an interval $\pm$ T, and is suppressed outside. Evidently one can think of many other rules of continuation. Perhaps the most natural is that proposed by Fourier himself: continue the signal as a periodic function. Again it may be a useful exercise for the reader to prove that this assumption, introduced into Eqs. II.6 and II.7, leads to the representation of the signal as a Fourier cosine series

$$s(t) = \sum_{n=0}^{\infty} S_n \cos(\pi n t/T) \qquad \text{II.9}$$

$$S_n = \frac{2}{T} \int_{-T}^{0} s(t) \cos(\pi n t/T)\, dt \qquad \text{II.10}$$

This series has the period 2T, because Eqs. II.6 and II.7 contained the assumption of even continuation. Dropping this and going back to Eqs. II.1 and II.2, one obtains Fourier's sine-cosine series, with period T.

We can thus represent a signal, by chopping it into intervals T and writing down for every interval the set of Fourier components. This is not a very elegant method, but it is a useful one, which we shall later apply repeatedly in problems of speech analysis. Instead of the up-to-date spectral function, which is continuous and changes continuously with time, we now have a discontinuous line spectrum, which changes abruptly when we step from one interval T into the next. Both are equally valid and complete specifications of the same signal.

II.3. The Energy Spectrum and the Autocorrelation Function.

There are important cases in which the Fourier spectrum is either not defined physically, or without interest, while the concept of energy spectrum or power spectrum still remains valid and important.

The spectrum of light as taken with a spectrograph is a clear example of the first case. The phase of light waves is unmeasurable, except at extremely high intensities, which never occur in practice, but their intensity is easily found. Noise is an example of the second case. The phases are not unmeasurable, but irrelevant. Acoustical communication, where the ear is the final receiving organ, is another example. The situation here is a little more complicated, and we shall return to it later.

If the complex Fourier transform S (f) of a signal is known, the "spectral energy density" is, by definition, the absolute square of S (f) or

$$S(f)\, S^*(f)$$

If the signal is real, this can be written by Eq. II.3 as

$$S(f)\,S(-f)$$

In the case of cosine transforms, which are even functions of f, this is simply the square of S_c. In every case we have the energy theorem

$$\int_{-\infty}^{\infty} s^* s\, dt = \int_{-\infty}^{\infty} S^* S\, df \qquad \text{II.11}$$

(This theorem, often called by the name of Parseval, who established it for series, is really attributable to Rayleigh, 1889.) It is interpreted as the equality of "energy" in the time description and in the Fourier description. (Some caution is needed, as ss* does not always exactly coincide with the physical definition of power.)

Thus if we know the Fourier spectrum of a signal, we also know its energy spectrum, but the converse is not true; the phases have been lost, and an infinity of different signals may have the same energy spectrum. It is of interest to inquire what these signals of equal power spectrum have in common in the time description.

We get an answer to this if we transform the energy spectrum back into time language. To do this we require a rule for the transforms of products, called "the theorem of resultants". Thus for complex transforms (Eqs. II.1 and II.2)

$$\text{The left-hand transform of } s_1(t)\, s_2(t) \text{ is } \int_{-\infty}^{\infty} S_1(-x)\, S_2(f+x)\, dx$$

$$\text{The right-hand transform of } S_1(f)\, S_2(f) \text{ is } \int_{-\infty}^{\infty} s_1(x)\, s_2(t-x)\, dx \qquad \text{II.12}$$

Apply this rule to the case $S_1 = S$, $S_2 = S^*$; moreover, assume a real signal for which Eq. II.3 holds. There is now no difference between right-and left-hand transforms. Both become Fourier cosine transforms, and the result is:

The spectral density of energy SS* is the Fourier cosine transform of the convolution integral

$$\Psi(\tau) \equiv \int_{-\infty}^{\infty} s(t)\, s(t+\tau)\, dt \qquad \text{II.13}$$

and vice versa. This function $\Psi(\tau)$ we will call the autocorrelation function of the signal. Some caution is needed, since this term is familiar in the theory of stochastic processes; but there the autocorrelation function is defined as

$$\phi(\tau) \equiv \lim_{T=\infty} \frac{1}{2T} \int_{-T}^{+T} s(t)\, s(t+\tau)\, dt \qquad \text{II.14}$$

and this $\phi(\tau)$ stands in the same relation to the power spectrum $P(f) = SS^*/2T$, as our $\Psi(\tau)$, defined by Eq. II.13, to the energy spectrum:

$$\phi(\tau) = \int_{-\infty}^{\infty} P(f) \cos(2\pi f \tau)\, df \qquad P(f) = \int_{-\infty}^{\infty} \phi(\tau) \cos(2\pi f \tau)\, d\tau \qquad \text{II.15}$$

This last theorem, from N. Wiener (1930), was also discovered independently by Khintchine in 1934 and is often called the Wiener-Khintchine Theorem. Its importance is in its application to chaotic processes, where no Fourier spectrum can be defined. This theorem is a cornerstone of the statistical theory of communication, but we shall not need it in signal analysis, where the formally similar theorem has been known for a much longer time. We will use the autocorrelation function as defined by Eq. II.13 with finite energy because we always speak of signals of finite duration only.

The autocorrelation function and the energy spectrum are two equivalent descriptions of the same thing, but in many cases one may be more practical than the other. For instance, the two-dimensional autocorrelation functions of X-ray diffraction diagrams, called "Patterson diagrams" or "vector diagrams", are much used by crystallographers because they make it possible to pick out, at one glance, all the periodicities which occur in a crystal. Statisticians have started to use autocorrelations in time series which are too short for forming a reliable function $S(f)\, S^*(f)$, which they call a "periodogram", and they find that significant periodicities can be picked out more reliably from the peaks of the autocorrelation function than from the peaks in the periodogram. There are also indications (9, 10) that the "short-term autocorrelation function" or "correlogram" has advantages for the recognition of speech sounds over "short-time spectra" or "sonograms".

In defining a "short-term autocorrelogram" we have the same freedom of choice as in the matter of "short-time spectra", that is, we are free in the choice of our perspectivic weighting function. Once we have chosen this, the only reasonable way is to define the short-time correlogram as the convolution integral of the weighted signal. This will always be the cosine Fourier transform of the similarly defined short-time energy spectrum.

One might ask whether the convolution integral of the Fourier spectrum might not also be an interesting quantity, since it would show up periodicities in the spectrum, which, one would suspect, might play some part in speech recognition. But the answer is in the negative. This convolution integral would be the cosine Fourier transform of the squared signal, and it is known that squared or two-way rectified speech loses most of its intelligibility. This reveals a difference between time and frequency which is physically obvious, but which is somewhat obscured by the deceptive symmetry of Fourier's formulas.

II.4. Analytical Signals.

We now return to the study of Fourier analysis in the infinite time interval, in order to introduce certain new concepts which will be useful later on. We have seen that in general, that is, excepting even functions of time, the transform $S(f)$ of a real signal $s(t)$ is complex and extends over both positive and negative frequencies. One of these is redundant, since for real signals $S(f) = S^*(-f)$. We could simply ignore the negative part of the spectrum, but it may be more instructive to show how we can suppress it.

In the complex Fourier analysis every harmonic component appears in a form

$$\cos \omega t = \tfrac{1}{2}(e^{j\omega t} + e^{-j\omega t}) \qquad \sin \omega t = \frac{1}{2j}(e^{j\omega t} - e^{-j\omega t}) \qquad \text{II.16}$$

That is to say, it has one component at $+\omega$, one at $-\omega$. The electrical engineer is accustomed to visualizing this as the sum of two vectors, of equal value, rotating in opposite sense. But it is well known that one can consider an oscillating vector also as the projection of one rotating vector:

$$\cos \omega t = R(e^{j\omega t}) \qquad \sin \omega t = R(-j\, e^{j\omega t}) \qquad \text{II.17}$$

This means that we can consider the real signal

$$s(t) = a \cos \omega t + b \sin \omega t$$

as the real part of the complex signal

$$\psi(t) = s(t) + j\sigma t = (a - jb)\, e^{j\omega t} \qquad \text{II.18}$$

This contains one rotating vector only, and we have produced it by adding to the real signal one in quadrature to it. Evidently we could now apply the same process to any arbitrary signal $s(t)$ of which we know the Fourier transform, and carry out the corresponding operation on every Fourier component. There is no need, however, to go into the Fourier transform first, since the signal in quadrature to a given signal s (t) can be directly calculated by the formula

$$\sigma(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} s(\tau) \frac{d\tau}{\tau - t} \qquad \text{II.19}$$

(This is an improper integral; it is to be understood as taken over the indented real axis, that is, around the pole at $\tau = t$. The integral thus taken is also called the "Cauchy principal value".) It is easy to verify this by showing that it converts $s(t) = \cos \omega t$ into $\sin \omega t$ and $\sin \omega t$ into $-\cos \omega t$. Conversely s (t) can be expressed as

$$s(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \sigma(\tau) \frac{d\tau}{\tau - t} \qquad \text{II.20}$$

s (t) and $\sigma(t)$ are a pair of Hilbert transforms (11).

By this method we can convert any real signal into one with the same real part, but without a negative spectrum. The study of these complex signals was started by the author (12), and was continued by J. Ville (13), who gave them the name of analytical signals.

Apart from signal analysis, Eqs. II.19 and II.20 are also of interest because they give an answer to the questions, "Why can we not add or subtract frequencies? Why do we always get both sidebands in heterodyning?" The answer is that we could, in fact, shift the whole spectrum in one direction, but only if we could construct the in-quadrature function to the signal, and this we could do only if we knew its future!

II.5. Specifying a Signal by its Moments.

We have started by considering the signal as a function. The Fourier-series description of a signal in a limited interval was a first example of expressing it by an enumerable set of numbers (the coefficients of the Fourier-components). We now discuss another method, which, in principle, is applicable to any signal so long as its energy is finite. This is the specification of a signal by its moments. We profit from our previous results by carrying out the discussion simultaneously and symmetrically in time language and in frequency language.

Consider a signal $\psi(t)$, which may be real or complex or, as a special case, analytical, and its complex Fourier transform $\phi(f)$. These are connected by

$$\psi(t) = \int_{-\infty}^{\infty} \phi(f)\, e^{2\pi j f t}\, df \qquad \phi(f) = \int_{-\infty}^{\infty} \psi(t)\, e^{-2\pi j f t}\, dt \qquad \text{II.21}$$

If $\psi(t)$ is analytical, the integration over f can be restricted to positive frequencies only. The notations $\psi(t)$ and $\phi(f)$ have been so chosen as to emphasize the similarity of the formulas which follow with those of wave mechanics. Analytical signals have an advantage in this connection, as in wave mechanics the frequency is considered a positive quantity.

We now introduce two kinds of moments: in time and in frequency. For the weighting function of the first, we adopt $\psi^*(t)\,\psi(t)$, the "instantaneous power"; for the second, the "spectral energy density" $\phi^*(f)\,\phi(f)$. Adopting an arbitrary zero point in time we write down the double sequence of moments.

$$M_t^{(o)} = \int \psi^* \psi \, dt \quad M_t^{(1)} = \int \psi^* t \psi \, dt \ldots M_t^{(n)} = \int \psi^* t^n \psi \, dt \qquad \text{II.22}$$

$$M_f^{(o)} = \int \phi^* \phi \, df \quad M_f^{(1)} = \int \phi^* f \phi \, df \ldots M_f^{(n)} = \int \phi^* f^n \phi \, df \qquad \text{II.23}$$

Writing the conjugate first and the power of t or f between the other two factors are conventions in wave mechanics; the reason for this rule will be seen in a moment.

We now define mean epochs (epoch means "time of occurrence") t_n of different orders n by

$$t_1 = \overline{t} = M_t^{(1)}/M_t^{(o)} \quad (t_2)^2 = \overline{t^2} = M_t^{(2)}/M_t^{(o)} \ldots (t_n)^n = \overline{t^n} = M_t^{(n)}/M_t^{(o)} \qquad \text{II.24}$$

and similarly, mean frequencies f_n of different order by

$$f_1 = \overline{f} = M_f^{(1)}/M_f^{(o)} \quad (f_2)^2 = \overline{f^2} = M_f^{(2)}/M_f^{(o)} \ldots (f_n)^n = \overline{f^n} = M_f^{(n)}/M_f^{(o)} \qquad \text{II.25.}$$

It may be noted that the phase drops out in $\phi^* \phi$, hence the frequency moments do not completely specify the signal, though they specify its energy spectrum or its autocorrelation function. On the other hand, the two rows of moments taken together completely specify the signal, under very general assumptions.

The mean frequencies of even order form a monotonically increasing sequence. They are equal if, and only if, the signal consists of a single harmonic wave train.

There exist important reciprocal relations between moments in time and in frequency. The first is

$$\int \psi^* \psi \, dt = \int \phi^* \phi \, df$$

This is the energy theorem, which we have already met in Eq. II.11. We can write it $M_t^{(o)} = M_f^{(o)} = M_o$; that is, there is only one moment of zero order, and this is the energy of the signal. It is a special case of the two general relations

$$\int \phi^* f^n \phi \, df = \left(\frac{1}{2\pi j}\right)^n \int \psi^* \frac{d^n}{dt^n} \psi \, dt \qquad \text{II.26}$$

and

$$\int \psi^* t^n \psi \, dt = \left(\frac{-1}{2\pi j}\right)^n \int \phi^* \frac{d^n}{df^n} \phi \, df \qquad \text{II.27}$$

This means that if we want to translate a moment or mean value in frequency from "frequency language" into "time language", we must replace ϕ by ψ, and the quantity f by the operator $(1/2\pi j)(d/dt)$. Conversely, if we want to translate a time moment into frequency language we must replace ψ by ϕ and the quantity t by the operator $-(1/2\pi j)(d/df)$.

Those conversant with wave mechanics will recognize the formal identity of these rules with the passage from "coordinate language to momentum language".

II.6. The Uncertainty Relation.

There exists an important relation between the zero-order moment M_o and the two second-order moments in time and in frequency, $M_t^{(2)}$ and $M_f^{(2)}$, in the form of an inequality

$$(4\pi)^2 M_t^{(2)} M_f^{(2)} \geq (M_o)^2 \qquad \text{II.28}$$

This can be derived from Schwarz's inequality (14, 15)

$$4\left(\int F^* x^2 F \, dx\right)\left(\int \frac{dF^*}{dF} \frac{dF}{dx} \, dx\right) \geq \left(\int F^* F \, dx\right)^2 \qquad \text{II.29}$$

by substituting either ϕ or ψ in place of F, f or t in place of x, using Eq. II.26 or II.27 applied to n = 2, and taking into consideration the fact that we admit in our analysis only such functions ϕ and ψ which vanish at infinity. The detailed proof may be left to the reader.

Equation II.28 can also be written in the simpler form

$$(4\pi)^2 \, \overline{f^2} \cdot \overline{t^2} \geq 1 \qquad \text{II.30}$$

Thus the product of the mean-square frequency and of the mean-square epoch of any signal cannot be smaller than a certain minimum. This remains true wherever we put the zero point of time or of frequency.

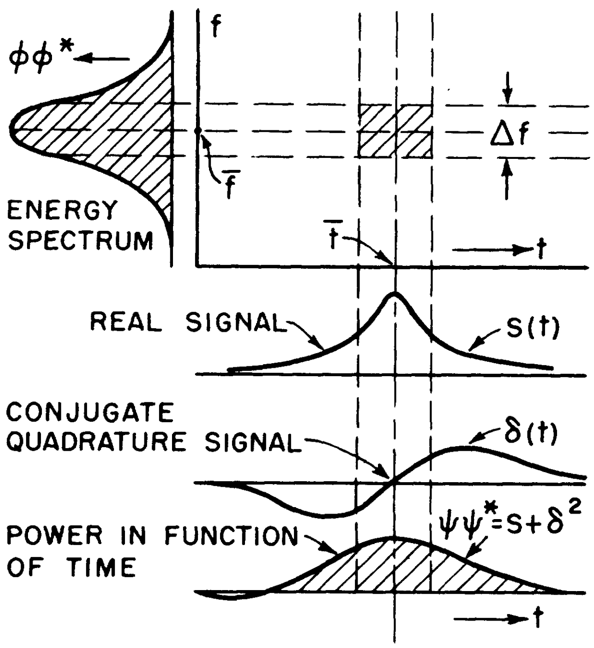

Fig. 2.

Defining the effective duration and effective bandwidth of a signal.

Now $f_2 = (\overline{f^2})^{1/2}$ and $t_2 = (\overline{t^2})^{1/2}$ can be interpreted as the inertial radii of the two portions shown shaded in Fig. 2, relative to arbitrary axes. But it is well known that these become smallest when taken relative to axes drawn through the center of gravity, i.e. through $\bar{f}$ and $\bar{t}$, respectively. Hence the inequality II.30 appears in its sharpest form if we replace

$$\overline{f^2} \text{ by } \overline{(f-\bar{f})^2} \text{ and } \overline{t^2} \text{ by } \overline{(t-\bar{t})^2}$$

Thus the rectangle, shown shaded, which is limited by the inertial radii at both sides of $\bar{f}$ and $\bar{t}$ has a minimum area of $1/\pi$. (This "area" is, of course, a pure number.) In order to avoid the factor π, it is convenient to define the "effective duration" Δt of a signal, and its "effective spectral width" Δf by

$$\Delta t = 2\pi^{1/2}\,\overline{(t-\bar{t})^2}^{\,1/2}$$
$$\Delta f = 2\pi^{1/2}\,\overline{(f-\bar{f})^2}^{\,1/2} \qquad \text{II.31}$$

We can now write the relation II.30 in its sharpest form as

$$\Delta t \cdot \Delta f \geq 1 \qquad \text{II.32}$$

This is the quantitative formulation of the uncertainty relation between time and frequency, which expresses the intuitively obvious fact that the concepts of "epoch" and of "frequency" of a signal cannot both be defined simultaneously beyond a certain limit. If the frequency is sharply defined, the signal must be an infinite harmonic wave with entirely undefined "epoch". If the time is sharply defined, the signal must be a δ- function whose frequency is entirely undefined.

II.7. Representation of Signals by Gaussian Elementary Signals.

Every physical signal is limited in time and in frequency (finite observation time, finite bandwidth). This obvious limitation does not fit in with either of the two descriptions which we have previously discussed. In the "time description" as s (t) we have really analyzed the signal into δ-functions, one infinitely sharp peak to every instant. In the Fourier description the parameter was the frequency, and a sharp frequency means infinite duration. What we require are intermediate descriptions, linear expansions in terms of elementary signals which are finite both in frequency and in duration.

If we have an observation time T and a bandwidth F, how many elementary signals do we need for a full description? An approximate answer can be given on the basis of the Fourier series description in Eq. II.9. There we had a section of the signal, of length T, and we expanded it in terms of functions $\cos \pi(nt/T)$, i.e. with frequencies n/2T, spaced by 1/2T. Thus the number of coefficients in a frequency range F is

$$2FT \qquad \text{II.33}$$

and this is what we can call, according to our previous definition, the "degree of freedom" of the signal in the interval T, F. (If instead of the cosine series we had used the sine-cosine series, the frequencies would have been spaced by 1/T, but now there are two coefficients to every frequency, and the result is again 2FT.) This answer was given independently by Nyquist, and by Küpfmüller, in 1924.

The uncertainty relation now allows us to give this statement a more precise meaning. It is evidently not quite admissible to say that because we have Fourier-analyzed the signal in the time interval T into a series, the frequencies in it are n/2T, up to F. There is a discontinuity on passing from one interval into the next; the change in the Fourier series coefficients produces extra frequencies, a widening of the lines into bands. By Eq. II.32 we can confine the effectively covered area of an elementary signal at best to unity, and the discontinuously changing truncated sine waves are very far from the optimum. (They cover, in fact, an infinite effective area.) Therefore, let us do the analysis rather in terms of elementary signals which come nearer to the optimum or which, if possible, represent the optimum itself.

Such signals exist, and they have a particularly simple structure. They are the gaussian elementary signals (Fig. 3), whose time description is

$$\psi(t) = e^{-a^2(t-t_o)^2} \cdot e^{2\pi j f_o t} \qquad \text{II.34}$$

and whose spectrum is

$$\phi(f) = e^{-\left(\frac{\pi}{a}\right)^2 (f-f_o)^2} \cdot e^{-2\pi j t_o (f-f_o)} \qquad \text{II.35}$$

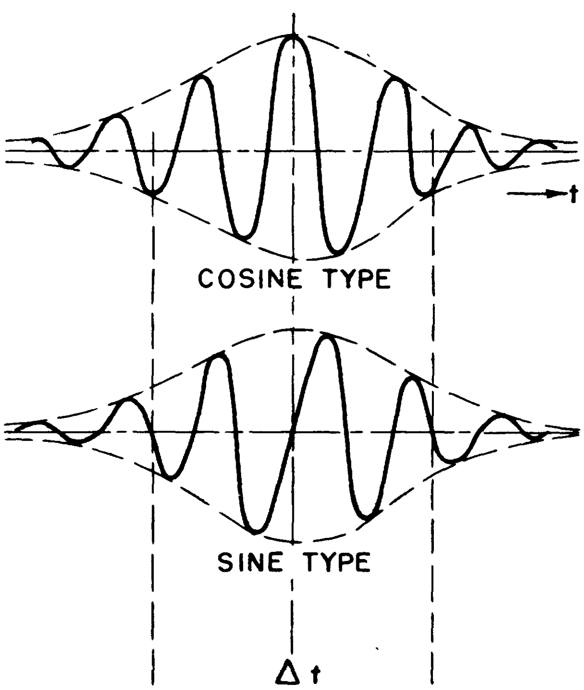

Fig. 3.
Gaussian elementary signals.

Thus both descriptions have the same mathematical form. (It may be left to the reader to prove directly that these functions satisfy the relation $\Delta t \cdot \Delta f \approx 1$ with $\Delta t = \pi^{1/2}/a$, $\Delta f = a/\pi^{1/2}$.) They are harmonic waves with a gaussian probability envelope. They center on the point t_o, f_o in the time-frequency plane, and extend effectively over a rectangle of area unity. The shape (aspect ratio, $\Delta t/\Delta f$) depends on the parameter a, which can be given any value.

The procedure of expanding an arbitrary signal in terms of these elementary signals is this: We divide the time-frequency plane into rectangles or "cells" of unit area, and we associate with every one a gaussian signal ψ_{ik} with a complex coefficient c_{ik}, as shown in Fig. 4. These coefficients are so determined that if we draw a vertical line at any instant, that part of the signal which is inside the frequency band F is given by the expansion

$$\psi_F(t) = \Sigma\Sigma\, c_{ik}\, \psi_{ik}(t) \qquad \text{II.36}$$

Formally the summation has to be carried out over all cells, but it can be shown that only the next columns at either side have an appreciable influence on the sum (16). Similarly, we can draw any horizontal line and obtain the expansion

$$\phi_T(f) = \Sigma\Sigma\, c_{ik}\, \phi_{ik}(f) \qquad \text{II.37}$$

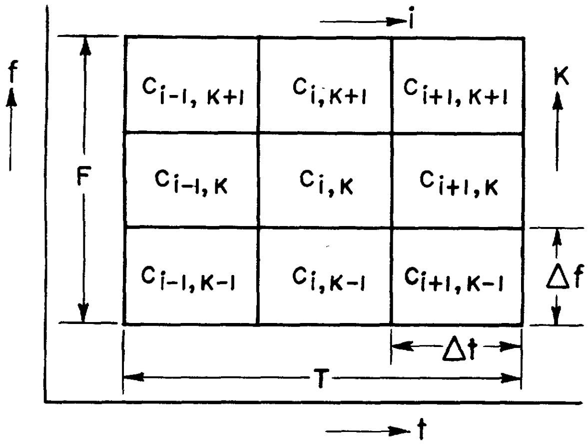

Fig. 4.

Matrix representation of a signal. A complex amplitude c_{ik} is associated with each cell of unit area.

where the ϕ_{ik} are the gaussian elementary signals in the frequency language. The coefficients c_{ik} are the same in both descriptions.

The gaussian signals are the most "compact" of all possible signals, their "spill-over" is a minimum, and the uncertainty of description is just unity, for any area FT. They also have the advantage of complete symmetry in time and frequency, which they share with the Hermitian functions, of which we shall talk later. Another advantage is that the familiar concepts of "amplitude" and "phase" can be taken over from infinite wave trains and can acquire a meaning which is not only more general, but also more precise than in the Fourier description. We will come back to this point when we discuss Ohm's "Law of Acoustics". In fact, the description by gaussian signals is intermediate between the time-description and the Fourier-description. It goes over into the first for $a = \infty$, and into the second for $a = 0$.

The gaussian signals have, on the other hand, the disadvantage that the analysis cannot be carried out easily by electrical means, since the signals are difficult to produce by electronic circuits; nor can they be carried out mathematically, since they are not orthogonal. For this reason, it is worth while to consider other sets of elementary signals, which are somewhat off the optimum, but have practical advantages.

II.8. Representation by "Signals with Limited Spectrum".

An elementary signal, convenient for many purposes, is given by

$$u_{t_o}(t) = (2F)^{1/2} \frac{\sin 2\pi F\,(t-t_o)}{2\pi F\,(t-t_o)} \qquad \text{II.38}$$

Its Fourier transform is

$$U_{t_o}(f) = (2F)^{-1/2}\,[H(f+F) - H(f-F)]\,e^{-2\pi j f t_o} \qquad \text{II.39}$$

H is Heaviside's unit step-function, which is $-\frac{1}{2}$ for negative arguments and $+\frac{1}{2}$ for positive. (An arbitrary constant can be added.) Hence Eq. II.38 represents a function whose spectrum has the constant amplitude $(2F)^{-1/2}$ inside the frequency range $-F < f < F$, and is zero outside it. It may be left to the reader, as a useful exercise, to find the signal with limited frequency whose nonzero band does not center on zero.

From the point of view of "compactness", these signals are not very suitable because their effective duration Δt is infinite. On the other hand, they possess the advantage of orthogonality. The "inner product" of two u functions, whose mean epochs are t_1 and t_2 is

$$\int_{-\infty}^{\infty} u_{t_1} \cdot u_{t_2}\, dt = 2F \int_{-\infty}^{\infty} \frac{\sin 2\pi F\,(t-t_1) \cdot \sin 2\pi F\,(t-t_2)}{(2\pi F)^2\,(t-t_1)\,(t-t_2)}\, dt = \frac{\sin 2\pi F\,(t_2-t_1)}{2\pi F\,(t_2-t_1)} \qquad \text{II.40}$$

This vanishes if the epochs t_1, t_2 are spaced by multiples of $1/2F$, except if $t_1 = t_2$, in which case it is unity. Hence a set of u functions whose epochs are spaced by 1/2F is an orthonormal set. It is also a complete set because it takes just 2F data per unit time to describe any signal inside the bandwidth F.

By using the property of orthogonality, it can be shown immediately that the expansion of an arbitrary real signal s(t) in terms of this set is

$$s_F(t) = \sum_{-\infty}^{\infty} s_n \frac{\sin(2\pi Ft - n\pi)}{2\pi Ft - n\pi} \qquad \text{II.41}$$

with

$$s_n = s(n/2F)$$

That is to say the expansion coefficients are simply the values assumed by the signal at the equidistant points $t = n/2F$. This is therefore an interpolation formula, from E. T. Whittaker (17), 1915, who called s_F the "cardinal function". Whittaker was also the first to prove that the cardinal function contains no Fourier components outside $\pm$ F. If F is increased beyond any limit, the u functions approach delta functions, and the cardinal function becomes identical with the original function.

Shannon (18) and particularly Oswald (19) have made many interesting applications of these "signals with limited spectrum".

II.9. Other Elementary Signals.

In 1938, Wheeler and Loughren (20), in the course of a historically important and still valuable investigation of television problems, discussed signals of the form of a "cosine half-wave", or alternatively, signals whose spectrum has this shape. The author showed later (12) that if one prescribes that the spectrum must be limited to a certain bandwidth, signals with this spectrum will have the smallest effective duration, that is, these are the signals which one can drive in minimum effective time through a certain waveband. The uncertainty product $\Delta t\ \Delta f$ is 1.14, only 14 percent larger than the absolute minimum.

If one wants to record only the absolute squares of the amplitudes, as in a "sound spectrogram" or "sonogram" like that developed by Potter and his collaborators in the Bell Laboratories, one is led to the same type of filter. This has the additional advantage that an even field will appear even, if the cosine half-wave filters are staggered by half their width. In analyzing a sonogram or playing it back, for example, through photocells, one should for the same reason use a squared cosine half-wave filter. In addition to evenness, this will give the best resolution in the direction of the scan.

Wheeler and Loughren have also discussed signals with "parabolic spectrum". These have a slightly larger uncertainty product, 1.194.

Further elementary signals have been recently investigated by van der Pol (21), who has produced a set of functions intermediate between signals of limited spectrum and gaussian signals. Hermitian signals will be discussed in a later lecture.

II.10. Generalized Linear Signal Analysis.

All these representations were based on the two ordering parameters "time" and "frequency", which have become suffixes capable of assuming integer values only, instead of being continuous parameters. One can also say that the analysis was based in every case, whatever choice of elementary functions we have made, on the pair of operators t and $(1/2\pi j)\ \partial/\partial t$. (A quantity like t can also be considered as an operator; it means that the operand at its right must be multiplied by t.)

The question now arises whether we could not have associated, instead of the frequency, another operator with t; or whether we could not have replaced both of them by some other operators.

This is a question of more than mathematical interest, because it throws some light on all multiplex systems of communication. Dividing a channel into frequency bands is one way of separating channels; time-division multiplex is another; but there must exist an infinity of other methods.

The answer can be given in a general way. Representation by two quantities A and B is possible if these correspond to noncommuting operators A and B. This means that if they applied to a symbol at their right in the order A,B, the result must be different from that obtained by interchanging A and B.

In symbols $$[A,B] \equiv A\cdot B - B\cdot A \neq 0 \qquad \text{II.42}$$

If this condition is satisfied, there will exist an uncertainty relation between the quantities A and B, and a representation similar to our joint time-frequency representation will be possible. It will also be possible to separate transmission channels, by allotting to them different domains of A and B.

Space forbids our going into more detail concerning these interesting problems. For those who want to study them, the following* are references to works on Quantum Mechanics, where the general mathematical part of the problem has already been explored to a considerable extent.

III. PHYSICAL SIGNAL ANALSYIS.

III.1. Introduction.

All methods discussed in the previous section lead to a system of "free coordinates" in the form of coefficients c_{ik}. We must now go on with the process of breaking down a signal into a selection from a finite set of possibilities. This means that we can no longer consider the c_{ik} as exact quantities in a mathematical sense.

While the previous analysis was based on mathematical identities, on purely logical processes, further progress can be made only by taking at least a few basic experiences from physics. This is evident, as the c_{ik} now become physical quantities. Before we can reduce them to numbers and start a mathematical discussion, we must have some sort of unit, and this can come from physics only.

The physical situation is that if we measure a quantity c_{ik}, this will be different by some uncertain amount from the value c_{ik} which represents one datum of the message. The situation is particularly simple in the case of communication between human minds, or between machines of the digital type if c_{ik} is known to be quantized. In this case the problem is of the type: "A value c has been measured. What are the probabilities $p_i\,(c_i\,|\,c)$ that c_i has been sent?" Thus we have the choice of a discrete set of values; only their probabilities are continuous.

In practice, we can always reduce the problem to this simple case, because the sender will not attempt to transmit data so finely graded that there is no reasonable chance of getting his original meaning. Moreover, we will arrive at a very similar result even in the case in which the sender is not a human being or a digital machine, but some natural phenomenon, because it is meaningless to talk of quantities which exist if they are too fine to be measured, and because there are physical limits to the precision of measurement.

Uncertain quantities, which cannot be calculated from the data available, are termed "noise". The term also has another meaning; noise is what we do not want in a communication channel. This second definition is more or less the accepted basis in Wiener's filtering theory, which starts

* For the association of operators with physical quantities, see any textbook on Wave Mechanics. Max Born in his "Atomic Physics" gives a particularly simple and elegant proof of the theorem that there exists an uncertainty relation between quantities whose operators are noncommuting. Further investigations on the general uncertainty relations can be found in reference 15, p. 235. The general theory of representation is known in Wave Mechanics as the Dirac-Jordan transformation theory, which is also very clearly explained in Tolman's book. Compare also reference 12, Appendix. I.

from some statistical regularities in the signal source, and in the noise source, which are assumed as known. There need not be any physical difference between the two; the crosstalk may be just as intelligent talk as the message, only we do not want it. In the physical theory, on the other hand, we must disregard this type of noise, since we cannot define "what is wanted" in physical terms.

If crosstalk and the like are eliminated, there remain three basic physical causes of uncertainty:

1. Temperature. We do not know the state of our apparatus (transmitter, channel, receiver) exactly. We know only its macroscopic parameters, such as the mean energy contained in a thermoelement, or the mean voltage drop across a resistor. In order to have a basis for the calculation, we assume that at least each part with which we are dealing separately is in thermal equilibrium, so that we can assign to it a temperature T. We can call this cause of uncertainty the "classical ignorance of detail".

2. The Finite Size of the Quantum of Action. Even if we eliminated the classical uncertainty by cooling all our instruments to the absolute zero, there would remain an uncertainty due to the basic physical fact that the exact measurement of energy requires infinite time. If there is a time interval Δt at our disposal, the error ΔE made in the determination of an energy will be of the order as given by the relation

$$\Delta E \cdot \Delta t \gtrsim h \qquad \text{III.1}$$

where h is Planck's constant, 6.54×10^{-27} erg sec.

It is convenient for our purposes to consider this uncertainty relation as the basic formulation of quantum theory. It can be interpreted by Planck's relation

$$E = hf \qquad \text{III.2}$$

where f is a quantity of the dimension of a frequency. With this substitution III.1 goes over into the mathematical identity II.32. But it must be made clear that by making the substitution III.2 we have by no means explained or derived a fundamental fact of experience. In most cases the frequency f associated with an energy E has no verifiable physical meaning. Even in the case of light we cannot measure the frequency directly; we only infer it from measurements of the wavelength and of the velocity of light. Hence Eq. III.1 is the physical statement, not Eq. III.2. It is important to realize from the start that the direct physical manifestations of the quantum are fluctuations and other uncertainties.

3. The Finite Size of the Elementary Charge. It may be remembered that electromagnetic signals become observable and measurable only through the forces which they exert on charged bodies. Moreover, weak electric signals are always amplified by electron streams, either in vacuo or in solids. The atomic nature of electricity manifests itself as shot noise, which, as will be shown, is one of the factors determining the ultimate limits of signal detection.

III.2. The Information Cell as a Planck Oscillator.

All electromagnetic signals are conveyed by radiation. Even if conductors are used in the transmission, by the Maxwell-Poynting theory the energy can be located in empty space.

The simplest model which we can use as a physical illustration of one-dimensional transmission is a cable, or a line, in which all signals are propagated with the velocity of light c. A signal of duration Δt extends over a length $c\Delta t$ of the cable and travels with speed c towards the receiver. In addition, there may be another wave travelling in the opposite direction; this does not concern us.

We can now apply to the forward wave in the cable exactly the same mathematical considerations as those used in the last section, with the only difference that the time interval Δt is now replaced by the length $c\Delta t$. We thus find that this physical system, the cable, has $2FL/c$ degrees of freedom for the forward wave on a length L. (L/c is the time interval T in which the wave of length L passes a fixed point, say the terminals of the receiver.) We can now apply to this physical system the general results of the Statistical Theory of Radiation. These can be expressed in the form:

(a) Two joint degrees of freedom (one amplitude with its phase, or simply one complex degree of freedom) can be considered as a "Planck oscillator".

(b) At a temperature T the mean energy of each Planck oscillator is

$$\bar{\mathcal{E}}_T = \frac{hf}{\exp(hf/kT) - 1} \qquad \text{III.3}$$

This is Planck's law. Thus we can say that at thermal equilibrium each information cell, $\Delta t\ \Delta f = 1$ will contain this energy.*

We can generalize this somewhat to the case when there is no thermal equilibrium. By the general law that "bodies radiate independently of one another, as if they radiated into a surrounding of zero absolute," it is evident that if there is no thermal equilibrium, T means the temperature of the transmitter, not of the receiver. If the cable has a third temperature, the situation is more complicated, but the solution can be inferred from our later results.

At large values of the parameter kT/hf, Planck's law (Eq. III.3) goes over into Boltzmann's law

$$\bar{\mathcal{E}}_T = kT \qquad \text{III.4}$$

The range of validity of this "classical" approximation can best be seen by expressing the parameter in terms of wavelength $\lambda = c/f$ instead of in terms of frequency

$$\frac{kT}{hf} = \frac{kT}{hc}\lambda = 0.7\lambda T \qquad \text{III.5}$$

where λ has to be substituted in cm, T in degrees absolute. When this number is large we are in the classical region. As an example for $\lambda = 1$ cm and $T = 300°K$, it is 210, still a large number. The change-over into the "quantum region" occurs rather abruptly when the parameter becomes of order unity.

III.3. Thermal Fluctuations in the Information Cell.

The expression III.3 represents the thermal energy in the cell, but it does not in itself mean a "noise". If it were constant, we could simply subtract it from the received signal. It becomes a disturbing factor only by its fluctuations. (This fact is usually forgotten in technical texts on noise, but without consequence, since it so happens that in the classical region the rms fluctuation of the noise power is equal to the noise power itself.)

Gibbs and Einstein have shown independently that the mean-square fluctuation of energy in any system at thermal equilibrium, whose specific heat is $d\bar{\mathcal{E}}/dT$, is given by the formula

$$\overline{\delta\mathcal{E}_T^2} \equiv \overline{(\mathcal{E}_T - \bar{\mathcal{E}}_T)^2} = kT^2 \frac{d\bar{\mathcal{E}}}{dT} \qquad \text{III.4a}$$

*The physical laws, which we have adopted here without proof or explanation, are from Planck and from Jeans. It was Jeans who first counted the degrees of freedom of a cavity in terms of standing waves or "modes" and interpreted them as "Planck oscillators". Von Laue, in 1914, carried out the analysis in terms of travelling waves, in a way similar to the one used here. In 1928 Nyquist used the cable model to derive his well-known law of thermal noise.

(The bar notation of mean values requires careful writing. $\overline{x^2}$ means the mean square of x; $\overline{x}^2$, the square of the mean.) Applying this general equation to Planck's law, one finds

$$\overline{\delta \mathcal{E}_T^2} = hf\,\overline{\mathcal{E}}_T + \overline{\mathcal{E}}_T^2 \qquad \text{III.5a}$$

To this equation Einstein, in 1905, gave an extremely illuminating interpretation. It is seen that the mean-square fluctuation of energy is the sum of two terms, as would be the case if the fluctuation were due to the superposition of two independent causes. The second is the "classical" or "wave term", because, as H. L. Lorentz has shown in detail, it can be ascribed to the superposition of wave components (the Fourier components of the radiation field, classically analyzed) with random phases. But the first term is evidently of quantum origin, since it would vanish if we could neglect h. We see its significance better if we express the energy $\overline{\mathcal{E}}_T$ in terms of units hf, putting

$$\overline{N}_T = \frac{\overline{\mathcal{E}}_T}{hf} = \frac{1}{\exp(hf/kT) - 1} \qquad \text{III.6}$$

We now obtain the fluctuation Eq. III.5a in the form

$$\overline{\delta N_T^2} \equiv \overline{(N - \overline{N}_T)^2} = \overline{N}_T + \overline{N}_T^2 \qquad \text{III.7}$$

If $\overline{N}$ is a large number, the second, classical, term predominates; at small $\overline{N}$ the first predominates, so that in this domain

$$\overline{\delta N_T^2} \approx \overline{N}_T \qquad \text{III.8}$$

But this is Poisson's Law of Rare Events. We should obtain it if we imagined that the energy $\overline{\mathcal{E}}_T$ is present in the form of particles, "photons", each with an energy hf, and that these photons moved about independently of one another, like the particles in a rarified gas.

We see again that it is by the fluctuations that the quantum aspect is introduced into radiation theory. Equation III.5a is also a warning against a too naive interpretation of the photon or particle aspect of radiation. Neither the classical field concept nor the particle concept can explain the phenomena, but only the two together.

Equation III.7 shows very clearly the division between the classical and the quantum domain, which occurs where $\overline{N}_T$ is of the order unity. If $\overline{N}_T$ is large, it can be approximated by

$$\overline{N}_T = kT/hf = 0.7\,\lambda\,T \qquad \text{III.9}$$

an expression which we have already discussed, while for small values

$$\overline{N}_T = \exp(-hf/kT) = \exp(-1/0.7\lambda T) \qquad \text{III.10}$$

is a good approximation. For visible light, $\lambda = 5.10^{-5}$ cm. This gives the extraordinarily small figure of 10^{-39} photons per cell, if T = 300 °K. At room temperature the division between the classical and quantum domain can be put roughly at $\lambda = 3.10^{-3}$ cm, or 30 microns. Wavelengths only about three times as long have already been produced electronically.

It may be noted that in the classical domain the rms power fluctuation is equal to the mean thermal power, but in the quantum domain it is always larger and may become very much larger.

III.4. Fluctuations in the Presence of a Signal.

So far we have considered thermal fluctuations only. It would be wrong to assume that in the presence of a signal we have only to add this to the signal, as if they were independent. We can, however, obtain the resulting noise by the following heuristic consideration (22).

We start from Einstein's result that the classical and the quantum parts of the mean-square energy fluctuation are additive, as if they were independent. Let us first calculate the classical part.

Consider a complex Fourier amplitude E_s of the signal at a frequency ω_1, and the Fourier amplitude E_T of the thermal noise at ω_2. The instantaneous energy density, resulting from the interference, is proportional to

$$E_S E_S^* + E_T E_T^* + \left\{ E_S E_T^* \exp\left[j(\omega_1 - \omega_2)t\right] + E_S^* E_T \exp\left[-j(\omega_1 - \omega_2)t\right] \right\}$$

The first term is the energy of the signal; the second, the noise energy; the rest is the result of interference or beats. Let us write this in the form

$$\mathcal{E} = \mathcal{E}_S + \mathcal{E}_T + \mathcal{E}_{ST} \qquad \text{III.11}$$

The mean value of the interference energy $\mathcal{E}_{ST}$ is nil, but not its mean square, which is

$$\overline{\mathcal{E}_{ST}^2} = \overline{2E_S S_S^+ \cdot E_t E_T^+} = 2\mathcal{E}_S \overline{\mathcal{E}}_T \qquad \text{III.12}$$

Using this, one obtains for the resulting mean-square fluctuation

$$\overline{(\bar{\mathcal{E}} - \bar{\mathcal{E}})^2} = \overline{[(\mathcal{E}_S + \mathcal{E}_T + \mathcal{E}_{ST}) - (\bar{\mathcal{E}}_S + \bar{\mathcal{E}}_T)]^2} = \overline{\mathcal{E}_T^2} - \bar{\mathcal{E}}_T^2 + 2\mathcal{E}_S \bar{\mathcal{E}}_T \qquad \text{III.13}$$

Here we have made use of the relations

$$\overline{\mathcal{E}_S \mathcal{E}_{ST}} = \overline{\mathcal{E}_T \mathcal{E}_{ST}} = 0$$

which are evident, as there is no correlation between the signal and the noise. We know the first two terms in Eq. III.13 from Lorentz's calculation, which gives

$$\overline{\mathcal{E}_T^2} - \bar{\mathcal{E}}_T^2 = \bar{\mathcal{E}}_T^2$$

so that the classical part of the energy fluctuation becomes, expressed in terms of signal energy and noise energy,

$$\overline{(\mathcal{E} - \mathcal{E})^2} = 2\mathcal{E}_S \bar{\mathcal{E}}_T + \bar{\mathcal{E}}_T^2 \qquad \text{III.14}$$

It may also be convenient, for some purposes, to write this in terms of the total mean energy $\bar{\mathcal{E}} = \mathcal{E}_S + \bar{\mathcal{E}}_T$ in the cell. This gives

$$\overline{(\mathcal{E} - \bar{\mathcal{E}})^2} = (2\bar{\mathcal{E}} - \bar{\mathcal{E}}_T)\bar{\mathcal{E}}_T \qquad \text{III.15}$$

To this classical part we must now add the quantum component, which by its interpretation as photon noise can depend on $\bar{\mathcal{E}}$ only, irrespective of how it is divided between thermal and signal energy. This term is therefore taken over without alteration from Eq. III.5, and we obtain the complete formula for the energy fluctuation in the presence of a signal

$$\overline{\delta\mathcal{E}^2} \equiv \overline{(\mathcal{E} - \bar{\mathcal{E}})^2} = hf\,\bar{\mathcal{E}} + 2\,\bar{\mathcal{E}}\,\bar{\mathcal{E}}_T - \bar{\mathcal{E}}_T^2 \qquad \text{III.16}$$

III.5. The Ladder of Distinguishable States of an Information Cell.

If we carry out a measurement of the energy in a cell, how much of the measured energy belongs to the signal? The answer is evidently uncertain, and the rms value of the fluctuation suggests itself naturally as a measure of this uncertainty. This suggestion receives its justification from a theorem of Bienaymé and Tchebycheff: Whatever the law of fluctuation, the probability of an error k times exceeding the rms value of the fluctuation is always smaller than $1/k^2$.

Thus, if we divide the scale of a physical quantity into intervals equal to the rms fluctuation, we can say that, a measurement having given a certain value, there will be about a 50–50 chance that the signal contained in the value differs only by at most $\pm\frac{1}{2}$ division from it. In general such a division will not be even. We can make it uniform by expressing the value directly by the number of divisions or steps which lead to it. Following the useful terminology of D. M. MacKay (23) we call this the proper scale of the physical quantity.

Equation III.16 now enables us to calculate the proper scale of any electromagnetic signal, associated with one information cell. It is advantageous to express this result first in terms of photon numbers N, since in this writing it can immediately be seen when we are in the classical domain ($N \gg 1$), and when in the quantum domain ($N \ll 1$),

$$\overline{\delta N^2} \equiv \overline{(N - \overline{N})^2} = N(1 + 2\overline{N}_T) - \overline{N}_T^2 \qquad \text{III.17}$$

Assuming that we have to deal with a fairly large number of steps S up to a level N, as is always the case, this will be approximately

$$s = \int_{\overline{N}_T}^{N} \frac{dN}{(\delta N^2)^{\frac{1}{2}}} = \int_{\overline{N}_T}^{N} \frac{dN}{[N(1+2N_T)-\overline{N}_T^2]^{\frac{1}{2}}} =$$

$$\frac{2}{1+2\overline{N}_T}\left\{[N(1+2\overline{N}_T)-\overline{N}_T^2]^{\frac{1}{2}} - [\overline{N}_T(1+\overline{N}_T)]^{\frac{1}{2}}\right\} \approx \frac{2N^{\frac{1}{2}}}{(1+2\overline{N}_T)^{\frac{1}{2}}} \qquad \text{III.18}$$

The last expression is valid for large signals $N \gg \overline{N}_T$. In the classical case, in which $N_T \gg 1$, this simplifies further to

$$s \approx (2N/\overline{N}_T)^{\frac{1}{2}} \approx (2\mathcal{E}/\overline{\mathcal{E}}_T)^{\frac{1}{2}} \qquad \text{III.19}$$

This shows that in the classical case the proper scale is the square root of the energy scale, i.e. the amplitude scale. This may be somewhat surprising, since it is usually assumed that as the signal and the noise are uncorrelated, the noise energy has to be added to the signal energy. But though this is true if mean values are concerned, it is not true for the fluctuations, as we have seen. The phenomenon of beats between the signal and the noise has the consequence that the uncertainty increases with the square root of the signal energy, while on the amplitude scale it remains constant. The factor $2^{\frac{1}{2}}$ in Eq. III.19 is a matter of convention; it could have been avoided if we had adopted $2\frac{1}{2}$ times the rms value of the energy fluctuation as the unit step.

We can now complete our program of breaking down signals into discrete elements. We have only to change our previous two-dimensional representation into a three-dimensional one (Fig. 5). On each information cell we erect a ladder of distinguishable steps, which will be evenly spaced if we take the signal on the proper scale. If now we have at our disposal for the signal a time T, a frequency band F, and an energy which corresponds to a maximum number of s_m steps, any signal in these limits corresponds to TF selections, each from s_m possible choices. Thus, by the Hartley-Shannon definition we can at once write down the information capacity

$$S = FT \log s_m \qquad \text{III.20}$$

With our definition of noise as due to unavoidable physical causes, this expression contains the temperature as the only parameter in addition to those exhibited in the equation.

It is also noted that Eq. III.20, of course, contains the assumption that the energy disposable in each cell is prescribed. If the total energy in the signal is given, we should obtain another expression, though of similar build. (Cf. ref. 11, p.1168.) In every case, the problem is reduced to a purely mathematical one: calculating the possible selections of lattice points with the prescribed limitations.

III.6. Classical Analysis and Quantum Analysis.

We have now brought our mathematical scheme into agreement with physics, but it might appear that we have lost half of the data. Originally we started out with two data per cell: an amplitude and a phase. Now we have only one, which can be expressed in terms of a real positive number, the number N of quanta in the cell.

To explain this paradox it must first be understood that the formulation, "we have lost half the data", is wrong. In fact, we have lost all degrees of freedom; instead of a 2FT-fold infinity we now have only a finite set of possible choices. But there remains a physical question.

We know that it is possible to analyze an electromagnetic wave in terms of amplitude and phase. We can do this if we have first recorded it, for example, by means of a cathode-ray oscillograph. We also know that we can count photons by means of scintillation counters, Geiger-counters, and the like. Ordinarily we never count the quanta in electromagnetic radiation, though this is possible to some extent, at least for microwaves, by the molecular beam methods developed by I. I. Rabi and his collaborators. Nobody has yet succeeded in taking an oscillogram of visible light. Yet the wave aspect and the photon aspect always co-exist, and they may even become simultaneously important in the extreme microwave region. Thus the problem is: If we analyze the same electromagnetic signal alternatively by "counter" methods and by "oscillograph" methods, do we get the same number of discrete data, or is one method of analysis inherently superior to the other?

In previous sections we have calculated the number of data which we could ascertain with ideal counters, which determine the photon number N accurately. We must now explore an ideal amplitude-phase or wave-analyzing device and compare the results. Only a very abridged summary of this investigation can be given here. (For details refer to Gabor, 1950, 1951.)

If we want to analyze a signal to its smallest details, we must first amplify it as nearly proportionally as possible. Amplification means that the signal is used for modifying a stream of energy. As the signal is in electromagnetic form, the stream must be one of charged particles. For simplicity we assume a stream of electrons *in vacuo,* which crosses the stream of electromagnetic energy. This is a special choice; the electron stream could as well move with the wave. But there is no essential difference once one goes to the limit, and the cross-stream device is simpler to analyze.

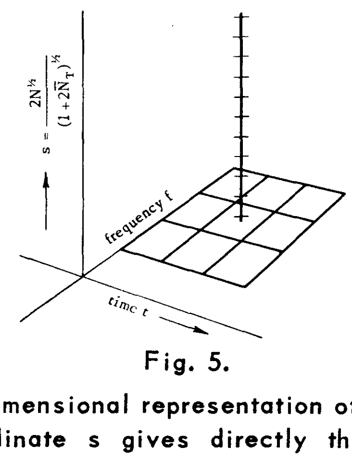

Fig. 5.

Three-dimensional representation of signals. The ordinate s gives directly the number of distinguishable steps.

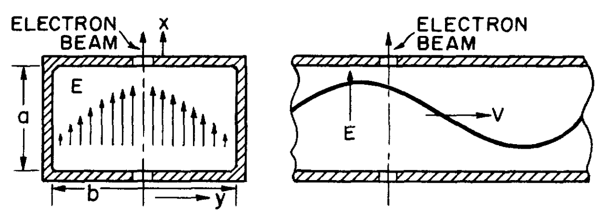

Fig. 6.

Imaginary experiment for the detection of an electromagnetic signal.

If the electromagnetic signal is introduced into a rectangular wave guide in the TE_{10} mode, as shown in Fig. 6, the electron beam passes through two small holes in the direction of the electric field E. One could just as well assume that the beam is running at right angles to field E, so that the beam is deflected instead of velocity-modulated. It is known how the velocity modulation can ultimately be translated into current modulation, by bunching, or by submitting the beam to a constant deflection field, so that it is made to play between two collecting electrodes, which are close together. In either case, the result is an alternating current which is proportional to the electron current and to the accelerations and retardations suffered by the electrons in the waveguide.

The first problem is to calculate the energy exchange between the wave and the beam. This is best done in units hf, where f is the mean frequency in the waveband admitted. Let $\bar{N}$ be the mean number of photons in an information cell $\Delta f \, \Delta t = 1$, and $\bar{n}$ the mean number of photons which a beam electron picks up or loses in passing through the guide. It may be emphasized that this is not to be understood in the naive sense that the interchange goes in steps hf. $\bar{n}$ connotes just the mean interchange energy, always a positive number, divided by hf.

The calculation has to be carried out differently in the classical region, where $\bar{N}$ and $\bar{n}$ are large, and in the quantum region. In the classical region one obtains

$$\bar{n}^2 = \frac{32}{\pi^3}\left(\frac{2\pi e^2}{hc}\right)\frac{\sin^2\theta}{\theta}\,\frac{v}{bf}\left[1-(c/2bf)^2\right]^{-\frac{1}{2}}\frac{\Delta f}{f}\,\bar{N} \qquad \text{III.21}$$

This means that the energy exchange is proportional to $\bar{N}^{\frac{1}{2}}$, i.e. to the field amplitude. In the proportionality factor, e is the electron charge, v the electron velocity, $\theta = \pi af/v$ is one-half of the "transit angle" of the electron in the guide, the guide dimensions a and b are as shown in the sketch. The numerical factor, shown separately in brackets, is of interest; its reciprocal

$$hc/2\pi e^2 = 137$$

is the fundamental number which connects photons with elementary charges.

In order to make the analysis efficient, the numerical factor of $\bar{N}$ in Eq. III.21 must be made as large as possible. The factor $\text{sine}^2\ \theta/\theta$ has a maximum of 0.723 at $\theta = 67°$, i.e. at a transit angle of 134°. This disposes of the depth a of the guide. The width b is determined by the consideration that the quantity

$$cb\,[1-(c/2bf)^2]^{\frac{1}{2}} = bU$$

must be as small as possible. This is b times the group velocity U in the guide. The smallest possible value of this is achieved if the cutoff is at the low-frequency edge, $f - \frac{1}{2}\Delta f$ of the waveband Δf. Substituting the corresponding value of b into Eq. III.21, one now obtains for optimum conditions the energy transfer

$$\bar{n}^2 = \frac{1.5}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta f}{f}\right)^{\frac{1}{2}}\bar{N} \qquad \text{III.22}$$

This is an interesting formula; it shows that the exchange in a single passage cannot be pushed beyond a certain figure because the velocity v of the beam electrons is always smaller than c. One can obtain better exchange by repeated passages. If the waveband Δf is sufficiently narrow, one can make the electron pass the guide P times, but only so long as P is not larger than $f/\Delta f$, else the passages will fall out of phase with the wave, and there will be loss instead of further gain. But if $\Delta f/f$ is a sufficiently small number, the number of passages can be increased until one reaches

$$\bar{n}^2 = \bar{N} \qquad \text{III.23}$$

which, as will be shown later, is the optimum adjustment.

In the quantum domain, that is, for small $\bar{n}$ and $\bar{N}$ the result is different. One finds for the optimum, instead of Eq. III.22,

$$\bar{n} = \frac{1.0}{137}\left(\frac{v}{c}\right)\left(\frac{\Delta f}{f}\right)^{1/2}\bar{N} \qquad \text{III.24}$$

This is quite similar to the classical equation, but with the fundamental difference that the energy exchange $\bar{n}$ is now proportional to $\bar{N}$ instead of to its square root. Thus in the quantum region it is as if we had single collisions of electrons with photons. We could try to interpret the classical Eq. III.22 similarly as the result of many collisions, each of which can have a positive or a negative result. But this does not take into account the fact that the positive and negative exchanges are not randomly distributed but ordered in time, i.e. it would not explain the formation of a field. This is a reminder that one cannot understand radiation on the basis of the naive particle picture; the photon picture and the field picture must be kept in mind simultaneously and not used as alternatives.

In the quantum domain, analysis into amplitude and phase is meaningless, but in the classical domain, to which Eq. III.22 refers, we can now ask the question: How many steps in the photon number N can be distinguished if we apply the output of the modulated beam, for example, to a cathode-ray oscillograph, which traces a curve of N or of $N^{1/2}$, and permits separation of amplitude and phase?

The problem is again one of calculating the rms fluctuation of the quantity which will be recorded as the "amplitude of the field", and to count the number of steps up to a certain level. Again only the bare outlines of the calculation can be given.

We want to make the rms error as small as possible, and to do this we have two variables at our disposal. One is the coefficient of $\bar{N}$ (multiplied by the number of passages P). Let us write this as

$$\bar{n}^2 = K\bar{N} \qquad \text{III.25}$$

and call K the exchange parameter. The other variable still at our disposal is the electron current density. Let us express this by the number M of electrons which pass through the guide while it is traversed by one information cell of duration $\Delta t = 1/\Delta f$. Thus we are still free to chose K and M.

There must be an optimum both for K and for M, for these reasons. A too weak interchange (small K, small M) would leave the cell unexplored. A too strong interchange, on the other hand, would spoil the object of observation. Though in the mean, electrons are as often accelerated as retarded, so that no energy is taken out of the signal, the electron numbers which go through in positive and negative phases will necessarily fluctuate, because of the finite size of the elementary charge (shot effect). Thus a strong interchange can, for example, extract energy which is not necessarily given back in the next half-cycle, and a spurious signal has been produced. Thermal noise we will disregard; it is not essential in this connection.

Taking these effects into consideration, and assuming normal shot effect,

$$\overline{\delta M^2} = \bar{M} \qquad \text{III.26}$$

i.e. without a smoothing factor, one obtains two simple rules for the optimum. One is K = 1, i.e. $\bar{n}^2 = N$. This determines the apparatus, apart from the electron current. The current in turn is given by

$$\bar{n}.\bar{M} = 2\bar{N} \qquad \text{III.27}$$

or, combining the two conditions,

$$\bar{M} = 2\bar{N}^{1/2} \qquad \text{III.28}$$

Comparing this with Eq. III.18, in which we put $\bar{N}_T = 0$, we see that this can also be expressed by the simple rule: Take one electron per distinguishable step.

If we give K and M their optimum values, we find that the relative error in the modulated electron current which measures the field amplitude is much larger than the relative error in the photon scale. The number of distinguishable steps on the amplitude scale is only the square root of the steps in the scale which we could have obtained by an ideal counter, which measures N only, but disregards the scale altogether. The reason is that now we have a second variable: the phase, which we can also measure, with the same accuracy. By combining the two independent scales, "the amplitude ladder" and the "phase ladder," we obtain $N^{1/4} \times N^{1/4} = N^{1/2}$, the same as before. Thus, ideally, the classical analysis gives us in the limit the same information as the photon analysis. All we have done by applying classical analysis was to order the information points in a two-dimensional lattice instead of in a row.

We can now answer the puzzling question regarding the phase of a ray of visible light. How is it that it has never been measured? The answer is that it can never be measured until we have light sources with very much higher spectral densities than at present. The relative error in the determination of the phase is the inverse fourth power of the photon number N per cell. Thus we require 10^8 quanta to determine phase to 1 percent. But sunlight at its visible maximum has only 0.03 photons per cell, and even the most intense line-spectrum sources have hardly more than 100 times as much. On the other hand, there is, in principle at least, no difficulty in determining the phase of the radiation emitted by a 10-cm, 1-Mw pulsed magnetron, because this sends 10^{24} quanta into one cell.

It can be seen, however, even from this last example, that with the progress towards shorter and shorter waves, and considering the weakness of the signals which are available in the reception, the limit will soon be reached when the present methods of dealing with signals as waves will no longer pay, and instruments of the "counter" type will have practical advantage.

III.7. The Limits of the Shannon-Tuller Law.

C.E. Shannon (24) and W.G. Tuller (25) have given a formula for the transmission capacity of a communication channel in the presence of white noise, basing their arguments on classical physics. We will now discuss the manner in which the validity of this law may be affected by the fundamental phenomena studied in the last section.

In Shannon's exact formulation, a channel of bandwidth F, in time T, in the presence of white noise of mean power $\overline{\mathcal{E}}_T$, has a capacity of

$$C = FT \log_2 (\overline{\mathcal{E}} / \overline{\mathcal{E}}_T) = [FT \log_2 (1 + S/N)] \text{ bits} \qquad \text{III.29}$$

assuming that the signal has a Gaussian amplitude distribution, with a mean power which is $\overline{\mathcal{E}}_S = \overline{\mathcal{E}} - \overline{\mathcal{E}}_T$, with our symbols. In the second expression for C we have used the more usual symbols S for the mean signal power, N for the mean noise power. By Shannon's theorem this is the upper limit for the amount of information which can be transmitted in these circumstances, and this upper limit can be approached with an error arbitrarily small, if an ideal code is used and the transmission time is very long.

In Eq. III.19 we have found that the number of distinguishable steps up to an energy level $\mathcal{E}$ in a cell is $(2\mathcal{E}/\overline{\mathcal{E}}_T)^{1/2}$ in the classical case. This allows us to give a simple interpretation of Eq. III.29, which is, of course, no full substitute for Shannon's more exact theory. Disregarding the factor $2^{1/2}$ and similar factors which arise if we substitute a Gaussian distribution instead of an upper limit for the amplitude, we see that the number of distinguishable amplitude configurations is the square root of the expression under the logarithm sign in Eq. III.29. This gives a factor of one half, if we take the power before the logarithm. The other half is accounted for by the equal number of independent phase configurations.

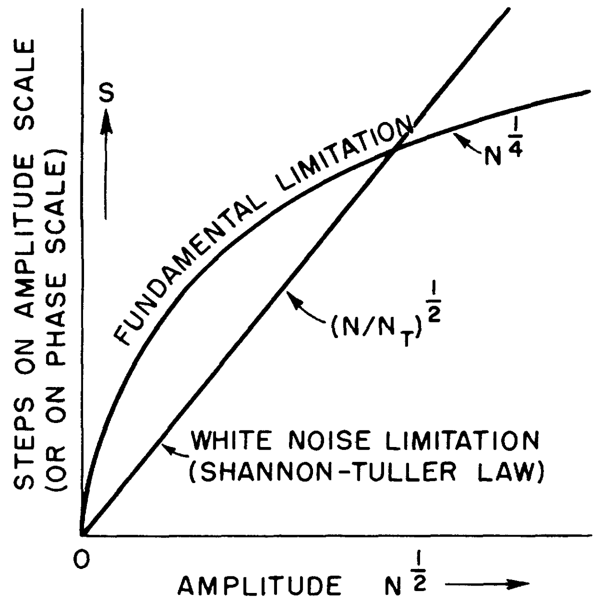

Fig. 7.
The limits of the Shannon-Tuller law for the capacity of a noisy channel.

We have seen, however, in the last section, that ultimately there is just as much information in the energy steps alone as in the amplitude-phase analysis. It must be emphasized that this is true only if we go to the limit, at which we can disregard the thermal noise (or white noise of any origin). The calculations in the last section were based on the assumption $\bar{n}_T = 0$. In radiofrequency communication the white noise is usually so strong that the fundamental limitation will make itself felt only at very high signal intensities. This is illustrated in Fig. 7. So long as white noise is dominant, the number of distinguishable steps is approximately $(N/\bar{N})^{1/2}$. This is represented by a straight line. The Shannon-Tuller Law holds until this line intersects the parabola representing the fundamental limitation, whose equation is approximately $s = N^{1/4}$. Hence the Shannon-Tuller Law is valid up to about

$$N/\bar{N}_T \lesssim N^{1/2} \qquad \text{III.30}$$

that is, up to the point when the signal-to-noise power ratio reaches the square root of the photon number. For example, at 60-db signal-to-noise the Shannon-Tuller Law holds so long as the signal itself corresponds to more than 10^{12} photons per cell.

Actually this breakdown would be well within the range of observation in the microwave region, were it not for the fact that the white noise is usually considerably in excess of the thermal noise, calculated by Eq. III.9 for room temperature. For instance for $\lambda = 10$cm the number of photons at $T = 300°$ is only 2100 per cell. Thus we could realize a transmission with 60-db signal-to-noise with about 4.10^{12} photons per cell, where the fundamental limitation is just about reached. The power P in a waveband Δf, at a mean frequency f is connected with the photon number N by the relation

$$N = P/hf\,\Delta f = 1.5\times10^{34}\,P/f\,\Delta f \qquad \text{(P is expressed in watts)}$$

At $f = 3.10^9$ cycles, corresponding to $\lambda = 10$ cm, and a bandwidth Δf of 1 Mc, this shows that the noise power in this band is only about 6.10^{-13} watt. Thus one could test the Shannon-Tuller Law with a power upwards of approximately 1 $\mu\omega$, if the noise could be really kept so low. Yet it can be seen that the modified law might soon come into the range of an experimental test. This modification was pointed out to the author privately by Léon Brillouin.

IV. Information Theory Applied to Light.

IV.1. What is Information Theory?

As we stated in the Introduction, information theory is a fairly recent attempt to introduce the point of view of communication theory into scientific method. Before we start enumerating these guiding principles, we must start with a disclaimer. It is not claimed that all or even any of them are novel. On the contrary, all of them have been known for a more or less long time, especially in physics, and it is through modern physics that they have found their way into communication theory. Yet it is possible, and even likely, that after a certain process of systematization and logical clarification, these principles will do good service as guides in scientific research.

The first is that everything we can learn by experiments can be expressed in finite form, by a finite set of data. Continua exist only as mental constructions, and are used to simplify the analysis, but they can have no physically ascertainable meaning.

Second, in analyzing any experiment, we must first lay down a system of "free coordinates", each corresponding to a quantity which is *a priori* independent of the others, which allow a complete and nonredundant description of anything observable under the given conditions. In other words, the free coordinates must take care of all of the limitations which are imposed on the phenomena by the observation system: the *a priori* limitations. Forgetting these might lead to errors like those of Eddington's ichtyologist, mentioned in the first lecture.

It must not be thought, however, that the *a priori* information about our experimental setup can be obtained by pure thought, as the words *a priori* might perhaps suggest. It means only "what we knew before the experiment started", including both knowledge of the particular apparatus, and of the general laws of physics, and so forth, as known until that moment.

The "free coordinates" embody only a part of the *a priori* information, the part which is nonstatistical. (In MacKay's terminology this is the "structural" *a priori* information.) Statistical knowledge means that though we know nothing certain about what we are going to observe, we know in advance certain mean values which would emerge if we repeated the observation a very great number of times. The experimenter can make use of this knowledge, to some extent, in designing his experiments so as to reserve as much as possible of his information capacity for what is really new. This is a problem to which statisticians have given much attention long before the advent of communication theory, but interesting results may be expected from the impact of the Shannon-Wiener theory on this art. In the physical sciences statistical experiments are rather rare, and where correlations occur (e.g. between the disintegration products of nuclei), they are usually taken care of by special techniques. Hence we will leave out of account this potentially important contribution of communication theory to the natural sciences.

Finally, the information which arrives in the free coordinates or "logons", each of which can be considered as an independent communication channel (called by MacKay the "metrical" or *a posteriori* information) has to be broken down into selections from a finite number of possibilities. We saw an example of this when we constructed the proper scale for one cell of an electromagnetic signal. Thus the result of any experiment so treated will appear as a finite set of pure numbers, with certain confidence limits.

It may be observed that the basic element in this view of science is not an object, a "thing which exists", but the observation. The "thing" is defined only by the observational results. As a philosophical system this view has existed for a long time (at least since Kant) but its consequent application to science is still largely a task for the science of tomorrow. We shall not be able to add much to it here; our present task is only to justify confidence in the method by showing its heuristic power in problems whose solution is already known.

IV.2. The Free Coordinates of Light Signals.

The historically first, "geometrical" theory of light is completely unsatisfactory from our point of view. The basic element in the geometrical theory is the "ray of light", a geometrical line which may be straight or curved, according to the refractive properties of the medium. If this were taken seriously, point-to-point imaging of a plane on any other would be possible, simply by means of a *camera obscura,* whose aperture is itself a geometrical point.

This is evidently an over-abstraction of the same nature as the substitution of an analytical function for a signal, and we need not consider it further, since the difficulty of an infinite number of free coordinates is at once avoided in the classical wave theory.

To simplify matters, let us first consider an optically homogeneous medium, say the vacuum. In the wave theory of the vacuum the most convenient basic element of discourse is the plane, monochromatic wave. It will be sufficient to consider "scalar light" first, which is described by a scalar amplitude u. (The passage from this to "vector light", according to the electromagnetic theory is trivial; it is done by considering u as a vector at right angles to the propagation direction, i.e. by describing it by two numbers instead of by one.) The classical description of a plane wave is

$$u(x,y,z,t) = u_o \exp [j(k_1 x + k_2 y + k_3 z - \omega t)]$$
$$= u_o \exp [j(\mathbf{k}\cdot\mathbf{r} - \omega t)] \qquad \text{IV.1}$$

where in the second line we have used vector notation. The absolute value of the vector k is $2\pi/\lambda$; its points in the direction of the wave normal whose direction cosines are k_x/k, k_y/k, k_z/k. Where r is the radius vector (x,y,z).

Equation IV.1 is an undoubtedly correct description of an electromagnetic radio wave. Its application to light is open to the objection that (as we have seen in the previous section) no measurement of the temporal factor $\exp(-j\omega t)$ is possible, except at extraordinary intensities. Thus we can expect that it will drop out in the final description, but it would lead to great difficulties if we wanted to do without it from the start.

Equation IV.1 contains only one, very general assumption, and this is that the amplitude u, where it is measurable, behaves as if it were a function of space and time everywhere (or at least in certain domains of space and time) independently of the fact that it is observed. If one grants this, Eq. IV.1 means only that we want to apply four-dimensional Fourier analysis to this space-time phenomenon.

We now take from experience the fact that light propagates *in vacuo* with the constant velocity c. Mathematically this finds its expression in the wave equation

$$\Box\, u = \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right] u = 0 \qquad \text{IV.2}$$

Applied to a plane wave this gives

$$\omega = kc \qquad \text{IV.3}$$

In optics the frequency is never measured directly, but always calculated from the wavelength $\lambda = 2\pi/k$ by the last relation.

We can now apply this to the analysis of optical imaging if we only take one more fact from experience. This is, that at least for a certain class of objects (not too small and nonfluorescent), the wavelength of the light does not change appreciably in being transmitted or reflected by the object.

For convenience, let us take a plane object in the plane $z = 0$ and illuminate it, again for simplicity only, by a plane monochromatic wave incident from the z-axis. We can now apply Fourier analysis to the light field behind the object and find that in addition to the original $k_x = k_y = 0$, $k_z = k$, plane waves in all directions have appeared. Let us interpret this by assuming an amplitude transmission coefficient $t(x,y)$ such that if the amplitude of the incident plane wave immediately before the object was

$$u(x,y,-0,t) = u_o \exp(-j\omega t)$$

it has changed immediately behind the object into

$$u(x,y,+0,t) = t(x,y)\, u_o \exp(-j\omega t) \qquad \text{IV.4}$$

The function t(x,y) is"the object", by definition. It may be noted that it is defined for one method of illumination only. How it has to be modified if the illumination is changed does not concern us here, though we know from experience that for a certain class of objects (almost plane, with not too fine details) t is almost independent of the illumination. What we want to investigate is only how much of the function t (x,y) is transmitted by the light wave which issues from it. The simplest way to do this is to consider instead of t (x,y) its Fourier transform $T(\xi,\eta)$ in the "Fourier coordinates" ξ,η which will soon receive a physical interpretation. That is to say we put

$$t(x,y) = \int\int_{-\infty}^{\infty} T(\xi,\eta)\, e^{2\pi j\,(x\xi + y\eta)} d\xi d\eta \qquad \text{IV.5}$$

Geometrically this means that we imagine the density t of the object as composed of sinusoidal stripes of all periods and directions. We can now write down the relation IV.4 for z = 0 for each Fourier component $T(\xi,\eta)\, d\xi d\eta$ separately. Its contribution to the light amplitude immediately behind the object, at z = + 0, is

$$u_o T(\xi,\eta) \exp [2\pi j\,(x\xi + y\eta + ft)]\, d\xi d\eta$$

We can immediately extend this to all positive values of z, by writing instead

$$u_o T(\xi,\eta) \exp [2\pi j\,(x\xi + y\eta + z\zeta - ft)]\, d\xi d\eta \qquad \text{IV.5a}$$

This wavelet must satisfy the wave Eq. IV.2, which gives the condition

$$\zeta^2 = \frac{1}{\lambda^2} - (\xi^2 + \eta^2) \qquad \text{IV.6}$$

For the Fourier components of the object for which $\xi^2 + \eta^2 < 1/\lambda^2$ this gives plane waves, propagating in a direction whose direction cosines are

$$\cos a = \lambda\xi \quad \cos\beta = \lambda\eta \quad \cos\gamma = \lambda\zeta = [1 - \lambda^2(\xi^2 + \eta^2)]^{1/2} \qquad \text{IV.7}$$

This is the geometrical interpretation of the Fourier coordinates ξ,η. Apart from a factor they represent the direction cosines of the corresponding plane wavelet.

On the other hand, for those Fourier components of the object for which $\xi^2 + \eta^2 > 1/\lambda^2$ we obtain a quite different result. ξ is now imaginary, and we obtain evanescent waves, which are rapidly damped out in a matter of a few wavelengths from the object. This means that transmission by light always acts as if it contained a lowpass filter; the periodicities of the object finer than a wavelength are not propagated, i.e. detail below a certain minimum can never be imaged. This is a first limitation. We see that the transmission function t (x,y) has objective existence only within a certain "waveband" of spatial periodicity.

Thus we have the amplitude for the light wave behind the object in the form

$$u(x,y,z,t) = u_o e^{-2\pi jft} \int\int_{o\ o}^{\lambda^2(\xi^2+\eta^2)=1} T(\xi,\eta) \exp\left\{2\pi j\left[\frac{1}{\lambda^2} - (\xi^2 + \eta^2)\right]^{1/2} z\right\} e^{2\pi j\,(x\xi + y\eta)}\, d\xi d\eta \qquad \text{IV.8}$$

This is the expression for the propagation in free space without any aperture to restrict the plane waves. If small apertures are introduced at close distance, there results a complicated diffraction problem. If, however, the aperture is large enough and sufficiently far away, each plane wave can be considered approximately as a "ray", that is, replaced by its wave normal, starting, for example, from the center of the object; a result which we give without proof. In this case the only change in the above expression is that the integration limits must be replaced by those given by the aperture, that is, the summation is to be extended only over the directions given by Eq. IV.7, corresponding to rays which are not cut out.

It can now be seen that the situation is mathematically the same as the one which we had in the one-dimensional communication problem, only now we have to deal with two dimensions (in addition to time). Each object coordinate, x or y, now takes the place of time; the corresponding Fourier coordinates ξ or η take the place of the frequency. Thus, generalizing our previous results, we can say at once that the number of complex data which can be transmitted in a stationary transmission is

$$\iiiint dx\,dy\,d\xi\,d\eta$$

where the integration has to be carried out over the object x,y, and over these Fourier coordinates which are not cut out by apertures. In terms of the solid angle Ω admitted by the imaging system, we can also write for the degree of freedom

$$\iint dS\,d\Omega/\lambda^2 \qquad \text{IV.9}$$

where dS means the object area. This is the number of complex data; the number of real data is twice as much (26).

Now we can at once generalize this to a time-variable transmission by joining to it the factor 2FT, where F is the frequency band of the light used and T is the observation or transmission time. In order to unify the mathematical form, we write the final result:

$$\text{Number of degrees of freedom} = 2.2.(2) \iiint \frac{dS}{\lambda^2}\,d\Omega\,df\,dt \qquad \text{IV.10}$$

The first factor 2 is due to our counting real data, the second to the vector nature of light (two independent transmissions are possible by using polarized light), the last factor 2 belongs to the "temporal phase". In view of the difficulties of measuring this last factor, discussed in the previous section, we have put this in brackets.

In the derivation (or, rather, explanation of this expression we have used plane waves, but it is evident that we could have just as well used other "elementary waves", as we used other elementary signals in the section on signal analysis. "Gaussian light bundles" are again particularly simple, but their application to this case may be left to the reader.

The main result, so far, is that an optical image, by whatever instrument it is produced, is not essentially different from a television picture. It has only a finite degree of freedom; that is, it contains only a finite number of independent points. We see again that the continuum, in the mathematical sense, is only an auxiliary construction which disappears when we come to observable quantities.

IV.3. An Imaginary Experiment with Light.

By allowing us to formulate the degrees of freedom in an optical image, the wave theory of light has brought us a step further in the application of the principles of information theory, but not far enough. Doubts arise already from inspecting Eq. IV.10. We have counted the degrees of freedom, but have they really all the same value? We could make the division into elementary light bundles, for example, in such a way that each beam covers the whole object area but only a fraction of the aperture area. Now we cover this up, not entirely, but using a screen which cuts out 0.999 of the light. Can we still count this as a degree of freedom?

Classical light theory is entirely unable to answer questions such as this, since it contains no natural unit for amplitude or energy. This, and the answer, can come only from quantum theory. But while in the last section we have taken the results from physics, we will now argue the existence of photons from an epistemological consideration.

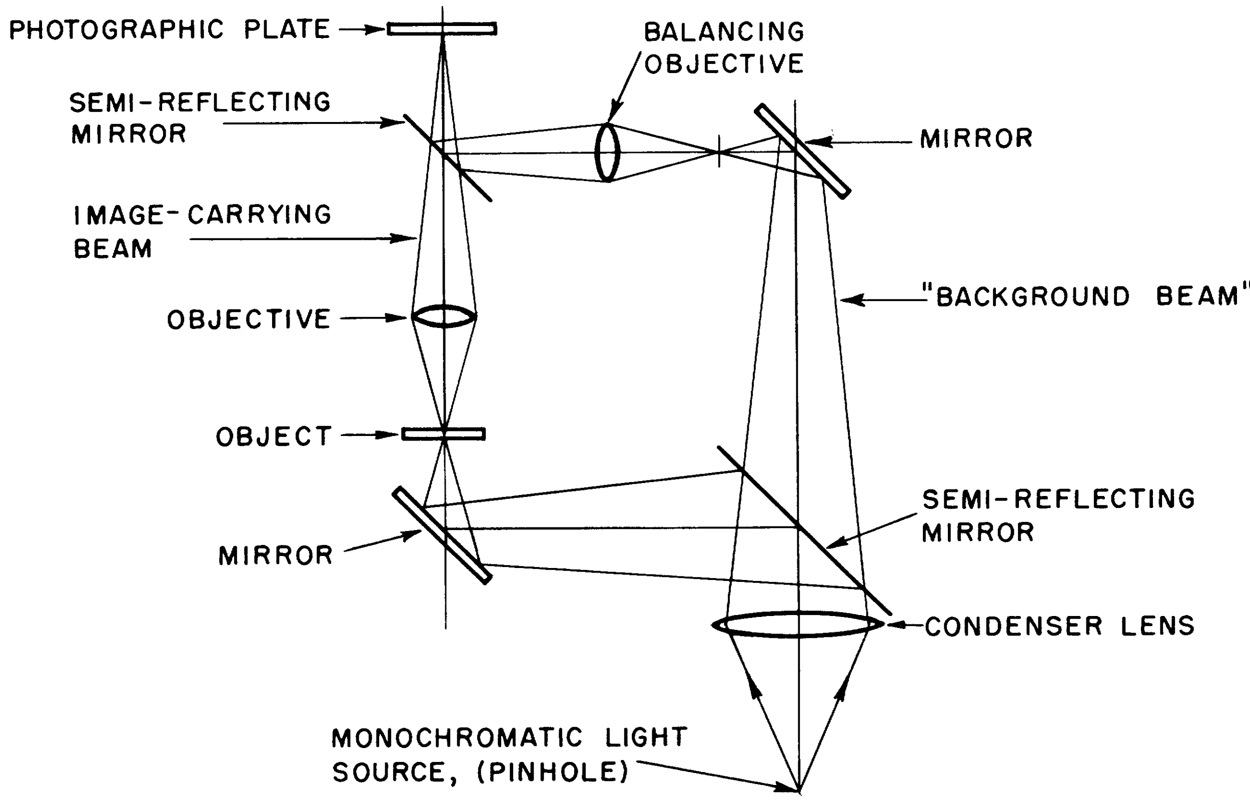

Fig. 8

Imaginary experiment, to lead *ad absurdum* the classical theory of light which allows "observation without illumination".

We take from experience the evident fact that a recording instrument, such as a photographic plate, can record a light signal only if it has some finite energy. Yet we will show that it can record an image of the object even if the object has been illuminated by light of any small energy, which can be made zero in the limit.

Consider the interferometric arrangement as shown in Fig. 8. It differs from an ordinary Jamin-Mach interferometer only in that we have introduced two lenses in its two branches. One is an objective, which images an object on a photographic plate; the other is only a "balancing" lens to make the two optical paths approximately equal. Also, contrary to ordinary interferometer practice, we make the light intensities in the two branches very unequal. Only a small part, say 0.0001, of the intensity goes through the object; the rest, the "background", which carries no information, goes through the other arm. The illumination is coherent, which means that the light source is approximately monochromatic, and a sufficiently small pinhole is used as the source.

For simplicity we consider only one "point" of the object, that is, a patch equal in diameter to the resolution limit; and we assume that this "point" is either black or white. In other words, we are asking only for one "bit" of information: "Is the patch covered or not?"

The light is coherent, hence we must form the vector sum of the amplitudes A (of the background) and a (of the image-carrying wave) and take the absolute square of this sum to obtain the resulting intensity.

$$A^2 + a^2 + 2Aa \cos \phi \qquad \text{IV.11}$$

where ϕ is the phase angle between the two waves. This is not the temporal, but the spatial phase, resulting from the geometrical path differences, a measurable quantity. Thus the intensity consists of three terms. The first, A^2, is known in advance and carries no information. The second, a^2, is negligibly small, since we have assumed $a^2 << A^2$. The third could be zero if ϕ happens to be $\pm\frac{1}{2}\pi$, but for safety we can carry out two experiments, in one of which we introduce a quarter-wave plate into one of the beams so that in one of the two experiments $\cos \phi$ is at least $\pm 1/\sqrt{2}$. Thus this term of the order $A \cdot a$ can be considered as the image-carrying term. The first large term A^2, being known and constant, can be subtracted.*

This term will betray the presence or nonpresence of the object point, provided that it carries enough energy in the observation time T. Let us write $A^2T = E$ and $a^2T = e$. The crossterm represents an energy signal of the order

$$(e \cdot E)^{1/2} \qquad \text{IV.12}$$

and this will be observed, if it is sufficiently large. But with our assumptions, however small we make e, (the light which has gone through the object), we can always make E (the background) so large that we get an observable signal. Thus we come in the limit to the paradox of observation without illumination.

The flaw in this argument evidently must be in the constant and accurately predictable background which we have subtracted. Let us now assume that the experiment is impossible because a limit will be reached at an energy e_o, where the fluctuations of the background E outweigh the image-carrying term; that is, we postulate that for $e < e_o$

$$\overline{\delta E^2} > E \cdot e_o \qquad \text{IV.13}$$

Dividing by e_o^2, we have at the limit

$$\overline{\delta (E/e_o)^2} = E/e_o \qquad \text{IV.14}$$

This means that the energy, measured in a certain unit e_o(i.e. expressed as a number $N = E/e_o$) obeys the fluctuation law

$$\overline{\delta N^2} = N$$

This, however, is Poisson's fluctuation law, which can be interpreted by saying that the energy behaves as if it were arriving in discrete units e_o, which have the property that at least one of these is required to carry a bit of information. We see that the photons have again made their appearance, as a result of the simple postulate that observation without illumination is impossible.**

By an extension of this experiment one can also prove that e_o must be inversely proportional to the wavelength, by postulating that a rotating mirror between the object and the plate cannot make something observable which would not be otherwise visible. This extension emerged in conversations of the author with Norbert Wiener.

*Electrical methods for subtracting a constant background are well-known, but there are also convenient optical methods, such as the "schlieren" method. The background can be subtracted before it reaches the photographic plate.

**One might make the objection that the classical theory also recognizes fluctuations; those of thermal origin. But a light source at the absolute zero of temperature is not classically impossible; for example, an electric dipole, tied to a rotating shaft, and kept at T=0 together with the rest of the apparatus.

At this point we have reached, by a heuristic argument, the starting point of the previous section, whose results we can now take over in order to complete the picture of the information carried by light.

IV.4. The Capacity of an Optical Transmission.

Let us first separate the "time-frequency cell" $\Delta t\ \Delta f$ from the integrand in Eq. IV.10, and put it equal to unity. By this we have imposed the coherence condition that all elementary beams which arise by diffraction from an illuminating beam satisfying the condition $\Delta t\ \Delta f = 1$ are coherent in the sense that they can be described by one time-function in the classical theory; they form only one complex wave. In quantum theory this means that the N photons of the illuminating beam are distributed over these subject only to the condition that their sum is equal to N. Thus the mathematical problem is that of counting the distinguishable configurations which can be produced by distributing up to N photons over F degrees of freedom.

Let n_i be the number of photons in the i-th degree or "logon". We have the condition

$$n_i + \dots + n_i + \dots + n_F \leq N \qquad \text{IV.15}$$

where all n_i are positive. By Eq. III.18, the number of distinguishable steps in i up to n_i is

$$s_i = 2(1 + 2n_T)^{-\frac{1}{2}}\, n_i^{\frac{1}{2}} \qquad \text{IV.16}$$

Let us now imagine the s_i as Cartesian coordinates in an F-dimensional space. The number of distinguishable configurations is now equal to the number of integer lattice points inside a hypersphere with the radius

$$R = 2(1 + 2n_T)^{-\frac{1}{2}} N^{\frac{1}{2}} \qquad \text{IV.17}$$

This radius is the maximum number of steps in i; it corresponds to the case when all the N photons appear in this one logon. The total number of lattice points is equal to the volume of that sector of the hypersphere in which all s_i are positive, which is, by a known formula,

$$P = \frac{(\pi/4)^{\frac{1}{2}F}}{\Gamma(\frac{1}{2}F + 1)} R^F = \frac{1}{\Gamma(\frac{1}{2}F+1)} \left(\frac{\pi N}{1 + 2n_T}\right)^{\frac{1}{2}F} \qquad \text{IV.18}$$

The information capacity is the logarithm of this number P. Using Stirling's formula

$$\log \Gamma(\tfrac{1}{2}F + 1) \approx \tfrac{1}{2}F\,(\log \tfrac{1}{2}F - 1)$$

valid for large F, we obtain

$$\log P = \tfrac{1}{2}F \log \frac{2\pi e N}{F(1 + 2n_T)} \qquad \text{IV.19}$$

This is the maximum capacity in which the appearance of photons in the elementary diffracted beams is restricted only by the conditions IV.15. Now, we can also decide the question, "What is the capacity of an optical channel which is partly obscured by absorbing screens?"* Let us assume that the screens have the effect that the transmission in the i-th logon is $\tau_i < 1$, so that the maximum energy which can appear in it is not N but $\tau_i N$.

*The expansion of the light wave into elementary beams must be fitted to the position of the absorbing screens. For instance, if there is a screen only in a certain aperture plane, it is convenient to use elementary beams which have a minimum cross section in this plane but cover the whole object area.

This problem is easily reduced to the previous one, if we replace the n_i by $n_i r_i$. We now have to consider the volume of a hyperellipsoid with semi-axes $R r_i$. The transmission capacity is

$$\log P = \tfrac{1}{2} \sum_{o}^{F} \log \frac{2\pi e N r}{F(1 + 2n_T)} \qquad \text{IV.20}$$

This formula at last gives an answer to our objections to the classical theory; the degrees of freedom are no longer counted, but are properly weighted. It may be noted that the formula is, of course, an asymptotic one; it must not be extended to transmissions r_i so small that an added degree of freedom appears to make a negative contribution to it. This happens if the argument of the logarithm falls below unity, that is, we must cut out those degrees of freedom for which

$$r_i \frac{N}{F} = \bar{n}_i \, r_i \leqslant \frac{1}{2\pi e} (1 + 2n_T) \qquad \text{IV.21}$$

The left side is the mean number of photons which will appear in this logon. The numerical factor is about 1/17.

It may be noted that the quantity log P differs only slightly from the physical entropy of the imaging light beam. By Max Planck's definition (27) the entropy of a system is k times the logarithm of the number P of ways in which a given energy can be distributed over the possible states of a system. The numerical difference between "a given energy" and "up to a given energy" is very small in the case of systems with a high degree of freedom; it is the difference between the surface and the volume of a hypersphere of F dimensions.

Space prevents us from going into the discussion of the connections and the differences between physical and selective entropy. In addition to the paper by Szilárd, reference may be made to two recent studies by Brillouin (28).

V. Problems of Speech Analysis.

V.1. Analysis, Compression, Recognition.

Telephony started with Alexander Graham Bell's idea of the faithful reproduction of pressure waves by "undulatory currents". This was indeed the only sound starting point so long as no adequate knowledge was available on the restrictions to which the message is subjected at the transmitter (the human voice organs) and at the receiver (the human ear plus brain). It is interesting that Bell's highly successful idea was preceded by other, unsuccessful ones, which to us appear much more modern. Bell's own "harmonic telegraph" was a forerunner of the vocoder; the Reis telephone was an on-off device, a "speech clipper"; and Bell's father, Melville Bell, spent many years on a phonetic writing based on the principal positions of the vocal organs. It appears a general principle that economy must step in towards the end of a development, not at its beginning.

The imperfect economy of the telephone was noticed as soon as Nyquist and Küpfmüller had established the principle: "The capacity of a line is proportional to the bandwidth". Good quality speech communication requires about 4000 cycles, and the intelligibility is still satisfactory if the band is restricted to the range of 400-2400 cycles, that is, to a band of 2 kc. This is sufficient for transmitting a maximum of about 20 letters per second, which could have been transmitted by telegraphy in a band of 100 cycles, using the 5-digit binary Baudot code. The economy figures even worse if we make the estimate in terms of "bits", as was first done by Hartley. Twenty letters (including spaces) represent 100 bits on the basis of the 32-letter alphabet. On the other hand, the capacity of a channel of 2000 cycles, with a signal-to-noise ratio of 30 db, is

$$2000 \log_2 10^3 = 2000 \times 3.32 \times 3 = 2.10^4 \text{ bits/sec}$$

Thus it appears that telephony utilizes the channel only to 0.5 percent.* It must not be forgotten, however, that here we have made the comparison by equating the spoken word with its message when it is printed. But the spoken word communicates, in addition, the "emotional" content of the message. (This term, coined by Harvey Fletcher, does not necessarily mean violent emotions; it comprises, for example, all the features by which we recognize the identity of the speaker.) It may well be asked whether the emotional content is worth in all cases 99.5 percent of the channel capacity, and whether we could not convey it in less.

Thus the first problem which faces us is that of condensation or "compression" of the message into a narrower band, and this again splits naturally into two types:

(1) "Commercial quality" speech compression, which could be used as a substitute for ordinary telephony;

(2) Transmission of the bare intelligence content of the speech.

One problem, of course, shades into the other, but it will be useful to keep them separate.

The second type of problem leads naturally to a third. If the speech is boiled down to its "printable" skeleton, why not print it? The message might be handed out in visual form, on a tape, or on a sheet. This problem can again be split into two parts:

(3) The message is visually recognizable, but it has a code of its own. This could be called the problem of the "mechanical stenographer".

(4) The message is spelled out in the letters of the alphabet. This means that the work of speech-sound recognition is carried out by the machine. This is the problem of the "mechanical typist" or of the "mechanical telegraphist".

We will deal with these problems briefly, though not necessarily in this order. There is an essential difference between the first two and the last two: in the first group the receiver is the ear; in the second, the eye. Hence, an approach to the first group must necessarily be preceded by a discussion of hearing. This is also indispensable for the second group, because we can spell out only what we hear (though not all of it).

Another approach to all four groups is through the vocal organs. The human vocal apparatus can produce only sounds of certain types, and not all of these are actually used in conversation, at least not in any one language. A preselection of the sound patterns to be transmitted is evidently a very powerful principle of condensation. It is seen that this problem involves two very different sciences: the physiology of the vocal organs, on the one hand; phonetics and linguistics, on the other.

The fourth problem, speech recognition, is the most difficult of all, since it involves all the problems of the previous groups, and in addition, a higher function of intelligence, that of pattern identification,** which has to be taken over from the human brain by the machine. None of the four problems has yet received a fully satisfactory solution (No. 3 is perhaps the most advanced), but it can be foreseen that speech recognition will involve the greatest amount of work.

V.2. Ohm's Law of Acoustics.

The first and most general simplification which we can apply to any audio signal is the elimination of the phases, because by what is usually known as Ohm's Law of Acoustics, the ear is not

* This becomes even four to five times worse in Shannon's statistical theory, if the correlations are taken into account in a given language, for example, in English.

** The German word *"Gestalt"* is often used for "pattern" in this connection, after the German school of *"Gestalt* psychologists" whose chief representatives were Köhler, Wertheimer, and V. Hornbostel.

supposed to be sensitive to it.

This law requires some investigation and restatement, all the more because in its usual interpretation it is notoriously wrong. The communication engineer will interpret amplitudes and phases in the sense of the Fourier spectrum. In this sense Ohm's law would mean: If we have a signal which is a function s(t) of time, which has a Fourier transform S(f), the ear cannot tell the difference between this signal and another, s'(t), whose transform is S(f) exp [jF(f)], where the phase shift F is any arbitrary function of f. But this is evidently wrong. Every communication engineer knows that unless F is made a linear function of the frequency the signal will not only suffer a delay, but it will be distorted. For example, the high frequencies will arrive before the lower ones, causing a chirping sound before every word. Sometimes there is an attempt to qualify Ohm's law by the uncertain statement, "The ear is insensitive to small phase differences", which is also wrong. However, we need not go into a discussion of this.

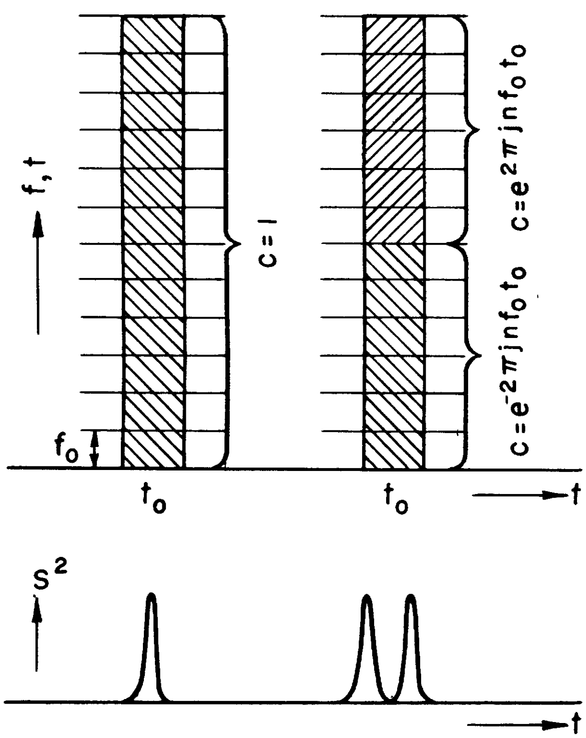

Fig. 9.

Disproving Ohm's Law of Acoustics. It is shown that a pulse can be split into two by varying the phases of its components.

It is evident that Ohm's law needs restatement. Let us try to replace Fourier analysis by analysis into gaussian elementary signals, to which, as we have seen, the concepts of "amplitude" and phase can be directly transferred. Ohm's law, if it is valid in this interpretation, must mean this: "There is no audible difference between any signal in a gaussian representation with properly chosen Δt and $\Delta f = 1/\Delta t$ and another whose expansion coefficients differ from the c_{ik} only in a unitary factor".

We will now show that Ohm's law is not exactly valid, even in this interpretation, with any choice of the cell aspect ratio. To show this, consider a sharp click whose matrix representation is particularly simple. (See Fig. 9.) Its expansion coefficient is constant in absolute value, say unity, in the column in which it occurs, and is zero outside. We can prove easily that such a column represents a sharp click. For simplicity let us write f_o in the place of $\frac{1}{2}\Delta f$. If the c_{ik} are all unitary in a column extending from $f_1 = n_1 f_o$ to $f_2 = n_2 f_o$, we can represent with them, for example, a time function

$$s(t) = e^{-a^2 t} \sum_{n_1}^{n_2} e^{2\pi j n f_o (t-t_o)} \qquad \text{V.1}$$

The sum here is a simple geometrical series, and its real part can be written in the form

$$s(t) = e^{-a^2 t} \cos \pi (n_1 + n_2) f_o (t-t_o) \frac{\sin \pi (n_2 - n_1 - 1) f_o (t-t_o)}{\sin \pi f_o (t-t_o)} \qquad \text{V.2}$$

It is easy to see that this represents a click, occurring at $t = t_o$, which will be sharp if $n_2 - n_1$ is a sufficiently large number. We see that by modifying the phases of the expansion coefficients, we can shift the click about. We cannot, however, shift it by much more than about the cell length $\Delta t = 1/\Delta f = 1/2 f_o$, because beyond this it will be cut out by the gaussian factor.

Let us now imagine, for instance, that we first make all coefficients equal and real over the whole column, from $f = 0$ to the maximum $f = F$. This gives one click, occurring at $t = 0$, in the middle of the column. Next let us split the column into two halves. In the upper half we modify the phases so that the click is shifted to $t_0 = +\frac{1}{2}\Delta t$; in the lower half we shift it to $-\frac{1}{2}\Delta t$. We have now produced two clicks instead of one, simply by modifying the phases.

The ear may not be able to tell these two clicks from one another if Δt is less than approximately 10 msec. But in this case we can test Ohm's law with the "dual" pattern, which consists of a single row in place of a column. By the perfect symmetry in time and frequency of the gaussian representation we can say at once that we can now produce, in the first case, a single pure tone; in the second, two tones succeeding one another and differing in frequency by $\Delta f = 1/\Delta t$. If the two clicks could not be distinguished from one another, because Δt was less than 10 msec, this frequency jump will be more than 100 cps, which can certainly be distinguished anywhere in the audio range.

Thus Ohm's law never holds exactly, whatever interpretation we put to the "phases", because we can always produce a test sound which by suitable manipulation of the phases we can change into something distinctly different.

The reason that Ohm's law is still almost generally accepted (and this finds its expression in the fact that the "visible sound" patterns which we discuss in the next section are considered as complete representations as far as the ear is concerned) is twofold. First, such abrupt phase changes as those that we have considered rarely occur in sound-reproducing apparatus. Second, in the case of speech, one can choose the cell so that both the time displacement of clicks and the shifts in frequency become almost if not entirely inaudible. It appears that the best choice is Δt = 25 msec, Δf = 40 cycles. Thus, at least with this interpretation, we can accept Ohm's law as a good practical rule which holds true for at least the great majority of the sound patterns and phase-distortion patterns with which we have to deal in practice.*

V.3. Visible Speech.

Accepting Ohm's law with these reservations, we can produce a representation of a speech signal which does not contain the phase simply by taking our previous "matrix" or "cell" representation and writing the energy $c_{ik}c_{ik}^{\dagger}$ into every cell, instead of the complex amplitude c_{ik}. This can also be done with elementary functions other than gaussian, provided they are sufficiently "compact"; that is, that they have an uncertainty product not much exceeding unity.

If instead of assigning a number to every point, we indicate the energy by the density of shading and merge the densities into one another by some interpolation method, we arrive at the "visible speech" representation which has been worked out experimentally by R.K. Potter and his collaborators at the Bell Telephone Laboratories from 1941 onwards (29). This has now become the most powerful tool of speech analysis, and the technical methods used in producing the "sound spectrograms" (now usually called "sonograms") are too well known to require description.

Sonograms taken with different analyzing filter widths are all equivalent; they correspond to representations with different cell aspect-ratios, but they have very different appearances. There is a maximum of visual detail in those taken with a bandwidth of approximately 45 cycles. This detail, however, appears to be mostly redundant, and even confusing, from the point of view of recognition, and a filter 300 cycles in bandwidth is used with preference both by deaf people for reading and for learning the correct pronunciation of speech sounds, and for the purposes of speech research.

* It can be proved very simply that sonograms are far from being complete representations once we go sufficiently far off the optimum division. "Amplitude clipped speech" is almost completely intelligible, but if we produced a sonogram of it with a filter width of the whole audio range, the result would be a uniform black surface!

We will return to this question later.

It may be asked whether or not a two-dimensional representation is necessary if one wants to eliminate the phase. The answer is that any number of signals can be substituted for s(t) in which the phase carries no information because it has been standardized in some way. One method of doing this would be, for instance, to use only elementary functions of the cosine or even type, whose amplitude is proportional to the square root of energy in the cell, and add these up to obtain the new signal with standardized phases. There is, however, no evidence that these signals would look simpler than the original.

The short-time autocorrelation function as a substitute for the short-time energy spectrum has already been discussed in section II.3. If it were plotted at right angles to time, and again shown in terms of densities, this would give a highly interesting alternative to visible speech.

V.4. The Vocoder.

After two decades of work by Harvey Fletcher and his collaborators in the Bell Telephone Laboratories, the science of speech and hearing resulted in the invention of the vocoder by Homer Dudley (30, 31). This instrument has given, in return, some of the most important contributions of the last decade to this science.

The vocoder is too well-known to be described in detail. It will only be recalled that it can be used in two ways: as an instrument which conveys intelligibility in the form of a hoarse whisper; and in a more perfect form, as an instrument in which the character of the voiced sounds is imitated, as if a larynx were in operation. In the first method the voice spectrum is analyzed by a number of filters, usually ten, covering the frequency band from approximately 250 cycles to 2750 cycles. The output of each channel is rectified; transmission of the rectified output (the envelope) requires a channel that is only 25 cycles (single sideband) wide. At the receiver this is remodulated with "white noise" filtered through the original channel. In spite of the unnatural character of the sound it can be 90 percent intelligible. This is called the "hiss" operation.

In the voiced or "buzz" operation the pitch of the speaker (the larynx frequency) is continuously measured whenever it steps above a certain minimum level. The pitch frequency is translated into a current intensity, which is transmitted again through a channel (the 11th) to the receiver. Here it is again translated into the basic frequency of a buzzer, which produces the pitch and all its overtones. This is used instead of the white noise as the energy source, modulated by the 10 channel outputs.

One of the most important results obtained with the vocoder is that the pitch contributes almost nothing to the intelligibility (apart from a certain improvement in the start and stop consonants; in some languages, like Chinese, this statement does not hold). Whether the buzzer is switched off, set to a constant pitch, or made to follow the speaker's larynx frequency makes hardly any difference, proving that the intelligibility is mainly conveyed by the spectral energy envelope, while the emotional content of the speech is mainly carried by the inflections of the pitch. Moreover, it is sufficient to analyze the envelope only in steps of about 250 cycles each, which checks in an interesting way with the observation that visible speech is perfectly legible with a filter bandwidth of approximately 300 cycles. This almost perfectly intelligible speech is, however, far from natural. It has a "machine-made" character, with noisy transients especially at the beginning and at the end of words. It becomes natural only when the number of channels is increased to 30, each with a bandwidth of approximately 80 cycles, as in the new vocoder of the Bell Laboratories. This can be considered as the first successful attempt at condensed speech of commercial quality. The gain in waveband is, however, not very large. Each channel (vestigial single sideband with guard band) requires approximately 40 cycles, making about 1200 cycles in all, a gain of little more than a factor of 2.

Intelligible but not natural speech can be transmitted in 400 cycles to 450 cycles.

V.5. Painted Speech.

Visible speech and the vocoder were combined in an important experiment. The "bars" in a sonogram were painted black where their density appeared more than a certain minimum; all the rest was left blank. This record was scanned simultaneously with 10 photocells, whose output was connected with the channels of a vocoder. The result was a certain degree of intelligibility, not sufficient for communication, but surprisingly high considering that the gradations in the amplitude had been entirely thrown away.

This line of research was continued, in particular by Cooper and by Vilbig. A 300-cycle sonogram proved to be a suitable starting point. This was painted in, or copied on a second sheet, with certain rules of schematization, as roughly indicated in Fig. 10. Playing this back with an instrument equivalent in effect to a 30-channel vocoder, i.e. with a bandwidth of about 100 cycles, the intelligibility reached about the same high level as in the 10-channel vocoder in the "hiss" operation, which is highly remarkable in view of the fact that the spectral envelope has now only a very superficial resemblance to the original, since it has only two levels: "something" and "nothing".

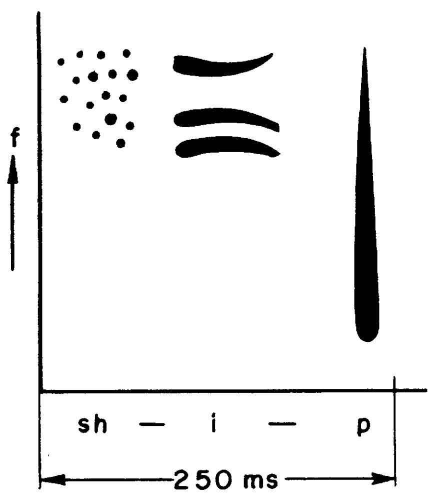

Fig. 10.

Example of a "low definition" sonogram.

On the basis of this experience we can now make a new estimate of the channel capacity required for intelligible, not natural, speech. Let us divide the time-frequency plane into cell of Δt = 25-msec duration, Δf = 40-cycle bandwidth. This cell size seems to have special significance for the analysis of audio signals. We have already met it when stating that Ohm's law appears to be best satisfied with this division, and a second time when noting that sonograms taken with a 45-cycle bandwidth have a maximum of visual detail. But a time interval of 25 msec appears to have an even more fundamental significance. It seems to mark the limit between spectral analysis and time analysis by the human ear, at least if the object presented to it is speech. It has been known for a long time that 25 msec is the lower limit at which a speech echo as such becomes noticeable; a time delay of less than this only gives a "fuller" character to the speech. The echo begins to be very disturbing at 50 msec; this is the point where the delayed message begins to become intelligible in itself. A further argument in support of this cell division is the fact that it was found that the limits of the bars in painted speech had to be accurate to within 30 to 40 cycles.

Thus, dividing the time into units of 25 msec, (see Fig. 11) and the frequency band of, say, 2640 cycles into 64 strips, it is found that in each time interval it is sufficient to transmit, at most, 6 numbers, under 64, which mark the upper and lower limits of the three formant bars or the limits of the "dots" by which, for example, the "sh" has been replaced in the previous figure. As each cell can be identified by 8 binary selections, this means that painted speech can be transmitted with good intelligibility in a channel of not more than

$$40 \times 6 \times 8 = 1920 \text{ bits / second}$$

capacity. This corresponds approximately to a channel 200 cycles wide, with 60-db signal-to-noise,

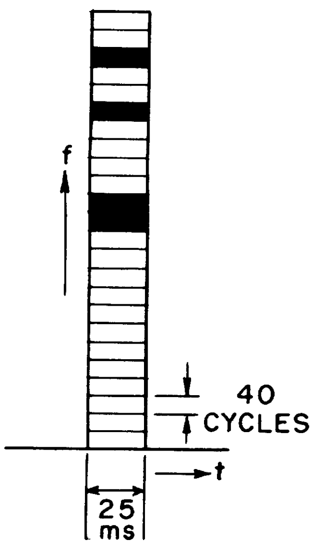

Fig. 11.

"Yes or no" simplification of the amplitudes transmitted in an elementary time interval.

i.e. to tenfold compression. A further gain by about the same factor would be possible if we took into account the fact that the configurations which will actually occur in speech, in a given language, are actually only a small selection among those possible. A machine, however, which could make use of this fact would be a speech-recognizing machine, and we will deal with this possibility in a later section.

V.6. Speech Analysis in Terms of Moments.

We have learned from this that the transmission of intelligible speech requires the transmission of six suitable selected data, 40 times per second. But the six data in the last section were obtained by a highly nonlinear method, via a visible speech pattern, which was intelligently simplified and schematized. Could we not obtain six significant data also by a linear analysis?

The first method which suggests itself is the Method of Moments, discussed in section II.5. We know that the short-time energy spectrum belonging to a time interval T (25 msec with preference), which we will call S_T^2, can be expressed by the sequence of its moments of even order. We recall their definition, for a real signal s(t) which exists only in the time interval T preceding the present instant t = 0, and some of their properties:

$$M_o = 2 \int_o^F S_T^2 (f)\, df = 2 \int_{-T}^o s^2 (t)\, dt \qquad \text{V.3}$$

$$M_2 = 2 \int_o^F f^2 S_T^2 (f)\, df = - \frac{2}{(2\pi)^2} \int_{-T}^o s \cdot \frac{d^2}{dt^2} s\, dt = \frac{2}{(2\pi)^2} \int_{-T}^o \left(\frac{ds}{dt}\right)^2 dt \qquad \text{V.4}$$

$$M_4 = 2 \int_o^F f^4 S_T^2 f\, df = \frac{2}{(2\pi)^4} \int_{-T}^o s \cdot \frac{d^4}{dt^4} s\, dt = \frac{2}{(2\pi)^4} \int_{-T}^o \left(\frac{d^2 s}{dt^2}\right)^2 dt \qquad \text{V.5}$$

and so on. The right-hand sides show the important property of these moments, that they can be determined directly without spectral analysis. We must only differentiate the signal successively, square the derivatives, and integrate them.

There is no doubt that the sequence of the moments contains the intelligence, but to what order must we go, and in what form is the essence of the information contained in the sequence? In order to decide this question, we apply the following method.

We know that speech can be processed in certain ways, without losing its intelligibility. If we subject a speech wave to such a nondestructive transformation, the intelligence must be contained in the invariants or near-invariants of the transformation.

The first test to which we submit the moments is differentiation of the signal. This is the same as passing it through a network whose amplitude transmission is proportional to the frequency. It is known that speech can be differentiated twice, and that it rather gains than loses intelligibility, especially in the presence of background noise. (It would be of interest to investigate how often one can repeat this process.) Now one can see immediately that by differentiation each moment steps up one in the sequence, that is, M_o takes on the value of M_2 and so on. This might suggest that the sequence of moments is indeed significant, but that it is only the relation of consecutive moments that matters, not their order. This, however, leads to difficulties which are more clearly seen if instead of the moments we consider the mean frequencies of different order:

$$f_2 = (M_2/M_o)^{1/2} \quad f_4 = (M_4/M_o)^{1/4} \quad \ldots \; f_n = (M_n/M_o)^{1/n} \qquad \text{V.6}$$

Differentiation of s(t) changes this into the sequence

$$f'_2 = (M_4/M_2)^{1/2} \quad f'_4 = (M_6/M_2)^{1/4} \quad \ldots \; f'_n = (M_{n+2}/M_2)^{1/n} \qquad \text{V.7}$$

Thus the sequence of mean frequencies is not only renumbered but significantly changed.

A second, and even more decisive test is that of speech clipping. As Licklider (32) and others have shown, speech retains a high degree of its intelligibility if its amplitude s (t) is clipped at a low level, so that only the zero crossings retain their position, but the amplitudes are all replaced by a standard amplitude. Moreover, as E.C. Cherry, in particular, has shown (33), the intelligibility is retained if after clipping the signal is differentiated, that is, if zero crossings are replaced by spikes, with signs according to the direction of the crossing. If, however, the wave thus obtained is rectified, most of the intelligibility is lost, and in particular, all vowels sound the same. (See Fig. 12.)

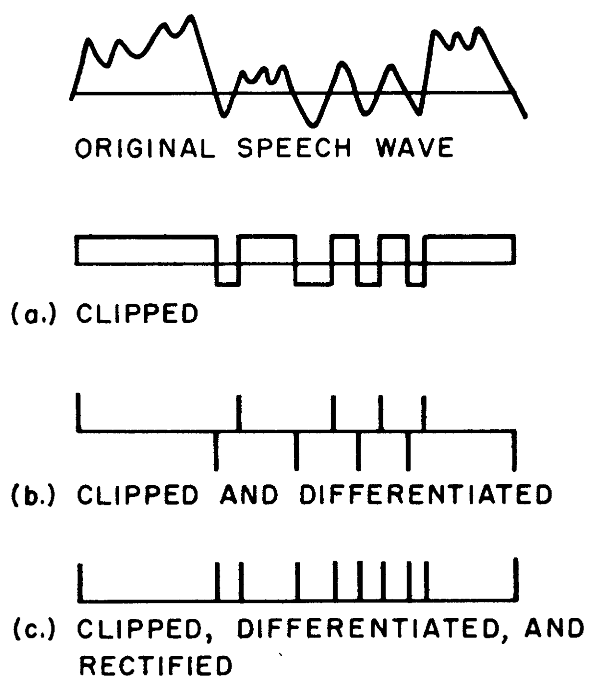

Fig. 12.

Varieties of speech clipping.

By a rather long calculation, of which only the results can be given here, the author has investigated what happens to the moments and to the mean frequencies of different order if the input is a single frequency f, and the waveband admitted is F. If the tone f were faithfully reproduced, all mean frequencies would be equal to one another and to f. Figure 13 shows the colossal distortion produced by clipping. Even at zero signal frequency the second-order mean output frequency is 0.58F; the fourth order, even 0.67F. Thus their whole range of variation is only a fraction of F if the input frequency runs from 0 to F. In case (c) the distortion is even worse. The mean output frequencies actually drop if the input frequency is rising Yet speech clipped, differentiated, and rectified still has a certain degree of intelligibility, especially for consonants.

We must conclude that the sequence of moments and of mean frequencies, which suggests itself so readily as a mathematical tool, is in fact very unsuitable for the analysis of speech. There is no doubt that it contains the intelligence because it contains everything, but to extract from it what is essential would require very complicated further operations.

V.7. Speech Analysis in Terms of Hermitian Functions.

Though the moment approach fails, this is no reason to give up linear analysis altogether. In the moment method we have formed a sequence of values by multiplication of the energy spectrum

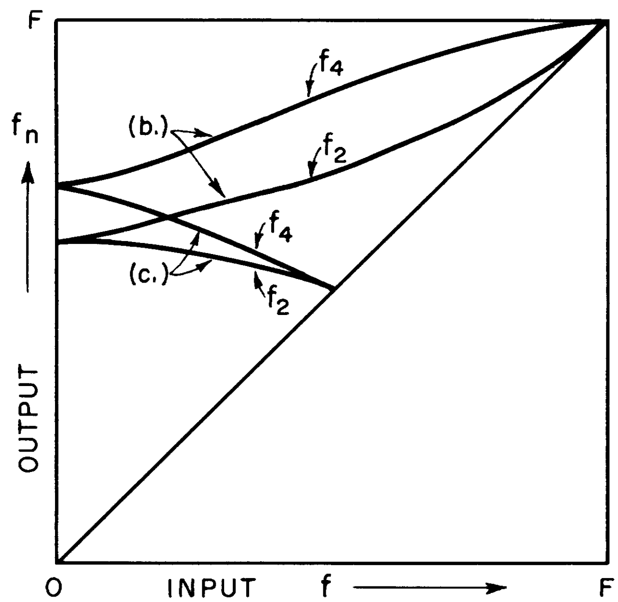

Fig. 13.

The first two effective frequencies resulting from two different methods of clipping a harmonic signal of frequency f.

with an even power of f, followed by integration over the spectrum. Let us now try to replace the powers by some more suitable set of functions. Orthogonal sets suggest themselves naturally because, in this case, the result of the integration gives the expansion of the energy spectrum in terms of the chosen set.

Hermite's orthogonal functions offer special advantages because of their self-reciprocity in Fourier analysis. It will be shown that if we expand the energy spectrum in terms of these, we obtain at the same time the expansion of the autocorrelation function.

It is convenient to start from the definition of the ϕ- functions (also called "parabolic cylinder functions") as listed by Campbell and Foster (ref. 7):

$$\phi_n (f/f_o) = e^{\pi f^2} \frac{d^n}{df^n} (e^{-2\pi f^2}) = (-)^n e^{-\frac{1}{4}x^2} (4\pi)^{\frac{1}{2}n} H_n(x) \qquad \text{V.8}$$

where $x = (4\pi)^{\frac{1}{2}} f/f_o$ and $H_n(x)$ is the n-th Hermite polynomial

$$H_n(x) = (-)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2} = x^n - \frac{n(n-1)}{2} x^{n-2} + \frac{n(n-1)(n-3)}{2.4} x^{n-4} - \dots \qquad \text{V.9}$$

(This definition is slightly different from that used in many textbooks on wave mechanics.) For these functions there exists the simple relation:

The right-hand Fourier transform of $\phi_n (f/f_o)$ is $j^n \phi_n (f_o t)$ V.10

The orthonormal Hermite functions follow from these by multiplication with a normalizing factor

$$h_n(x) = \frac{2^{\frac{1}{4}}}{(4\pi)^{\frac{1}{2}n} n!} \phi_n(x) \qquad \text{V.11}$$

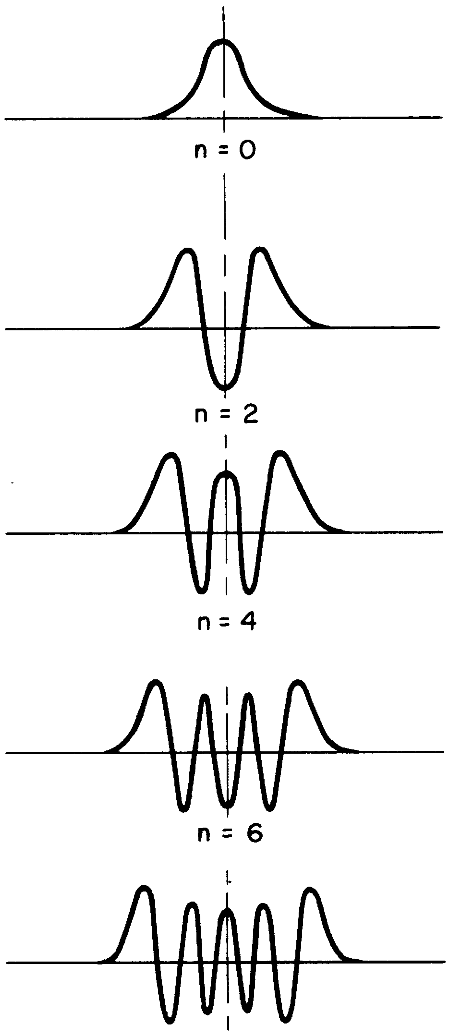

Fig. 14.
The first five hermitian orthogonal functions of even order.

These are also self-reciprocal, and have the property

$$\int_{-\infty}^{\infty} h_i h_k \, dx = \delta_{ik} \qquad \text{V.12}$$

where δ_{ik} is Kronecker's symbol; unity for $i = k$, otherwise zero.

If we now expand the energy spectrum (short-time or long-time) in terms of these functions in a form

$$S^2(f) = a_0 h_0(f/f_0) + a_2 h_2(f/f_0) + a_4 h_4(f/f_0) + \dots \qquad \text{V.13}$$

the corresponding expansion of the autocorrelation function will be

$$\Psi(f) = a_0 h_0(f_0 t) - a_2 h_2(f_0 t) + a_4 h_4(f_0 t) - \dots \qquad \text{V.14}$$

that is, it is a series with the same coefficients, but the sign of every second is reversed. We will call the sequence $a_0, a_2, a_4, \dots$; the Hermite spectrum. Only terms of even order are needed, since both the spectrum and the correlogram are even functions. Their shape is shown in Fig. 14.

Whether the Hermite spectrum is a useful tool or not remains to be seen, but a few conclusions can be drawn from existing experimental material. When subjected to the test of differentiation, the Hermite spectrum a_n (unlike the sequence of moments or of mean frequencies) reveals a simple property: Its zeros remain very nearly invariant. Hence we have some right to expect that the zeros, i.e. the sign changes of the sequence of the a_n will stand for characteristic data of the speech signal, which are essential for the intelligence.

It may be noted that the Hermite expansion also gives interesting results if it is applied to the signal itself instead of to its energy spectrum or its correlation function.

V.8. The Problem of Speech Recognition.

This difficult problem can be discussed only very briefly. Perhaps the most important result of research to date is the fact that deaf people and others can be trained to read visible speech. This, however, means only that sonograms, even of the simplified painted type, contain the information from which the eye plus the brain can carry out the identification which otherwise is carried out by the ear plus the brain. In both cases highly complicated processes are involved.

Identification of "universals" can be carried out in principle in two ways. One is to apply certain transformation groups to the single phenomenon and to check whether by these it can be reduced to a standard, with certain limits of tolerance.

This is the method which, according to W. S. McCulloch and W. Pitts of this laboratory, the visual cortex is using when it tries to decide whether a letter is an "A". The letter is translated to a standard position, rotated into a standard direction, and finally reduced to a standard size. If it covers the "key" or standard "A" of the mind, it will pass for an "A".

This also may well be the method, again according to McCulloch and Pitts, which the ear is using when it recognizes speech sounds. Just as an illustration let us think again of visible speech, but laid out on a logarithmic frequency scale, while the time axis remains linear. We want to recognize a speech sound, whether it is spoken by a man or a woman. In the second case the formants are about 10 percent higher. On the logarithmic scale this corresponds to a translation in the frequency direction. Let us imagine that we have a set of standard sonogram patterns. We can try these in turn, shifting them in the time direction and a little up and down, until one of them fits. We have intentionally used the vague term "speech sound". It may well be that our patterns will have to be two-letter combinations, syllables, and even words. But if we try long enough, and with a sufficient number of standards, we must succeed.

The second possible way of recognizing universals is through their invariants. Searching for an "A" we could ask a number of questions such as: "Is there a point in which two lines terminate in an acute angle? Are these two lines of about equal length? Is there a third line about halfway to their junction?" If we have an exhaustive set of questions for every "universal", and every one of them has been answered with a yes, then and only then have we made an identification.

We have been in possession of such a complete set of questions for the vowels for a long time (Helmholtz, Jones, Steinberg and others). A complete classification of all phonemes has been given quite recently by Roman Jacobson. One of the most important present problems of speech analysis is to translate Jacobson's criteria into mathematical and instrumental language.

In concluding these lectures, as at the beginning, it may again be emphasized that their purpose was not and could not be to present a complete picture of communication theory, which, though only a few years old, has already grown to impressive proportions. In the selection the guiding principle was to present with preference the methods of the British contributors, which are less well-known in this country, and which the author felt more competent to explain. Thus it could not be avoided that the methods and results of the American or "stochastic" school (Wiener, Shannon, Fano, and many others) have received much less than their fair share of space. The author had no wish to take owls to Athens and believes that he can leave the presentation of these theories to far more competent hands.

REFERENCES

1. R. V. L. Hartley: Bell System Tech. J. 7,535, 1928.
2. C. E. Shannon, W. Weaver: The Mathematical Theory of Communication, Bell. Labs. Series, Van Nostrand, New York 1950.
3. The concept of associative fields comes from A. Koestler: Insight and Outlook, Vol. I, London 1948; MacMillan, New York 1949.
4. The most extensive discussion of problems of this type may be found in J. Neumann, O. Morgenstern: Theory of Games and Economic Behavior, Princeton Univ. Press, New York 1944.
5. L. Szilárd: Z. Physik 53, 840, 1929.
6. D. M. MacKay: Phil. Mag. 41,189, 1950.
7. G. A. Campbell, R. M. Foster: Fourier Integrals for Practical Applications, Bell Labs. Series, Van Nostrand, New York 1949.
8. R. M. Fano: J. Acoust. Soc. Am. 22, 546, 1950.
9. R. M. Fano, P. E. A. Cowley: Quarterly Progress Reports,Research Laboratory of Electronics, M. I. T.
10. N. Stevens: J. Acoust. Soc. Am. 22, 677, 1950.
11. E. C. Titchmarsh: Introduction to the Theory of Fourier Integrals, Clarendon Press, Oxford 1937, p. 119.
12. D. Gabor: J. Inst. Elec. Engrs. (London) 93, 429, 1946.
13. J. Ville: Cables et Transmissions, 61-74, 1948.
14. H. Weyl: The Theory of Groups and Quantum Mechanics, p. 393, Methuen and Co.,London 1931.
15. R. Tolman: Principles of Statistical Mechanics, p. 235, Clarendon Press, Oxford 1938.
16. D. Gabor: New Possibilities in Speech Transmission, J. Inst. Elec. Engrs. 94,Part III, 369, 1947.
17. E. T. Whittaker: On the Functions which are Represented by the Expansions of the Interpolation-Theory, Dept. of Mathematics, University of Edinburgh Research Paper 8, 1915.
18. C. E. Shannon: Proc. I. R. E. 37, 10-21, 1949.
19. J. Oswald: Chem. Revs. 229, 21-22, 1949; Cables et Transmissions, pp. 197-215.
20. H. A. Wheeler, A. V. Loughgren: Proc. I. R. E. 26, 540, 1938.
21. B. van der Pol: Symposium on Information Theory, London, Sept. 1950.
22. D. Gabor: Phil. Mag. 41, 1161, 1950; Ed. Revue d'Optique, Paris 1951.
23. D. M. MacKay: Phil. Mag. 41, 189, 1950. (MacKay also suggests calling each degree of freedom in a signal or complex measurement a "logon", and that quantity of a physical parameter which induces us to move up one step in the proper scale a "metron".

24. C. E. Shannon: Proc. I. R. E. 37, 10, 1949.

25. W. G. Tuller: Proc. I. R. E. 37, 468, 1949.

26. This result is due to M. v. Laue: Ann. Physik. 44, 1197, 1914.

27. M. Planck: Ber. Preuss. Akad. Wiss. Berlin, 24, 442, 1924.

28. L. Brillouin: J. Appl. Phys. 22, 334 and 338.

29. R. K. Potter, G. A. Kopp, H. C. Green: Visible Speech, Van Nostrand, New York 1947.

30. H. Dudley: J. Acoust. Soc. Am. 11, 169, 1939.

31. R. J. Halsey, J. Swaffield: Proc. Inst. Elec. Engrs. (London) 96, Part III, 503, 1949.

32. J. R. C. Licklider: J. Acoust. Soc. Am. 22, 677, 1950.

33. E. C. Cherry: Imperial College of Science and Technology, London, unpublished work.