

0-21-47

Some Fundamental Considerations Concerning Noise Reduction and Range in Radar and Communication

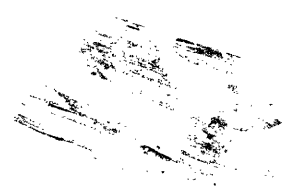
STANFORD GOLDMAN

TECHNICAL REPORT NO. 32

December 15, 1947

RESEARCH LABORATORY OF ELECTRONICS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY



The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research), and the Air Force (Air Materiel Command), under the Signal Corps Contract No. W-36-039 sc-32037.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Research Laboratory of Electronics

Technical Report No. 32

December 15, 1947

SOME FUNDAMENTAL CONSIDERATIONS CONCERNING NOISE REDUCTION
AND RANGE IN RADAR AND COMMUNICATION¹

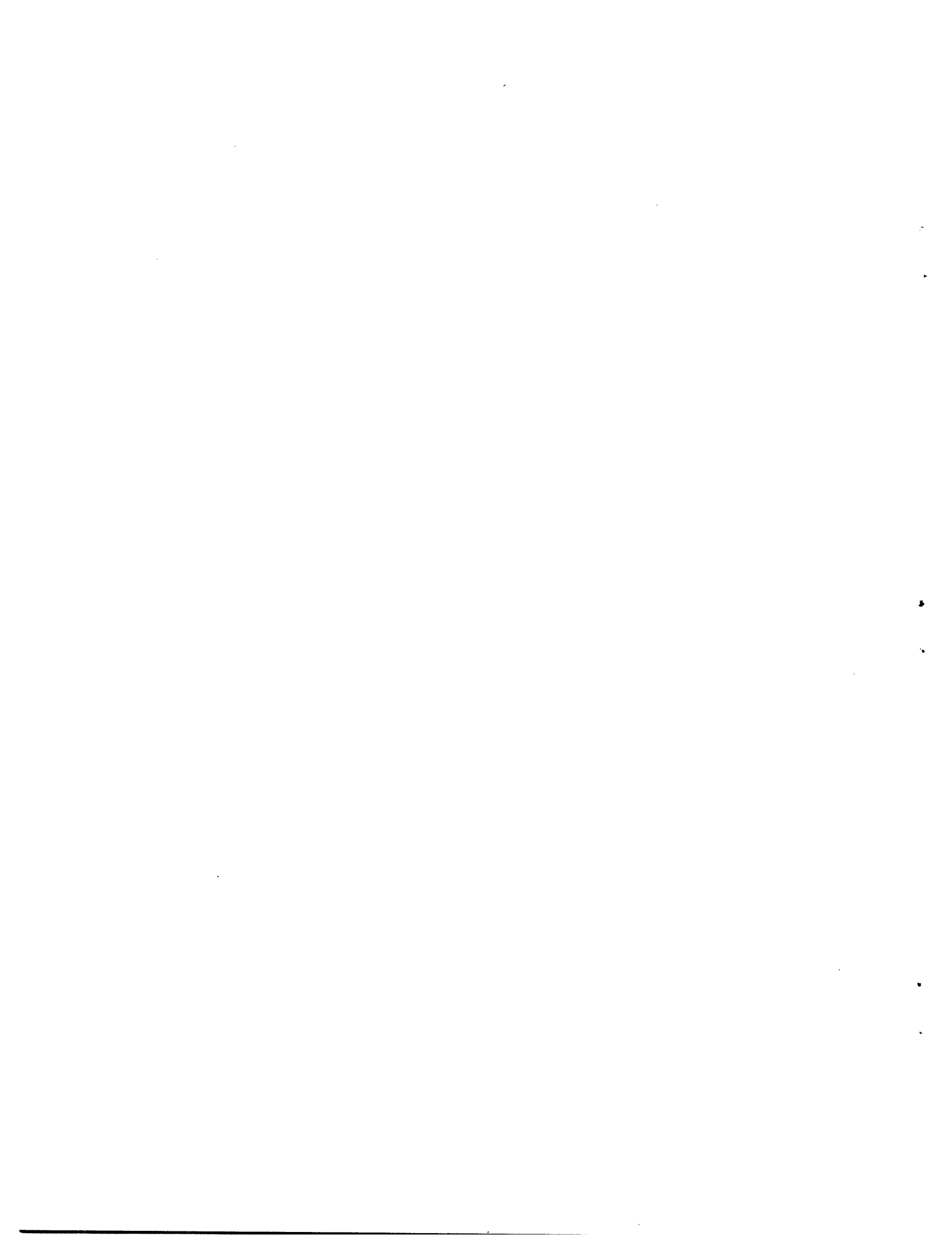
Stanford Goldman

Abstract

A general analysis based upon information theory and the mathematical theory of probability is used to investigate the fundamental principles involved in the transmission of signals through a background of random noise. Three general theorems governing the probability relations between signal and noise are proved, and one is applied to investigate the effect of pulse length and repetition rate on radar range. The concept of "generalized selectivity" is introduced and it is shown how and why extra bandwidth can be used for noise reduction. It is pointed out that most noise improvement systems are based upon coherent repetition of the message information either in time or in the frequency spectrum. It is also pointed out why more powerful noise improvement systems should be possible than have so far been made.

The general mechanism of noise improvement thresholds is discussed and it is shown how they depend upon the establishment of a coherence standard. The reason for, and the limitation of, the apparent law that the maximum operating range of a communications system, for a given average power, is independent of the type of modulation used, is then explained. General ways in which improvements in range of radar and communication systems may be made are also discussed. The possibility of using extra bandwidth to reduce distortion is then pointed out. Finally, some possible relations of this work to biology are described.

1. Presented at the National Electronics Conference, Chicago, Illinois, November, 1947.



Contents

1. Information Theory
 2. Modulation
 3. Some General Properties of Signals
 - 3.1. The Distinction Between Random Noise and Signals
 - 3.2. Generalized Selectivity
 - 3.3. Coherence
 4. Ultimate Noise Probabilities
 - 4.1. Probability Measure of Noise Level
 - 4.2. Some Useful Theorems
 5. Noise Improvement and Thresholds
 - 5.1. Improvement Based upon Coherence (by Means of Apparatus)
 - 5.2. Thresholds of Detection
 - 5.3. Perception Selectivity
 - 5.4. The General Question of Undesired Signals
 6. Range in Radar Systems
 7. Some Apparent Relations to Biology
- Appendix: Some Typical Examples of Noise-Reduction Calculations**
- A. Noise Reduction in Frequency Modulation
 - B. Noise Reduction in Pulse-Width Modulation (PWM)
 - C. Multichannel Operation with Coherent Audio

SOME FUNDAMENTAL CONSIDERATIONS CONCERNING NOISE REDUCTION
AND RANGE IN RADAR AND COMMUNICATION

1. Information Theory

The signals which are of interest in radio engineering may be represented graphically as functions of time. One such signal is shown in Figure 1. In a transmission system having L different significant amplitude levels, any particular signal such as that shown in Fig. 1 having a

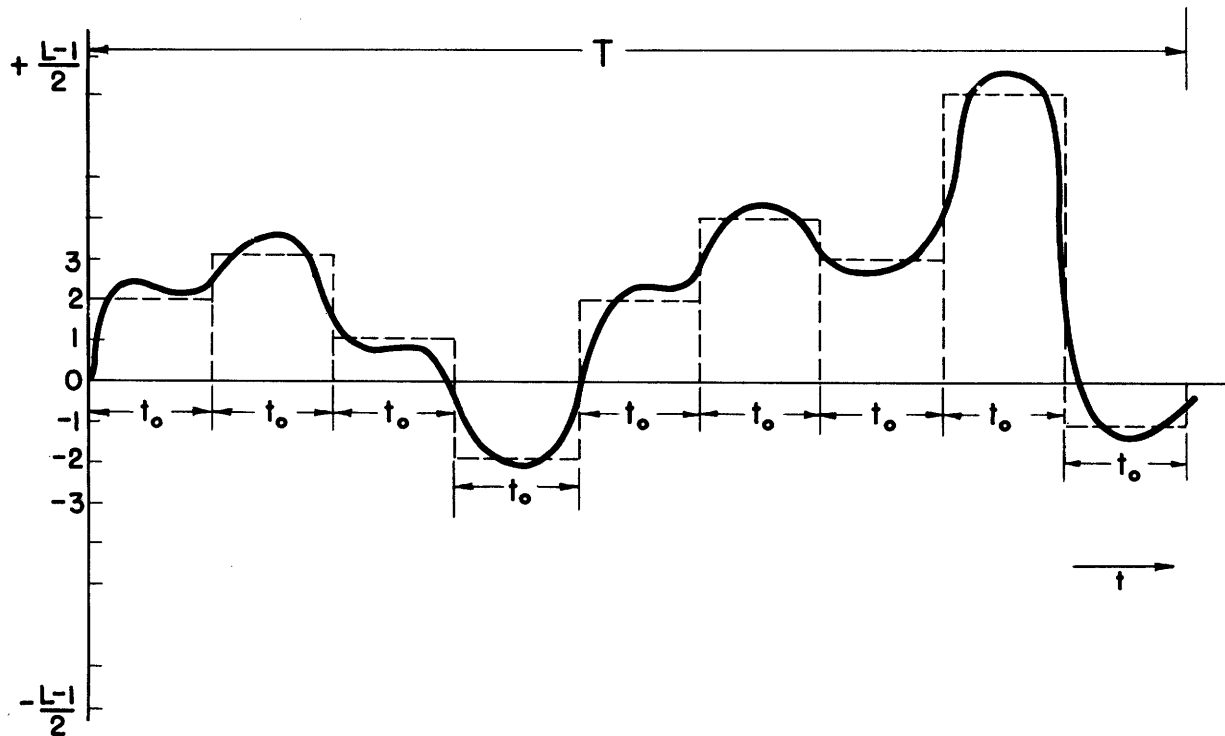


Figure 1. Diagram of a signal¹, showing its significant time intervals and amplitude levels. (Ideal signal shown by broken line. Solid line shows the same signal after passing through a transmission system of bandwidth B).

duration of n significant time intervals represents one out of L^n different possible signals of this duration which could have been transmitted by the system. With the foregoing meaning for the various symbols, we have as our

1. This signal is in a system in which there are both positive and negative levels. With noise also having both positive and negative levels, the spacing between signal levels must be the peak-to-peak value of noise namely $2N$, so that the number of different significant amplitude levels is still $L = S/N + 1$.

first equation

$$\text{Number of different possible messages} = L^n. \quad (1)^{1,2}$$

The number of significant amplitude levels is usually determined by the noise in the system. If the system is of a linear nature³, and the maximum signal amplitude is S, while the noise amplitude is N, then the number of significant amplitude levels is essentially

$$L = S/N + 1 \quad (2)$$

where the "1" in Eq. (2) is due to the fact that the zero signal level can be used.

The duration, t_o , of a significant time interval of the signal is determined by the inherent limited bandwidth of the signal. It is well known that if a signal has passed through a transmission system having more or less uniform transmission over a frequency bandwidth, B, then the smallest time intervals into which we can separate the portions of the signal such that amplitudes of the individual intervals shall be separately significant, will have a duration of approximately⁴

$$t_o = 1/2B. \quad (3)$$

Equation (3) may in any particular case be in error by several per cent. However, it will not be wrong by an order of magnitude. If the total duration of the signal is T, then the number of its significant time intervals is

-
1. This relation was derived by R. V. L. Hartley, B.S.T.J. VII, p. 535 (1928).
 2. For example, if there are three amplitude levels, designated as a, b, and c, and if there are two time intervals, then the $3^2 = 9$ possible signals are aa, ab, ac, ba, bb, bc, ca, cb, and cc.
 3. If the ultimate receiver is the eye or ear of a human being, the number of distinguishable signal levels is known to be proportional to the logarithm of the ratio of the maximum to minimum signal levels. This logarithmic ratio can also be expressed in decibels and is then called the dynamic range of the signal. The number of distinguishable amplitude levels is thus proportional to the number of decibels in the dynamic range of the signal, in case the ultimate receiver is a human being.
A system as just described, in which successive levels have logarithms which differ by constant values, has the special property that the relations between signal levels are unchanged if a signal is subjected to distortionless attenuation, (or distortionless amplification) before being received. Since all signals coming to the human eye or ear are subject to such attenuation by distance, and the human observer has no available amplification to make up for the attenuation, it is not surprising that human beings have logarithmic characteristics in their sensory receiving equipments.
 4. Stanford Goldman, "Frequency Analysis, Modulation and Noise", McGraw-Hill, New York, 1948, Chap. IV, especially Figure 7c.

$$n = T/t_0 = 2TB. \quad (4)$$

Consequently, a given message of duration T represents a particular choice of one out of

$$L^n = (S/N + 1)^{2TB} \quad (5)^1$$

different possible messages of the same duration which could have been sent through the system.

It is apparent that the amount of information in a signal increases monotonically with the amount of choice, L^n , available in choosing the signal. Hartley has given reasons for using

$$\log(L^n) = n \log L \quad (6)$$

as a measure of the quantity of information in a signal. It will, however, not be necessary for the purposes of this paper to choose a particular relation, such as Eq. (6), for the relation between quantity of information and (L^n) , and thus to assign a numerical value to information. We will, however, make use of the fact that, according to Eq. (5) and the monotonic relation between information and L^n , the quantity of information in a signal increases with T , B , and (S/N) . We will also use the fact that Eq. (5) shows the relative importance of these three factors.

Next, suppose that instead of plotting the signal as a function of time as shown as Fig. 1, we made a plot of its frequency composition, including magnitude and phase, and let us suppose that the information which it is desired to transmit is carried in this frequency plot. Using a similar method to that used in arriving at Eq. (3), it may be shown that the number of significant, independently specifiable quantities in the frequency composition of the signal is equal to the number of significant components in a Fourier series expansion of the signal using T as its fundamental period. Thus we have

$$\begin{aligned} \text{Signal} = f(t) = & a_0 + a_1 \cos 2\pi \frac{t}{T} + a_2 \cos 4\pi \frac{t}{T} + \dots \\ & + b_1 \sin 2\pi \frac{t}{T} + b_2 \sin 4\pi \frac{t}{T} + \dots \end{aligned} \quad (7)$$

Since the signal has no components of higher frequency than B , the coefficients in Eq. (7) become negligibly small after a_{TB} and b_{TB} . The signal, therefore has,

$$2TB + 1 = 2TB \text{ (approximately)} \quad (8)$$

1. Equation (5) has been derived independently by many people, among them Mr. W. G. Tuller, from whom the writer first learned about it.

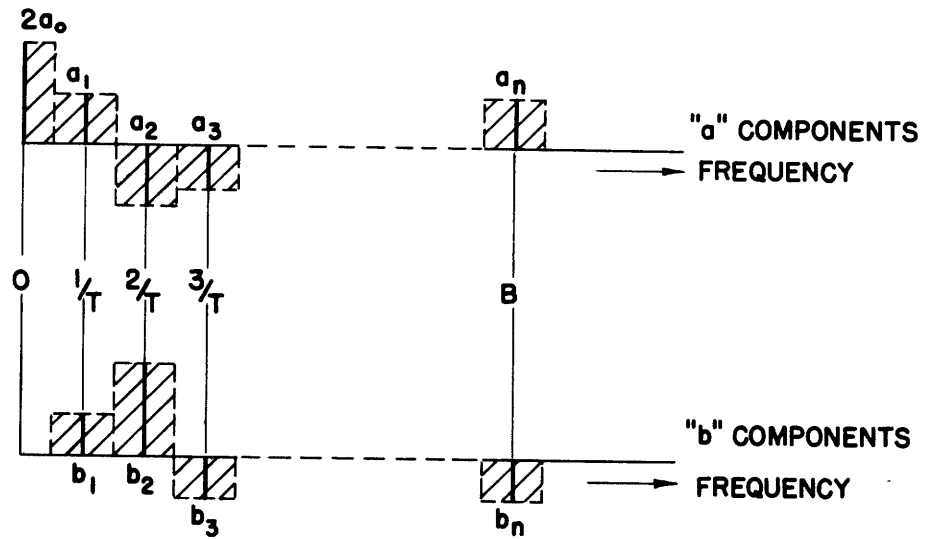


Figure 2. Frequency plot of a signal such as that of Figure 1.

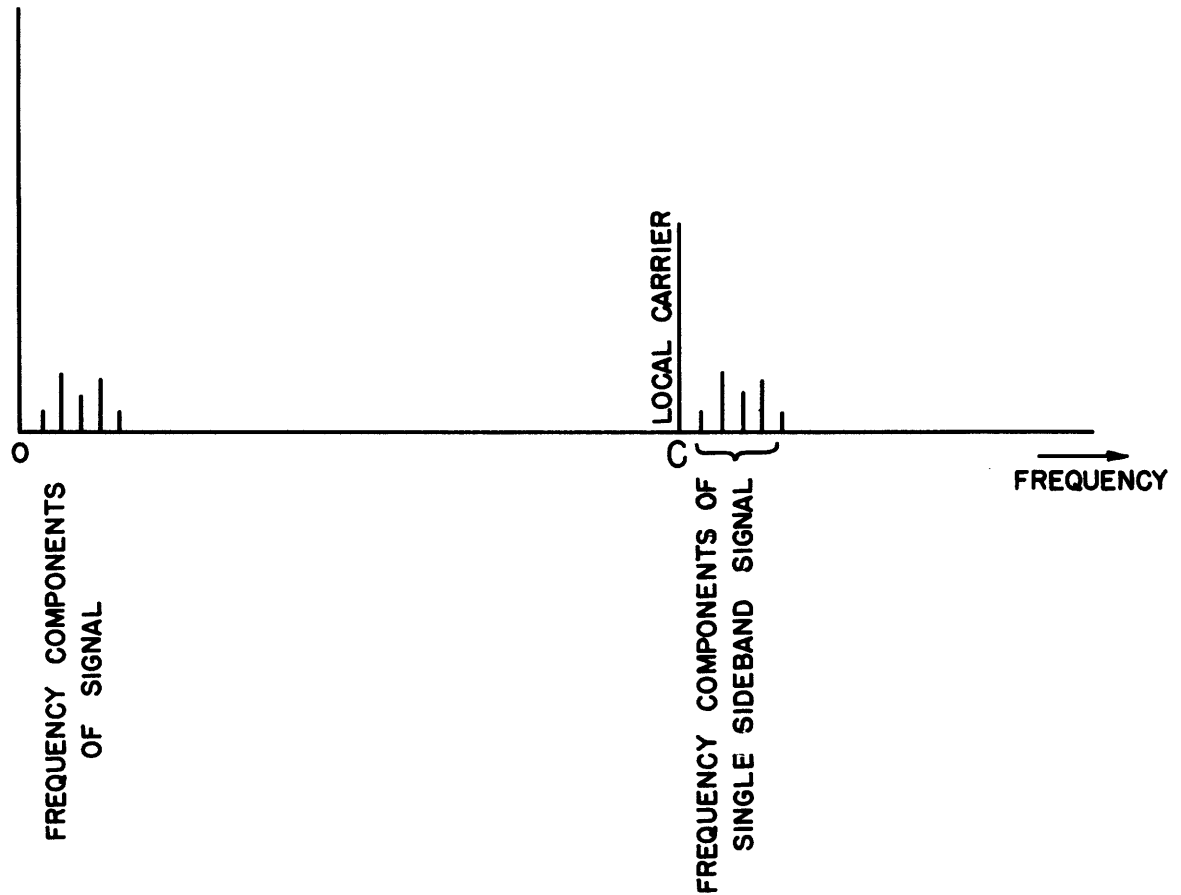


Figure 3. Single sideband transmission as displacement of the information components in the frequency spectrum. (This figure shows only one set, either the "a" or the "b" components. The other set would be displaced in the same manner.)

independently specifiable frequency components. Comparing Eq. (8) with Eq. (4) we see that the signal has the same number of independent information components, whether it is analyzed on a time or a frequency basis. A frequency plot of a signal is shown in Fig. 2. Instead of plotting a and b components, we could just as well plot magnitude and phase.

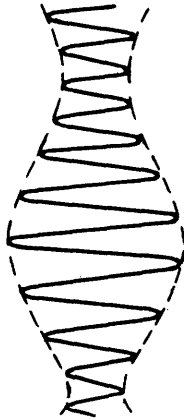
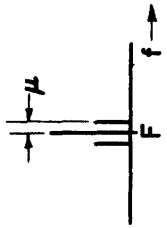
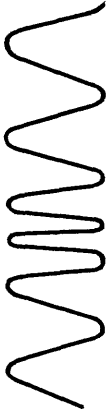
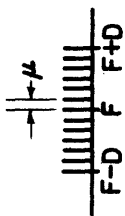
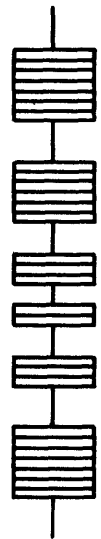
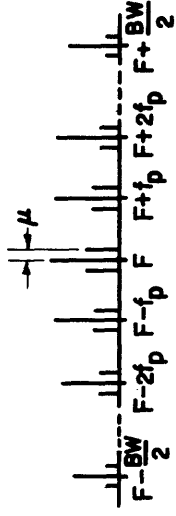
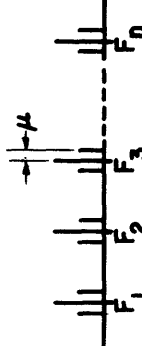
2. Modulation

For many purposes, it is desirable to send a signal in the form of a modulated carrier. The simplest method of doing this from an information theory point of view is to use single sideband transmission. In that case, the frequency components of the signal are all shifted by a constant amount, C, in the spectrum, as shown in Fig. 3. Then if a large carrier is added to the signal, the combination of signal plus carrier will be amplitude-modulated and its envelope will have the same shape as a function of time as did the original signal. Furthermore the number of significant time intervals and frequency components will still be given by Eqs. (4) and (8). (The combination of carrier plus sidebands will also be frequency-modulated, and if the proper frequency equalization is used, a frequency modulation detector will also detect the original signal.)

Although single sideband transmission with an added carrier, is the simplest type of modulation from an information theory point of view, there are many other types of modulation which are to be preferred in various applications. In Fig. 4, a group of these types is shown. In addition, Fig. 4(d) shows a method of transmission of a message as n separate amplitude-modulated carriers. The detected signals from the n channels are superimposed in phase in the final output.

Of special interest to us is the fact that the use of these various types of modulation, for a given average transmitted power, gives rise to different ratios of (signal/random noise) in the final output. The value of (S/N) for double sideband amplitude modulation divided by the (S/N) value for the type of transmission in question is called the noise improvement ratio of the latter¹. The noise improvement ratios of the various types of modulation are also shown in Fig. 4.

1. It would be more logical to use single sideband transmission as standard, but since double sideband AM has always been used in the past, we will retain it as the standard.

	TIME REPRESENTATION	FREQUENCY REPRESENTATION	NOISE IMPROVEMENT RATIO
<p>(a) PURE AMPLITUDE MODULATED SIGNAL. (DOUBLE SIDEBAND)</p>			<p>UNITY</p>
<p>(b) FREQUENCY MODULATED SIGNAL.</p>			<p>$\sim \frac{D}{\mu}$</p>
<p>(c) PULSE WIDTH MODULATED SIGNAL.</p>			<p>$\sim \sqrt{\frac{BW}{f_p}}$</p>
<p>(d) SEPARATE FREQUENCY CHANNELS WITH INDEPENDENT CARRIERS AMPLITUDE MODULATED BY THE SAME AUDIO PROGRAM.</p>			<p>UNITY</p>

$D \equiv$ peak frequency deviation in FM
 $f_p \equiv$ pulse repetition frequency in PWM
 $BW \equiv$ bandwidth of pulse-modulated signal

Figure 4.

$F \equiv$ carrier frequency
 $\mu \equiv$ modulation frequency

3. Some General Properties of Signals

3.1. The Distinction Between Random Noise and Signals. The outstanding difference between signals and random noise is that whereas there is no specified or predictable relationship between the amplitudes in the different specific time intervals of random noise, in the case of a signal there is some specified relation between the amplitudes in the time intervals. Correspondingly, in the frequency domain, whereas there is no specified or predictable relationship between the amplitudes of the Fourier a and b components in the case of random noise, in the case of a signal, there is a specified relation. In the case of random noise the amplitude of any time or frequency interval does not depend upon the values in any preceding or succeeding interval. The only thing which is specified is the probable average value taken over a large number of intervals, and even this is subject to random fluctuations. On the other hand, in the case of a signal, the relation between the time intervals is definitely specified. If the signal is simple, these relations may be specified by a simple equation such as

$$a = A(1 + n \sin \mu t) \cos \omega t \quad (9)$$

and the relation between the frequency components is thereby also definitely specified as being

$A \cos \omega t$ at frequency $\omega/2$,

$$\frac{An}{2} \sin [(\omega + \mu)t] \text{ at frequency } \frac{\omega + \mu}{2\pi},$$

$$\frac{An}{2} \sin [(\omega - \mu)t] \text{ at frequency } \frac{\omega - \mu}{2\pi},$$

and zero amplitude at all other frequencies. In case the signal is more complicated, the specification of the relations between the time intervals and between the frequency components is not so simple, but it is none the less definite.

3.2. Generalized Selectivity. The fact that a signal has a definitely specified relationship between its time components (and between its frequency components) allows a certain correspondence to be set up between signals and transmission systems. Thus it is well known that a tuned LRC circuit will transmit signals of the frequency $f = 1/(2\pi\sqrt{LC})$ with less attenuation than any other signal, whether the latter be a pure sinusoidal signal of a different frequency, or any other type of signal. As another example, an FM detector will give more output for signals of the form

$$a = A \cos \left[2\pi f t + \frac{D}{\mu} \sin 2\pi \mu t \right] \quad (10)$$

than for any other type of signal. However, whereas in the case of the tuned circuit the output signal is of the same form as the input, in the case of the FM detector, this is not true. In the case of the latter, the output just shows the modulation of the signal in Eq. (10) and not the whole signal.

As a general description of the degree to which a given transmission system will give a greater response to certain signals than to others, we will use the term generalized selectivity. Since random noise contains signals of all types in greater or lesser amounts, we will find it convenient to define the numerical value of the generalized selectivity of a transmission system as the ratio of the value of $\frac{(\text{Output})}{(\text{Input})}$ for the preferred type of signal to the value of $\frac{(\text{Output})}{(\text{Input})}$ for random noise. As already pointed out, the output need not be similar in wave shape to the input. Actually, we will use generalized selectivity as a descriptive term and will not need to calculate its numerical values.

Finally, we may point out that the more complicated the signal to which a given transmission system is fitted, the greater is the opportunity for a large value of the generalized selectivity. This is rather obvious, but we shall have occasions to demonstrate specific examples in which it is true.

3.3. Coherence. In Section 3.2 we described the relation between a signal and a transmission system in terms of generalized selectivity. We now wish to describe the relation between one signal and another signal, or between parts of the same signal. When two identical signals are superimposed in phase, it is well known that the energy of the resultant is equal to four times that of either signal taken separately. On the other hand, if they are superimposed 180° out of phase, the resultant has zero energy. In general if we have two signals $E_1(t)$ and $E_2(t)$, and the two are superimposed, the resultant energy is proportional to

$$\int [E_1(t) + E_2(t)]^2 dt = \int \{ [E_1(t)]^2 + [E_2(t)]^2 \} dt + \int 2E_1(t)E_2(t)dt. \quad (11)$$

The first integral on the right of Eq. (11) represents the energy of the two signals taken separately and the second integral represents the interaction energy. In the past, it has been customary to define two signals as incoherent if their interaction energy is zero. In accordance with this definition, two noise signals of independent origin are incoherent, and the different Fourier series components of the same signal (even the sine and cosine components of the same frequency) are incoherent. It is a general practical fact that any two signals of independent origin have a vanishingly small percentage of interaction energy when considered over a long period of time.

The definition of coherence on the basis of interaction energy, while it is quite useful for many purposes, is too specialized for our present

considerations. Thus the different sidebands¹ of a modulated wave, whether amplitude-modulated, frequency-modulated, or pulse-modulated, have definite phase relationships with respect to one another and with respect to the carrier, even though they are of different frequencies. We will therefore call them coherent, even though their interaction energy is zero. In general, we shall call any two signals (or any two parts of the same signal) coherent when there is a specified relationship between their detailed values. Incoherent signals will then be signals, or parts of the same signal, which are independent of one another. All incoherent signals will still have vanishingly small interaction energy, but the converse will not be true. Signals, whether having zero interaction energy or not, will still be called coherent if there is a specified relation between their detailed values (i.e., between the amplitudes of their time intervals).

4. Ultimate Noise Probabilities

4.1. Probability Measure of Noise Level. In any observation of a signal appearing above a level of random noise, there is always a certain probability that the signal is not a signal at all, but is just a fluctuation in the noise. When the signal-to-noise ratio by any definition is large, this probability will be very small indeed. However, when the signal and the noise are of comparable size, the probability becomes appreciable. In radar, when only one (or a few) significant time intervals are involved, the probability that what is believed to be signal may actually be noise is a matter of practical concern.

In the case of communication, when a large number of significant time intervals is involved, the probability that a received signal such as "The temperature at Dallas at 7 p.m. was 21 degrees" is merely a fluctuation of random noise and does not represent a transmitted signal is obviously almost infinitesimal. The reason for this is that the above signal includes a large number of significant time intervals. We know from probability theory, that the probability that each of the amplitudes in the independent time intervals should be a noise fluctuation is equal to the product of this probability for each interval taken separately. Since the probability for each interval is considerably less than unity, the final product is extremely small.

It would, however, not be necessary to change the sound amplitudes in many intervals of the above signal, to change "21 degrees" to "29 degrees", and if the noise level is high we might have some doubt about the accuracy of the reception of the above signal. This will serve to remind us that the correctness of the information in the above signal is a matter of probability, and that there is actually an extremely small probability that the entire signal may be merely a noise fluctuation and may not represent a transmitted

1. Arising from a particular modulation frequency.

signal at all. Despite the extremely small magnitude of this probability in almost all practical cases, we will shortly find that it will serve as a very useful measure of some of the noise properties of a signal. Accordingly we now propose the following universal measure of the noise level of a signal. Definition: The noise level of a signal is hereby defined as the probability that the observed signal does not represent transmitted information, but is just a fluctuation in the background of random noise.

With the aid of this definition we will now investigate some noise properties of signals.

4.2. Some Useful Theorems. Suppose that a signal is received for a time T by way of a transmission system of bandwidth B . Suppose also that the reception has a background of random noise of mean square value $\overline{I^2}$ (for the bandwidth B). According to random noise theory, the probability that the noise shall have the value I , in any particular significant time interval, is then, as a first-order effect,

$$P(I)\Delta I = \frac{1}{\sqrt{2\pi\overline{I^2}}} e^{-I^2/2\overline{I^2}} \Delta I \quad (12)$$

where ΔI is the increment between significant amplitude levels.¹

In Fig. 5(a) is shown the received signal plus background noise, of which the signal is shown in Fig. 5(b), and the noise in Fig. 5(c). The observer is of course unaware of the composition of Figs. 5(b) and (c). The amplitudes in the time intervals of Fig. 5(a) will be designated $A_1, A_2, A_3,$ etc., respectively. Now according to Eq. (12) the probability that the signal in the first interval, of amplitude A_1 , is random noise is

$$P(A_1) \Delta I = \frac{1}{\sqrt{2\pi\overline{I^2}}} e^{-A_1^2/2\overline{I^2}} \Delta I. \quad (13)$$

Similarly, the probability that the signal in the second interval is random noise is

1. In the following discussion, we shall assume that the distance ΔI between significant amplitude levels is the same for all amplitude levels. Very likely, this is an unnecessary limitation as far as Theorem B (Sec. 4 p. 11) is concerned.

The present section assumes that the superposition of signal and noise is a linear process. It therefore no longer applies after the mixed signal has gone through a nonlinear transmission system.

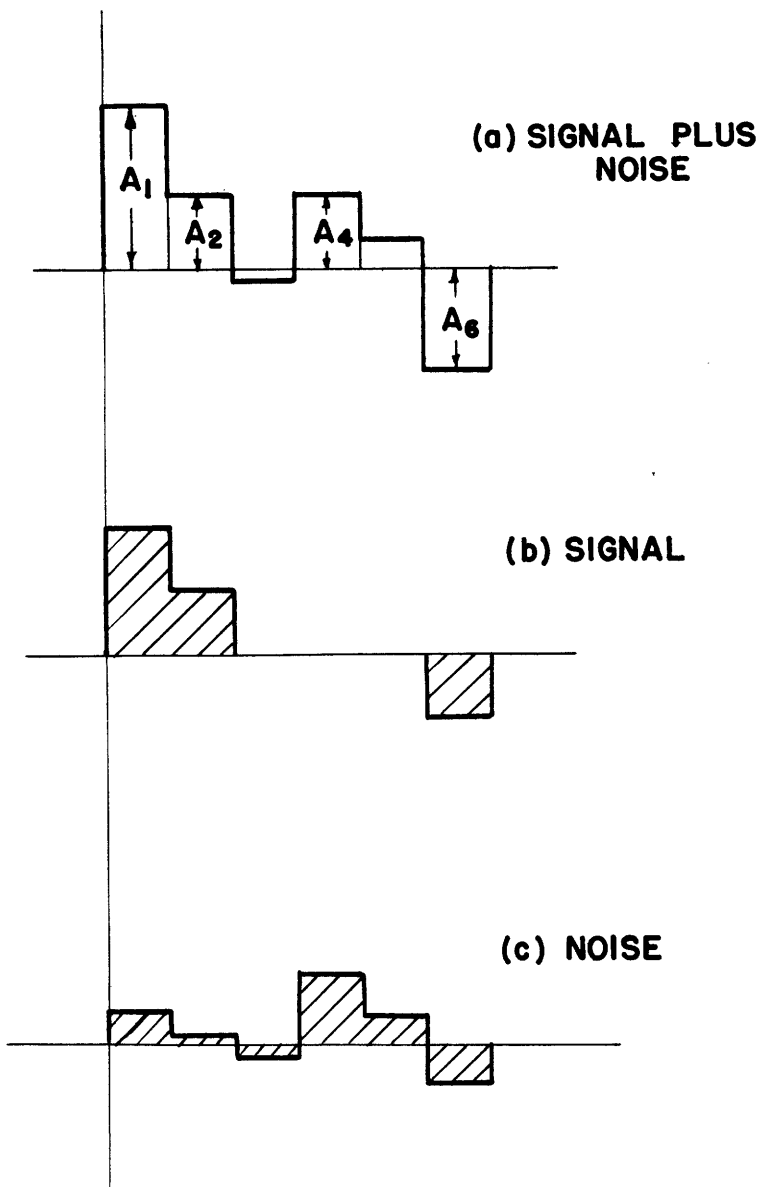


Figure 5. Superposition of signal and noise.

$$P(A_2)\Delta I = \frac{1}{\sqrt{2\pi I^2}} e^{-A_2^2/2I^2} \Delta I \quad (14)$$

and so on. According to a well-known law in probability theory, the probability that the entire signal is random noise is then

$$P(A_1)\Delta I P(A_2)\Delta I \cdots P(A_n)\Delta I = \frac{1}{\left\{ \frac{1}{2\pi I^2} \right\}^{n/2}} e^{-(A_1^2 + A_2^2 + \cdots + A_n^2)/2I^2} (\Delta I)^n \quad (15)$$

where $n = 2TB.$ (16)

It should be pointed out in passing that the probability given by Eq. (15) is not the probability that the amplitudes $A_1, A_2 \cdots A_n$ just occur, but rather that they occur in the sequence $A_1, A_2 \cdots A_n$. In other words, the signal shape as well as its magnitude is taken into account.

We have already pointed out that when a signal is superimposed upon random noise, the energy of the combination will be the sum of the energy of the signal plus the energy of the noise¹. Therefore, if we have a desired signal whose amplitudes in the n intervals are a_1, a_2, \dots, a_n respectively, and if this is superimposed upon noise of average value I^2 , the probability (to an observer who does not know the composition of what is being received but who only knows the average noise level I^2 of the transmission system and its bandwidth B) that the combination is merely a noise fluctuation is²

$$\frac{1}{\left\{ \frac{1}{2\pi I^2} \right\}^{n/2}} e^{-(a_1^2 + a_2^2 + \cdots + a_n^2 + nI^2)/2I^2} (\Delta I)^n. \quad (17)$$

We note immediately the remarkable fact that this probability depends only on the total energy ($a_1^2 + a_2^2 + \cdots + a_n^2$) of the signal and does not depend upon the way in which this energy is distributed among the different

-
1. Strictly speaking, what we mean is that if the number of intervals $2TB$ is large, the probability of an appreciable deviation from this addition of energies, is very small. (See Goldman, op. cit. Chap. VII). If $2TB$ is a small number, then in an individual trial there may be appreciable interaction energy, but averaged over a large number of trials, the average interaction energy will be zero, regardless of whether $2TB$ is large or small.
 2. More precisely "is" should read "will on the average be".

time intervals. We have thus arrived at the following important theorem:
Theorem A: In a system of given bandwidth, a signal of given duration and a given amount of total energy, in a background of white noise, will have the same probability of being a noise fluctuation regardless of how its energy is distributed in time.

A completely similar derivation based upon Fourier components instead of time intervals will give us the following alternative theorem.

Theorem A: In a system of given bandwidth, a signal of given duration and a given amount of total energy, in a background of white noise, will have the same probability of being a noise fluctuation regardless of how its energy is distributed among its Fourier components.

We can make Eq. (17) more universal by normalizing it. Thus

$$\overline{I^2} = M^2 B \quad (18)$$

where M^2 is a constant (called the strength of the background noise field) and

$$\frac{a_1^2 + a_2^2 + \dots + a_n^2}{2\overline{I^2}} = \frac{n\overline{a^2}}{2M^2B} = \frac{2TB\overline{a^2}}{2M^2B} = \frac{T\overline{a^2}}{M^2} \quad (19)$$

Consequently, we can rewrite Eq. (17) as

$$\begin{aligned} \left. \begin{array}{l} \text{(Probability that} \\ \text{(received mixed signal)} \\ \text{(is a noise fluctuation)} \end{array} \right\} &= \epsilon^{-\frac{(a_1^2 + a_2^2 + \dots + a_n^2)/2\overline{I^2}}{M^2}} \frac{\epsilon^{-n/2}}{\left\{ \frac{\overline{I^2}}{2\pi\overline{I^2}} \right\}^{n/2}} (\Delta I)^n \\ &= \epsilon^{-\frac{T\overline{a^2}}{M^2}} \frac{\epsilon^{-n/2}}{\left\{ \frac{\overline{I^2}}{2\pi\overline{I^2}} \right\}^{n/2}} (\Delta I)^n. \end{aligned} \quad (20)^1$$

$$\frac{\epsilon^{-n/2}}{\left\{ \frac{\overline{I^2}}{2\pi\overline{I^2}} \right\}^{n/2}} (\Delta I)^n = \frac{(\Delta I)^n}{\left\{ \frac{\overline{I^2}}{2\pi\epsilon\overline{I^2}} \right\}^{n/2}} = \left\{ \frac{(\Delta I)^2}{2\pi\epsilon\overline{I^2}} \right\}^{n/2} \quad (21)$$

1. The combination of desired signal plus random noise will be called the "mixed signal".

is the average probability that we would find that the mixed signal were noise if we made repeated trials and if the mixed signal were pure noise. Let us now define the normalized probability that a mixed signal is a noise fluctuation as the ratio of the actual numerical probability to the corresponding average probability for noise. Then Eq. (20) becomes

$$\left. \begin{array}{l} \text{(Normalized probability that mixed)} \\ \text{(signal is a noise fluctuation)} \end{array} \right\} = e^{-\overline{a^2 T} / M^2}. \quad (22)^1$$

Equation (22) is a fundamental equation. It shows that the normalized probability that the mixed signal is a noise fluctuation depends only upon the total signal energy $\overline{a^2 T}$ and upon the strength M^2 of the background noise field. It is independent of the time distribution of the signal energy as well as of the bandwidth B and duration T of the transmission.

Instead of deriving Eq. (22) by a consideration of the amplitudes in the time intervals we could by exactly similar mathematical steps dealing with the amplitudes of the Fourier series components, have derived the equation

$$\left. \begin{array}{l} \text{(Normalized probability that mixed)} \\ \text{(signal is a noise fluctuation)} \end{array} \right\} = e^{-\left[T \sum (a^2 + b^2) / 2 \right] / M^2} \quad (23)^2$$

where the summation is taken over all the Fourier series components of the signal. The quantity $T \sum (a^2 + b^2) / 2$ is, of course, just the expression for the signal energy in terms of the Fourier series components. The details in the derivation of Eq. (23) will not be reproduced here.

In addition to the conclusions drawn from Eq. (22), Eq. (23) shows us that the normalized probability that the mixed signal is a noise fluctuation is also independent of the frequency distribution of the signal. We can now include all these results in the following theorem:

Theorem B: The normalized probability that a mixed signal is a noise fluctuation, depends only upon the total energy of the desired signal and the strength of the background white noise field. It is independent of the time or frequency distribution of the desired signal energy, as well as of the bandwidth and duration of the transmission.

-
1. The strength M^2 of the white noise field, has the dimensions of energy. Therefore since $\overline{a^2 T}$ also has the dimensions of energy, the exponent of Eq. (22) is dimensionless, as it should be.
 2. \underline{a} in Eq. (23) stands for the amplitude of a cosine component in the Fourier expansion in accordance with the notation in Sec.1. It should not be confused with \underline{a} in Eq. (22), and the other equations of this section where \underline{a} stands for the amplitude of a time interval.

Theorem B is of fundamental importance. It tells us that no type of signal shape and no method of modulation can improve the normalized probability unless it increases the total signal energy or decreases the strength of the background white noise field¹. Although Theorem B is of universal applicability, it does not tell the whole story. The number of significant information components of the signal, $2TB$, does not depend upon either the amount of signal energy available or the background noise level. We shall show in a later section that when ample information is available, some of this information can be used to increase the signal-to-noise ratio and improve the noise probability despite the fact that the normalized probability is not improved. Even in radar reception, we will find that the actual range of operation is usually not pushed to the limit imposed by Theorem B, and certain types of noise reduction can be of practical value.

Finally, it is worth pointing out that Theorem B is a general mathematical theorem, i.e., a phenomenon of mathematics, and it may well have applications in the natural sciences in addition to its use in the field of communication engineering.

5. Noise Improvement and Thresholds

5.1. Improvement Based Upon Coherence - (by Means of Apparatus). In the present section we will deal with noise theory particularly as it relates to communication signals. We will be particularly concerned with the fundamental theory of noise improvement and the thresholds at which it begins.

The detailed wave shape of an audio signal carrying information cannot be predicted ahead of time, since it would not be carrying information if it could be predicted. There are, however, certain average characteristics of audio voice and audio music signals by means of which they can on the average be distinguished from random noise and on the basis of these characteristics a certain amount of average noise reduction can be obtained. Frequency pre-emphasis and de-emphasis as used in present day FM broadcasting may be cited as an example. The most potent types of noise reduction, however, such as that of wideband FM or of pulse modulation are based upon a method of generalized selectivity in which extra bandwidth is used outside the required audio range which carries the desired intelligence. This extra bandwidth is used to carry extra information which is used for purposes of generalized selectivity; that is, it makes a signal of the particular type of wave shape to which a prearranged detector is especially sen-

1. The strength of the background white noise field is determined in the early circuits of the receiver, and is unaffected by the noise reduction of frequency modulation or pulse modulation. However, a change of the carrier frequency, which allows the thermal noise and/or shot noise of the receiver (as usually measured by the noise figure) to be reduced will actually cause a practical reduction in the strength of the background noise field.

sitive. Due to this generalized selectivity, the signal-to-noise ratio in the output of the detector is much greater than in the input.

The extra information which is put into a signal for selectivity purposes is not really true information, since its characteristics must be known ahead of time. For this reason, we may call it "wave-shaping information" or "selectivity information" to distinguish it from what we may call the "message information". The "selectivity information" could theoretically be put into a signal by using either extra bandwidth or extra duration. However, in most cases of interest, such as audio or television broadcasting, the duration cannot be varied from the duration of the original "message information". It is therefore necessary to put the "selectivity information" into extra bandwidth. In a general way it is also clear that the more excess bandwidth there is available and used to carry "selectivity information", the greater will be the possible distinction between the signal and random noise, and consequently the greater will be the possible generalized selectivity.

The amount of noise improvement which is obtained by any particular method of "generalized selectivity" does not depend merely upon the amount of extra bandwidth used, since we found different answers for FM and PWM. In particular we found a noise-reduction factor proportional to D/F_a for FM and a noise-reduction factor proportional to $\sqrt{BW/f_p}$ for PWM. Now (D/F_a) and (BW/f_p) are in their respective cases each proportional to the number of complete sets of sidebands of the message modulation frequencies. It therefore appears that in these cases the number of times that the message information is repeated in the frequency spectrum is what determines the noise reduction. However, if we consider FM to operate on a linear basis, then PWM operates on a square-root law basis. In a similar manner, we would say that the case of multi-channel operation with coherent audio, which we found had no noise reduction for a given amount of total power, was a case in which the exponent was zero.

A general explanation of the way in which coherent repetition acts to reduce the noise level can also be based upon probability theory. If the original signal level gives a certain probability p that a particular information unit is a noise fluctuation, coherent repetition of the information q times is the equivalent of getting the same results q times out of q tries in probability theory. This reduces the probability that the signal is a noise fluctuation to p^q . A considerable correction must, however, be made in p , since the noise level rises with the extra bandwidth. Different types of transmission systems have different degrees of effectiveness in translating this decreased probability into noise reduction. We found, in the above-mentioned cases, that the noise reduction was proportional to q^m where m is a constant. The above p^q formula suggests, however, that the inherent possible noise reduction may increase exponentially with bandwidth, which is a

more rapid increase than is given by raising q to any constant exponent. The same possibility is suggested by Eq. (5). It therefore seems reasonable to believe that systems can be devised whose noise reduction greatly exceeds that of any of the systems which we have mentioned or which have yet been built.

The systems which we have so far described as noise-reducing systems gave noise reduction with a given amount of total energy. There are, however, practical cases in which it is not the total energy which is limited, but the energy per unit time or energy per unit bandwidth. In such cases, it is fair to describe a system which uses extra time or bandwidth for noise reduction as a noise-reduction system, even though the system uses extra energy. On this basis an integrating radar is a noise-reducing system which uses coherent repetition in time for noise-reduction purposes. On the same basis, the previously discussed case of multichannel coherent audio transmission would be a noise-reducing system, if extra power were available in the extra frequency channels. The noise reduction would be based upon coherent repetition in the frequency spectrum in the latter case¹.

5.2. Thresholds of Detection. All known noise improvement systems which are based upon "selectivity information" will operate effectively only above a certain threshold value of the S/N ratio. In the case of pulse phase modulation discussed above, the improvement threshold occurs when the signal pulses are twice as high as the peak noise fluctuations that are likely to occur during the reception of a message. At this level it is possible to use top and bottom limiters to remove all noise except that which occurs during the time of rise and fall of the pulses. If the signal falls below the improvement threshold level, the reception of noise between pulses causes a sudden great increase in the noise output, usually so great that it will blanket the desired signal.

In the case of FM, when the signal exceeds the noise entering the detector, the frequency modulation of the signal by the noise is relatively small. If, however, the noise exceeds the signal, then it is the noise which controls the phase of the signal-plus-noise combination, and the various sidebands of the desired signal are no longer coherent in phase with the effective carrier and can no longer operate effectively in unison to give large amounts of frequency modulation. The transition level between larger signal and noise is the improvement threshold in frequency modulation.

1. A standard double sideband system of pure AM uses twice the required bandwidth to carry its information. The reason for this is that in each significant time interval it gives the message information as AM, but it also gives the extra information that the FM in the interval is zero. This latter information results in an in-phase addition of the upper and lower frequency sidebands in their effect on the envelope, which has the effect of noise reduction. A similar situation apparently occurs in any type of double sideband transmission. Thus in FM the extra information is transmitted that the AM in each interval is zero.

In the case of double sideband amplitude modulation, there is also an improvement threshold. Above the threshold the upper and lower sets of sidebands of the desired signal due to their phase relation with respect to each other and with respect to the effective carrier, have double efficiency in causing amplitude modulation. Below this threshold, the noise is large enough so that the carrier is modulated more than 100 per cent most of the time by the noise sidebands. This eliminates most of the desired signal or transforms it into distortion.

To get an idea of what happens when the improvement threshold is reached, consider the case of pulse width modulation. Above the threshold value of 2 of the S/N amplitude ratio, only signal modulation or noise which changes the location of the time of rise or fall of the pulses is effective in producing output. From the point of view of probability theory, we can say that such signal modulation or noise is given preferred weighting above the improvement threshold. In a general way, we can say that below the improvement threshold, the receiving system cannot locate the signals which are to be given preferred weighting in order to obtain benefit from the "generalized selectivity".

The existence of a noise improvement threshold is apparently a characteristic of all types of modulation systems which use "selectivity information"¹. When the noise exceeds the improvement threshold value, the signal no longer controls the standard by which coherence is determined and on the basis of which the detector is designed to give preferred weighting to the desired signal. Under these circumstances, the detector increases the noise-to-signal ratio rather than the signal-to-noise ratio.

The maximum operating range of any communication system is determined by the location at which the signal falls below the improvement threshold. When this occurs, even in ordinary (double-sideband) amplitude modulation, there is a relatively sudden large rise of the noise level which effectively blankets the signal.² (See Fig. 6). This added noise is due to an irreversible process, and it cannot be removed by frequency selectivity. A realization of this fact is important in the design of communication systems. For example, in the design of an amplitude-modulated communication or radar receiver, if the signal-to-noise ratio is much higher than the improvement threshold, the bandwidth of the receiver prior to the second detector can

1. The only type of modulation which does not use "selectivity information" is single sideband transmission.
2. Below the improvement threshold, a major part of what comes into the receiver as desired signal is transformed by the detector into distortion or noise because of the loss of the coherence standard. This is the distinguishing characteristic of the improvement threshold.

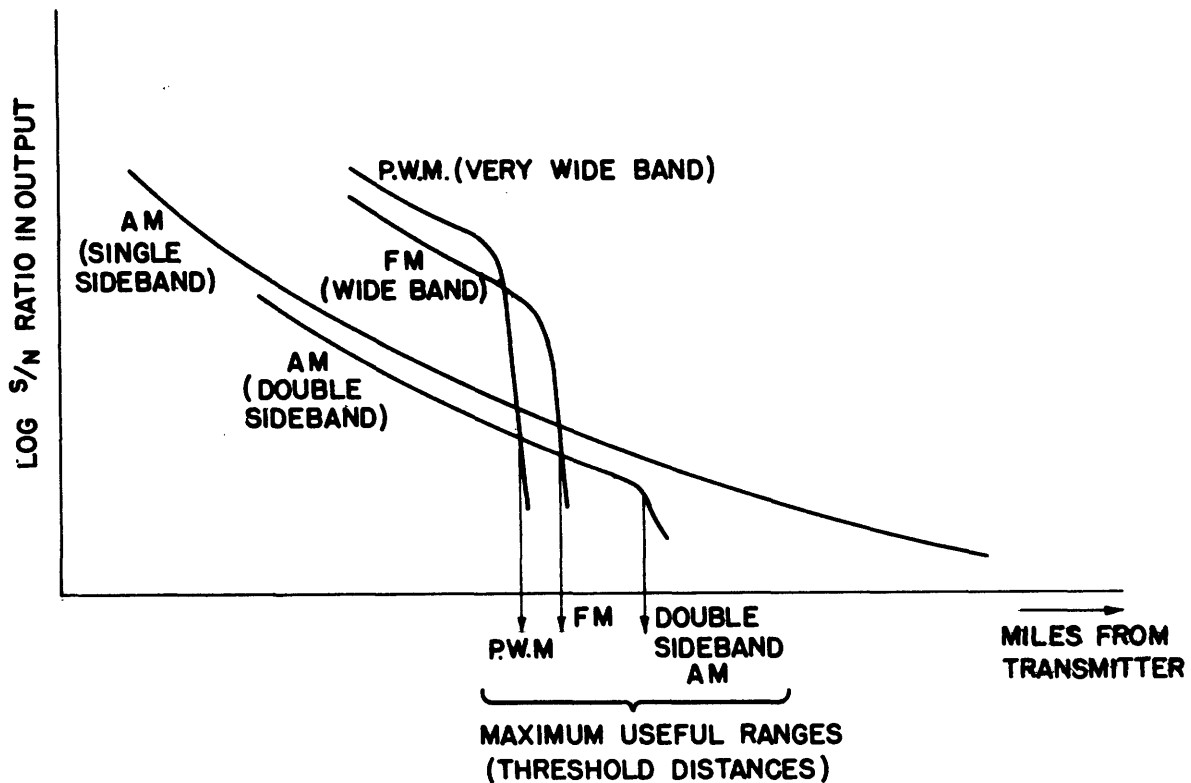


Figure 6. Threshold versus range for a given average power, with various types of modulation.

be increased many-fold in order not to lose the signal in case there is frequency drift of either the transmitter or the receiver oscillator. There will be no loss in signal-to-noise ratio in the final output, despite the increased predetection noise caused by the increased bandwidth, so long as the bandwidth is again narrowed to its optimum value by the audio (or video) amplifier. The situation, however, is quite different in case the signal-to-noise ratio is near the improvement threshold. In that case, widening of the predetection bandwidth to the extent that the signal-to-noise ratio falls into the range¹ of the improvement threshold will cause an irreversible rise in noise which cannot be erased by narrowing the bandwidth of the audio (or video) amplifier. For this reason, if maximum range is desired, the predetection bandwidth should not be increased beyond its normal value of twice the modulation frequency range except insofar as is absolutely necessary because of frequency drift.

The existence of the noise improvement threshold is the cause of the "law" that the maximum operating range of a communication system, for a given average power, is essentially independent of the type of modulation used. The

1. The improvement threshold has a narrow range of about 3 or 6 db, depending on the type of modulation. The coherence standard is gradually lost as the input signal-to-noise ratio falls to the bottom of this level.

explanation depends upon the fact that according to the 2TB formula for information components, the minimum possible bandwidth required, and therefore the minimum average noise, is fixed by the audio-frequency range¹ which it is desired to transmit. The minimum noise is thus independent of the type of modulation used. It is approximately true in the types of modulation which we have examined that when the average noise power exceeds the average signal power, the signal no longer controls the coherence standard used by the detector. When this occurs, the signal becomes unintelligible. The maximum range, since it is thus determined by the minimum noise, is consequently likewise approximately² independent of the type of modulation used. The above-mentioned maximum range "law" is thus a characteristic of all types of modulation systems in which the signal no longer controls the coherence standard of the detector when the average noise exceeds the average signal. If other type of modulation systems could be devised in which the coherence standard is not so controlled, there would be no reason to expect the "law" to hold for them.

The foregoing analysis indicates that if it were possible to supply a large local carrier in a double sideband AM receiver³ which is synchronized with the transmitted carrier, then the noise-reduction threshold of the receiver would be decreased and the operating range of the communication system would be extended, correspondingly. Radar synchronization is a closely related type of procedure, but for full utilization of the noise-reducing possibilities, local carrier would have to be added in the receiver with synchronization of the r-f phase.

5.3. Perception Selectivity. In the practical reception of signals, whether audio, television, or radar, the final human observer usually adds a large and important amount of effective noise reduction. A human observer will weigh intelligible speech much higher than "gibberish" as being parts of the signal, and the human observer will even fill in the gaps where noise makes the signal unintelligible. As soon as the signal can be recognized as belonging to a customary type of communication signal, such as speech or music, the human perception mechanism gives greatly preferred weighting, on the average, to true signals as compared with noise and greatly decreases the previous probability that the signal is a noise fluctuation. This human "perception selectivity" is thus another example of generalized selectivity.

1. ~~Systems, such as the Voder, which reduce the required bandwidth of transmitting speech, will, of course, increase range. This, however, will be true regardless of the type of modulation used.~~
2. We are here neglecting small constant factors in the neighborhood of unity, such as 2 or 1/2, etc.
3. This might be obtained with the aid of a very narrow-band high-gain receiver tuned to receive the carrier alone.

Perception selectivity, like other noise-reducing systems has a threshold below which it will not operate. The threshold at which perception selectivity begins may roughly be described as the S/N level at which the signal can be recognized as belonging to the transmission "language". At this level the human perception mechanism recognizes parts of the mixed signal as coming from a common origin; i.e., it recognizes them as parts of a coherent signal. This level may be described as the intelligibility threshold, and is the noise improvement threshold of a human being as a signal detector. It is probable that perception selectivity begins at a lower S/N ratio than any non-living type of noise improvement system so far devised.

Perception selectivity for speech does not require any extra bandwidth in addition to that normally used, because the normal audio range of say 10 kc/sec is many times more than adequate to carry the actual amount of information in a speech signal. The actual message information in speech is generated at a rate of about 5 to 25 significant time intervals per second. The audio bandwidth is thus of the order of one thousand times as much as necessary for the purposes of carrying the pure "message information" in speech. The excess bandwidth is thus available for the generation of characteristic human speech wave shapes which can be modulated by the "message information". The human detector then gives preferred weighting to these characteristic human speech wave shapes¹.

We have found previously in our study of nonhuman noise improvement systems, that coherent repetition of the information in many different frequency channels is a common characteristic of the signals used. At least part of the noise reduction of audio perception selectivity is obviously due to the use of signals with this same characteristic.

Perception selectivity is also used in radar. For example, in the use of a radar "A" presentation, the fact that a slightly greater signal amplitude is apparently characteristic of a particular location along the horizontal axis in many repeated intervals, is interpreted by the human observer as indicating an echo. It is interesting to note that repetition of the information in the time domain, rather than in the frequency domain, is here used for noise reduction.

The use of repetition in the time domain for reducing the probability that a signal is a random fluctuation is a common practice in perception selectivity. Repeating a message in order to be sure that all the information is correctly received is an example of applying this to every information interval in a signal, so as to reduce the probability that any part of the received signal is a random noise fluctuation.

1. The writer does not mean to imply that the wide audio range of human speech has noise reduction as its sole purpose. Obviously, the radiation of sound from the mouth, as just one example, could not take place efficiently at a syllabic frequency, but requires the use of higher audio frequencies.

Perception selectivity is an important part of almost every communication system in which a human observer is the final receiver.

5.4. The General Question of Undesired Signals. We can classify all undesired signals in the following three groups:

- a. Random noise - which is incoherent information.
- b. Signal interference - which is erroneous information, coherent in itself, but not coherent with the desired signal.
- c. Distortion - which is erroneous information, coherent with the desired signal.

Extra bandwidth can be used to reduce (a) and (b) relative to the signal. Frequency modulation systems are perhaps the best-known example of this fact. That extra bandwidth can be used to reduce distortion is not well known; but a comparison of double sideband versus single sideband AM transmission shows that it can be done. The use of large amounts of extra bandwidth to reduce distortion to a very low value presents interesting possibilities. Just as in the case of the reduction of random noise and signal interference, the reduction of distortion in the simplest cases would probably be accomplished by coherent repetition of the desired signal information in several different transmission channels; with the use of a type of modulation and detection which would give coherent addition of the desired signal but not of the distortion.

In this connection, it is interesting to note that feedback is a type of coherent repetition. It has certain drawbacks, especially instability. Perhaps some other type of coherent repetition will do the same job as feedback in some cases, without the danger of instability.

6. Range in Radar Systems

In analyzing the maximum range problem in radar we must distinguish between two cases. In the first case, the presence of a reply echo is determined both by observing excess signal at the location of the echo and by recognizing a characteristic shape of the reply pulse. This case requires bandwidth in excess of that necessary to give the best signal-to-noise ratio; the excess bandwidth being required to show the pulse shape. In the second case, the bandwidth is reduced to obtain the best signal-to-noise ratio, and the presence of the echo is recognized purely on the basis of excess signal. The first case involves perception selectivity. This case is not amenable to accurate quantitative calculation. The second case, however, can be handled in a quantitative manner, and will now be discussed.

In the operation of a radar system, the carrier is off except during short pulse lengths. These pulses are repeated at regular intervals, and the echoes after demodulation are displayed on a cathode-ray tube screen, or an equivalent circuit device. The cathode-ray tube screen is scanned

in synchronism with the pulse-repetition rate, so that successive echoes will be superimposed as long as the target does not move. If the pulse length is L and the period of repetition is T_o , then for a given average transmitted power P , the pulse signal length S , prior to detection, will follow the relation

$$S \propto \sqrt{\frac{PT_o}{L}} \quad (24)$$

Now the bandwidth B required to transmit the pulses is inversely proportional to L . Furthermore, the amplitude N of the background noise prior to detection is proportional to the square root of the bandwidth. Therefore, we can write

$$\frac{S}{N} \propto \frac{\sqrt{\frac{PT_o}{L}}}{\sqrt{B}} \propto \sqrt{PT_o} \quad (25)$$

The signal-to-noise ratio is thus independent of the pulse length or the corresponding bandwidth, for a given repetition rate and average power.

The ultimate operating range of a radar is determined by the distance at which reply pulses become so small that they cannot be distinguished from the background noise. In order to push the operating range of radar to its maximum, a special integrating device may be used. This consists of a time-gating device which is sensitive only for the duration of a pulse length, but whose intervals of sensitivity are repeated in synchronism with the radar's pulse-repetition rate. This is equivalent to moving a gate, of the same width as a reply pulse, along the scanning direction of the cathode-ray tube. If this gating device is connected to an rms meter of long time constant, it will be very sensitive in locating reply pulses as it is slowly moved along the axis. There will, of course, still be noise fluctuations, and it is necessary to try to distinguish true echo signals from noise fluctuations. According to probability theory, the ratio of the average fluctuation $\overline{\Delta N}$ of the rms noise from the long time average rms noise value \bar{N} will decrease as the square root of the number of scans included in the observation. Thus

$$\frac{\overline{\Delta N}}{\bar{N}} \propto \sqrt{\frac{T_o}{T}} \quad (26)$$

where T is the total time of observation.

The noise in Eq. (26) is post-detector noise whereas that in Eqs. (24) and (25) was pre-detector noise. At ultimate range the signal-plus-noise locations are only barely perceptible above the background of noise. Therefore S is small in comparison with N as they enter the detector. Because of the random r-f phase relations of the signal and noise,

the signal then causes only a second-order change in the amplitude of the envelope of the noise output, as we will show. Thus

$$N \cos \omega t + S \cos(\omega t + \phi) = \sqrt{(N + S \cos \phi)^2 + (S \sin \phi)^2} \cos(\omega t + \psi) \quad (27)$$

Here

$$\sqrt{(N + S \cos \phi)^2 + (S \sin \phi)^2}$$

is the amplitude of the envelope, which will become the detected signal, and is therefore the quantity of prime interest. The angle ϕ is a random r-f phase angle, and so is ψ . Now

$$\sqrt{(N + S \cos \phi)^2 + (S \sin \phi)^2} = [N + S \cos \phi] \left[1 + \frac{1}{2} \left(\frac{S \sin \phi}{N + S \cos \phi} \right)^2 \right] \text{ approximately}$$

provided that $\frac{S}{N} \ll 1$.

(28)

The average output envelope is then

$$\frac{1}{2\pi} \int_0^{2\pi} [N + S \cos \phi] \left[1 + \frac{1}{2} \left(\frac{S \sin \phi}{N + S \cos \phi} \right)^2 \right] d\phi = N + \left(\frac{S}{4N} \right) S, \text{ approximately.} \quad (29)$$

The effective value of the signal has thus been reduced by the factor $(S/4N)$.

The average output signal-to-noise ratio at echo locations will then be

$$\frac{\bar{S}}{\bar{N}} = \frac{\left(\frac{S}{4N} \right) S}{N} = \frac{1}{4} \frac{S^2}{N^2} = \frac{1}{4} P T_0 \quad (30)$$

The ratio $\bar{S}/\Delta \bar{N}$ determines the probability that the signal is a noise fluctuation; and from Eqs. (26) and (30) it follows that

$$\frac{\bar{S}}{\Delta \bar{N}} = \frac{\bar{S}/\bar{N}}{\Delta \bar{N}/\bar{N}} \propto \sqrt{P \tau} \cdot \sqrt{P T_0} \quad (31)$$

The ratio $\bar{S}/\Delta \bar{N}$ thus depends on $P \tau$, the total signal energy received, and upon $P T_0$, the energy per pulse. The range of an integrating radar for a given average power, can thus be improved by increasing the time of observation τ , and by increasing the pulse repetition period, T_0 . A PPI is also an integrating type of radar and the foregoing discussion will apply to it. In non-integrating types of radar, Eq. (31) will still apply, if τ is set equal to T_0 . This case is, however, not very important since in all radars in which there is visual observation, there is always considerable effective integration by the observer in the perception-selectivity process, even when there is no integration by the receiving apparatus.

In the case of the radar systems analyzed above, the predetection coherence standard of the double sideband AM radar signal has been lost at ultimate range; so that the noise is eliminating the signal. However the post detection coherence standard (synchronization) has been maintained by the pulses from the local transmitter so that coherent repetition in

time, utilized by integration, can be used to improve $(\bar{S}/\Delta N)$. The extreme simplicity of a radar signal is what has made it possible for us to carry out the above quantitative analysis of the detection process below the threshold of detection.

In many practical radar systems, the method of display is such that ultimate range is reached when the signal is of about the same magnitude as the noise. In that case the detection correction for small signals no longer applies, and

$$\frac{\bar{S}}{N} = \frac{S}{N} \text{ (approximately)} \quad (32)$$

and

$$\frac{\bar{S}}{\Delta N} = \frac{\bar{S}N}{\Delta N/N} \propto \sqrt{P\tau} \quad (33)$$

In this case, the range depends only on the total received energy.

Radar has the simplest type of information of any signal, since we are interested only in whether a signal is present or not. The situation in communication is more complicated, since we are there concerned with the probable percentage of information intervals which have correct information. In the simplest communication system, we will have only two levels, off and on. However, in most cases, such as voice or television communication, we have many amplitude levels. Thus we have the possibility that an interval which has signal may give the wrong information because a noise fluctuation has moved it to the wrong amplitude level.

The type of analysis used above for radar is inadequate to handle such problems. In the usual practical case, a small percentage of intervals carrying erroneous information can initially be tolerated. This erroneous information is usually then removed by perception selectivity. Although the matter is one of great importance, no attempt will be made in the present paper to handle the question of probable percentages of correct intervals in a communication signal.

7. Some Apparent Relations to Biology

The material discussed in the present paper, if properly developed, may have a fascinating field of application in the other sciences. Every indication points to the fact that the nervous system is an intricate communication network, and a good deal of the perception selectivity we have talked about probably takes place in a manner similar to the coherent repetition used in simple communication systems. The well-known phenomenon of a conditioned reflex is also closely related to coherent repetition. The

proper interpretation and analogies of bandwidth, random noise level, information intervals, and thresholds in these physiological domains may lead to new and worthwhile ideas in devising and interpreting new experiments.

A particularly interesting biological application of the ideas in this paper is in the explanation of a reason for the division of large organisms into cells¹. According to current biological theory, the characteristics of an organism are determined by the chromosomes in its cells, and there are essentially identical sets of chromosomes in each of the cells of the organism (excepting the germ cells). Now the life history of the organism is regulated by the interactions of the various parts of the organism which is living in an environment subject to a considerable amount of random fluctuation. These interactions, whether they are secretions by glands, the sending of nerve impulses, the process of digestion, or any other type of physiological action, represent perfectly good signals in the general theory of communication. The activity of the chromosomes, whatever it may be, may likewise be considered a set of signals.

In the case of a large organism, there is a large amount of structure built according to a detailed plan which must be kept running according to plan during the life cycle in spite of the general disintegrating processes of the physical world and the random fluctuations of the environment. The large amount of characteristic detail in the organism is the communication equivalent of a large amount of signal detail. The communication theory equivalent of retaining all this characteristic individual detail despite the presence of the fluctuating environment is the maintenance of a large signal-to-noise ratio. It is therefore not surprising to find that in the case of large organisms, the fundamental information contained in the structure or activity of the chromosomes which regulate the life history of the organism is set up in such a way that the information is repeated coherently, i.e., according to a planned lay-out. Thus the division of the organism into cells, each of which has an essentially identical set of chromosomes, may be a way in which nature can keep a large organism operating in a planned life cycle despite the disintegrating effects of the physical world.

1. There are, of course, also other reasons for the division of a large organism into cells besides the one suggested in these paragraphs.

Appendix

SOME TYPICAL EXAMPLES OF NOISE-REDUCTION CALCULATIONS.

A. Noise Reduction in Frequency Modulation. An elementary analysis¹ shows that a small component $B \cos 2\pi g t$ received simultaneously with the carrier $A \cos 2\pi f t$ causes amplitude modulation of the carrier at the beat frequency $(g-f)$ and of degree of modulation, B/A . It also causes frequency modulation of the carrier at the beat frequency $(g-f)$ but of degree of modulation $[(g-f)/D]B/A$, where D is the maximum permitted frequency deviation of the carrier, corresponding to 100 per cent frequency modulation. Thus all the noise components are equally effective in causing amplitude modulation of the carrier, but their effectiveness in causing frequency modulation is proportional to their frequency displacement from the carrier.

In Fig. 7 we have a diagram showing the relative efficiencies of noise components in causing amplitude and frequency modulation in a receiver. This figure also illustrates the fact that since the noise-modulation frequency is $g-f$, any noise component is ineffective if its frequency differs from the carrier by more than the audio cutoff frequency of the receiver. It is clear from Fig. 7 that the noise reduction of FM as compared with AM is

$$\text{Noise-Reduction Factor} = \frac{\int_{-F_a}^{+F_a} M^2 df}{\int_{-F_a}^{+F_a} \left(\frac{Mf}{D}\right)^2 df} = \frac{\sqrt{2M^2 F_a}}{\sqrt{\frac{2M^2}{D^2} \frac{F_a^3}{3}}} = \sqrt{3} \frac{D}{F_a} \quad (1A)$$

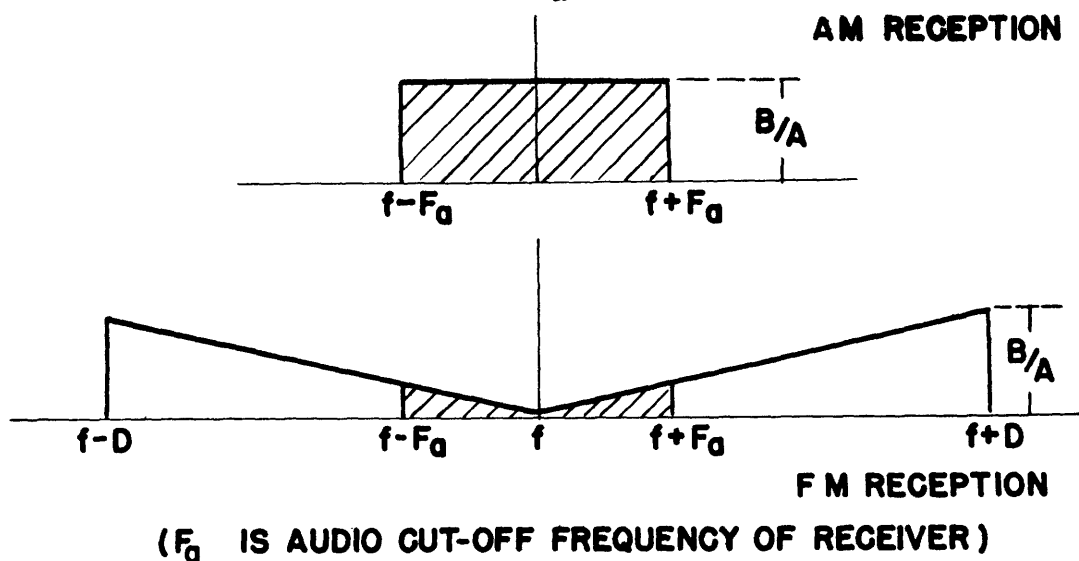


Figure 7. Effectiveness of noise components in causing modulation.

1. Goldman, op. cit., Sec. 5.5.

In order to get a clearer understanding of the reason for this noise reduction, let us consider a simple 100 per cent modulated f-m signal

$$a = A \cos \left[2\pi f t + \frac{D}{\mu} \sin 2\pi \mu t \right]. \quad (2A)$$

This is modulated at the modulation frequency μ , and the instantaneous frequency at any time t is

$$\frac{1}{2\pi} \frac{d}{dt} \left[2\pi f t + \frac{D}{\mu} \sin 2\pi \mu t \right] = f + D \cos 2\pi \mu t. \quad (3A)$$

The signal in Eq. (2A) is well known¹ to have a distribution of sidebands extending essentially from $f - D$ to $f + D$ and spaced from each other by μ .

The amplitudes, phases, and frequencies of the f-m signal sidebands, with respect to each other and with respect to the carrier, are such as to be very effective in causing a large deviation D of the carrier at the modulation frequency μ . If these sidebands had their amplitudes, phases, and frequencies distributed at random, as is the case with noise, the frequency deviation of the carrier would be greatly reduced. Thus the coherence of the f-m signal sidebands is one reason for noise reduction in frequency modulation.

Next we note from Fig. 7 that only those noise sidebands which are near the carrier frequency, and are therefore inefficient in producing modulation, generally cause noise that will pass through the audio system of an f-m receiver. This is in decided contrast with the signal, which has effective sidebands far removed from the carrier and all of whose sidebands act together to swing the carrier frequency at the rate of the signal modulation frequency.

The noise reduction properties of an f-m system require that the signal level shall exceed the noise level. When the noise exceeds the signal, the situation becomes reversed, and the noise tends to wipe out the signal.²

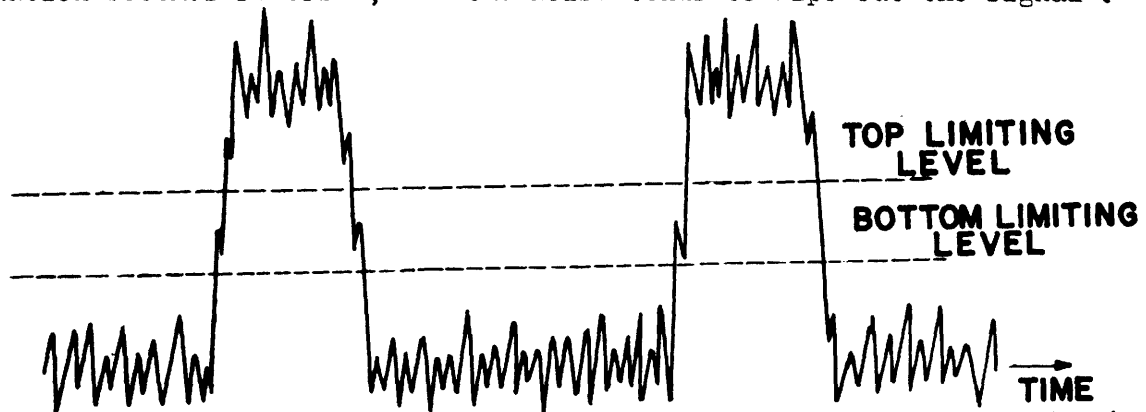


Figure 8. A pulse-modulated signal received in the presence of noise.

1. See for example, Goldman, op. cit., Chap. V.
2. Goldman, op. cit., Sec. 6.11d.

B. Noise Reduction in Pulse-Width Modulation (PWM). A second type of noise-reduction system is that using pulse-width modulation. If the width of the pulse is w , then for a single modulation frequency

$$w = w_0(1 + n \sin 2\pi\mu t) \quad (4A)$$

where n is the degree of modulation and μ is the modulation frequency. In Fig. 8 is shown a pulse-width-modulated signal received in the presence of some random noise and the elimination of noise between pulses by top and bottom limiting is indicated.

Since the bandwidth of the receiver is not infinite, the angle of rise and fall of the pulse is not 90° , but is some lesser angle θ , where

$$\tan \theta = \frac{\text{pulse height}}{\text{time of rise}} = 2bS = BS \quad (5A)$$

where S is the signal height before limiting, b is the bandwidth of the pulse amplifier, and B is the bandwidth of the (double-sideband) carrier-frequency amplifier. Now the presence of noise during the rise and fall of the pulse affects the pulse width. In Fig. 9 is shown a single pulse from Fig. 8. The dotted sloping lines represent the original signal pulse, and the solid lines

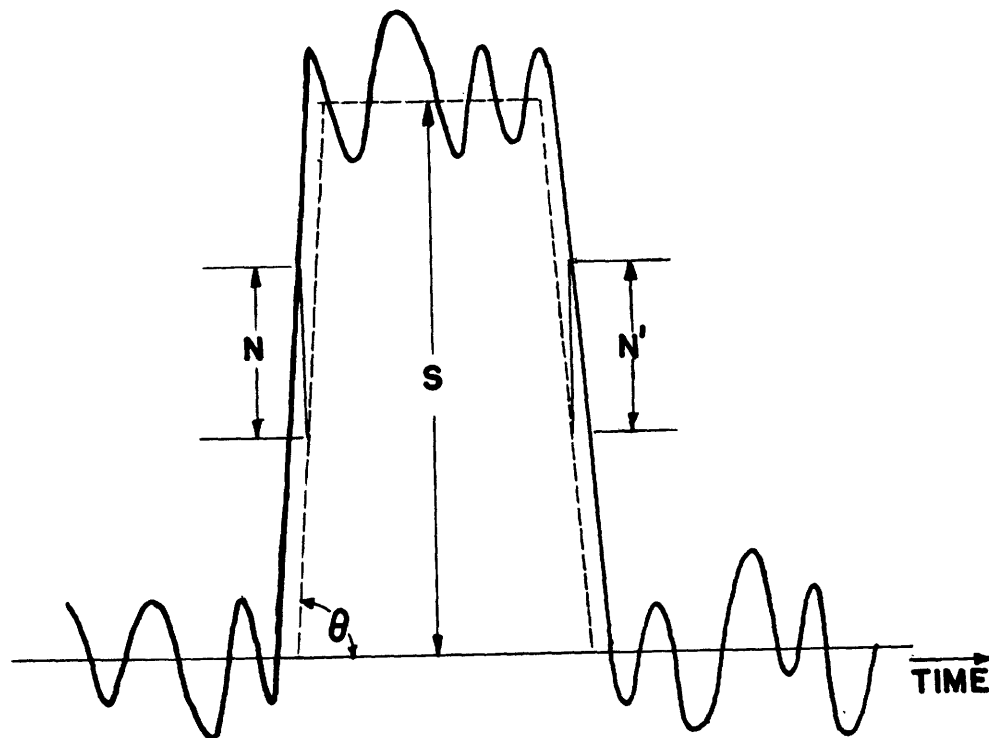


Figure 9. The effect of noise on the location of the edge of a received pulse.

represent the pulse with the noise superimposed. The distances N and N' represent the noise amplitudes during the pulse rise and fall, respectively. For noise components which will give rise to signal in the audio-modulation range, N and N' are essentially equal. If Δw represents the change in pulse width due to the presence of noise, then

$$\Delta w = 2N \cot \theta = \frac{2}{B} \left(\frac{N}{S} \right). \quad (6A)$$

The noise-modulation factor is thus

$$\frac{\Delta w}{w_o} = \frac{2}{w_o B} \frac{N}{S}. \quad (7A)$$

Of this noise, however, essentially only that which lies within $\pm F_a$ of the subcarriers at $f \pm nf_p$ will get through the audio amplifier of cutoff frequency F_a . The audio-noise-modulation factor will therefore be

$$\sqrt{\frac{2F_a}{f_p}} \frac{\Delta w}{w_o} = \sqrt{\frac{2F_a}{f_p}} \frac{2}{w_o B} \left(\frac{N}{S} \right). \quad (8A)$$

The factor 2 under the radical is due to the presence of audio sidebands on both sides of the subcarrier, and the radical itself is due to the random relations between noise sidebands, which makes the noise amplitude proportional to the square root of the bandwidth.

Since the carrier is only present during a fraction of the time, namely $(w_o f_p)$, if the transmitting power is determined by average power, then the ratio of signal height in the PWM case to that in the a-m case is

$$\frac{S}{S_o} = \frac{1}{\sqrt{w_o f_p}} \quad (9A)^1$$

-
1. Since a 100 per cent a-m wave has twice the amplitude but only 3/2 of the energy of an unmodulated wave, the ratio in Eq. (9A) would more correctly be

$$\frac{S}{S_o} = \frac{\sqrt{3/2}}{2} \frac{1}{\sqrt{w_o f_p}}$$

if the a-m comparison signal is 100 per cent modulated. It is, however, necessary for small percentages of modulation to be received, as well, in order to get the information transmitted. Because the proper average value to replace $\sqrt{3/2}/2$ is thus uncertain, we omit the factor in our equations, since it does not differ greatly from unity anyway.

where S_o is the a-m signal amplitude. We also note that because of the difference in receiver bandwidths, (that of the a-m receiver being $2F_a$) the corresponding noise amplitudes have the ratio

$$\frac{N}{N_o} = \sqrt{\frac{B}{2F_a}}. \quad (10A)$$

Substituting Eqs. (9A) and (10A) into Eq. (8A), we obtain for the audio-noise modulation-factor in PWM.

$$\begin{aligned} \text{Audio-Noise-Modulation Factor (PWM)} &= \sqrt{\frac{2F_a}{f_p}} \frac{2}{w_o B} \frac{N}{S} \\ &= \sqrt{\frac{2F_a}{f_p}} \frac{2}{w_o B} \sqrt{\frac{B}{2F_a} w_o f_p} \frac{N_o}{S_o} = \frac{2}{\sqrt{w_o B}} \frac{N_o}{S_o}. \end{aligned} \quad (11A)$$

Thus the greater $w_o B$, the greater the noise reduction. The maximum value of w_o which is possible without distortion at 100 per cent modulation is

$$w_o' = \frac{1}{2f_p}. \quad (12A)$$

Using this value in Eq. (11A), we have

$$\text{Best Audio-Noise-Modulation Factor in (PWM)} = 2 \sqrt{\frac{2f_p}{B}} \frac{N_o}{S_o}. \quad (13A)$$

Corresponding to Eq. (1A) for FM, we have for the noise reduction factor of PWM (as compared with AM)

$$\text{Noise-Reduction Factor} = \frac{1}{2} \sqrt{\frac{B}{2f_p}} = \frac{1}{2} \sqrt{\frac{b}{f_p}}. \quad (14A)$$

Noise reduction thus increases with the square root of the ratio b/f_p .

It thus appears that the amount of noise reduction depends upon the number of subcarriers produced, each of which individually carries the information as sidebands.

The amount of noise reduction obtainable increases with the bandwidth, and is potentially unlimited if the bandwidth can be made large enough. It should be pointed out, however, that the average level of the noise background also increases as the bandwidth increases, and since the top and bottom limiters cannot be made to operate until the S/N ratio is greater than 2, range and noise reduction are opposing factors in choosing the bandwidth in PWM, just as they are in FM.

C. Multichannel Operation with Coherent Audio. The foregoing examples suggest another type of operation which might at first be expected to give noise reduction. This is the simultaneous transmission by AM, for example, of the same audio program on separate carriers in several different frequency channels. The coherent audio programs are then to be combined after demodulation. Let us suppose that there are n channels, each separately having the same average signal-to-noise ratio which we will designate as S_o/N_o . If these outputs are combined after demodulation, the signal-to-noise ratio will be

$$\frac{nS_o}{\sqrt{n}N_o} = \sqrt{n} \frac{S_o}{N_o}, \quad (15A)$$

since the signals are coherent but the noise is not.

Next suppose that the same total energy is used to transmit the signal in one channel. Then the signal amplitude level will be increased to $\sqrt{n}S_o$ and the signal-to-noise ratio will again be $\sqrt{n} S_o/N_o$. Thus for a given total amount of energy available for transmission, the S/N ratio is not improved by separation into several different channels with coherent audio. However, if the energy per cycle is constant, then separation will improve the ratio, and the improvement will be proportional to the square root of the number of channels. (See Fig. 10).

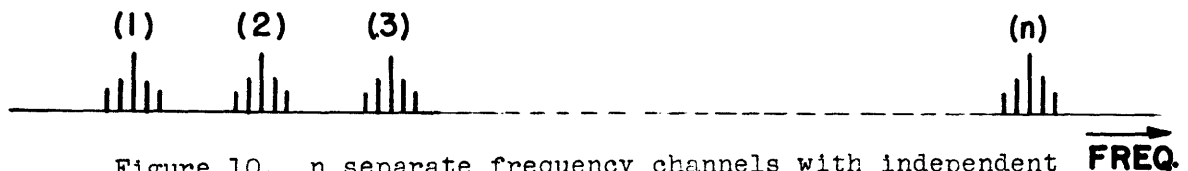


Figure 10. n separate frequency channels with independent carriers carrying the same audio program.