THEORY OF RADAR INFORMATION

by

P.M. Woodward

Radar is a system of measurement rather than communication, yet it is quite possible to apply information theory to it, in order to see whether the very small received signals inherently contain as much information as those of an ideal communication system working at the same signal-to-noise ratio. It turns out that they do, very nearly, but this is not really what makes radar a suitable topic for this symposium. The main interest is in the type of coding it represents. Shannon has pointed out [1] that when the natural number of dimensions of a message is artificially increased by mapping non-topologically into a signal space of higher dimensions, a marked threshold effect is produced. Radar exhibits such a threshold particularly well and it is to this that I wish to direct attention.

We shall consider only the most obvious radar problem, that of measuring the range of a stationary target. This is one-dimensional information, and it is important to realize that the way in which it is coded is almost entirely beyond control: it is determined by the very nature of radar. A known periodic waveform is transmitted, echoed, and received again. The time τ which elapses between transmission and reception represents the range of the target. Unlike most systems of electrical signalling, the choice of transmitted waveform does not represent the required information but forms part of the observer's a priori knowledge. All the required information is embodied in the time-delay of the received waveform. If we fix the received signal energy, this leaves the received signal with only one degree of freedom, but the noise which goes with it will have many degrees of freedom. In geometrical language, the received signal is treated as a point in a multi-dimensional waveform space, and the ensemble of received signals lie along a one-dimensional twisted curve embedded in this hyper-space, and incidentally lying on the surface of a hyper-sphere. One therefore expects a threshold effect as soon as the noise perturbation is sufficiently large to short-cut the convolutions of this message locus, and introduce wild ambiguities of range measurement.

However, there is really no difference in principle between interpreting a radar signal and any other kind of signal. The observer always has some a priori knowledge of what he is trying to receive, and it merely happens that in radar he knows, apart from noise, the exact shape of the waveform. Obviously his first step is, effectively, to place the transmitted and received waveforms side by side, and try to estimate the time-shift τ. If the possible values of τ are known a priori to be continuously distributed, it should be clear that in the presence of noise he cannot hope to determine τ exactly, because this would represent an infinite quantity of information and would require an infinite amount of signal energy. His best estimate is therefore bound to be subject to error. One might suppose, then, that the first problem is to determine theoretically the spread of his best guesses over an ensemble of received waveforms all resulting from the same true value of τ. The average quantity of information obtained in any determination would then be given by

$$I = H(y) - H_x(y).$$

Shannon [2] uses x to symbolise transmission and y to symbolise reception, but it must be understood that the "transmitted message" in radar has nothing to do with the radar transmitter; it refers to the true value of τ, while y refers to the observer's estimate of τ. The entropy $H(y)$ is a measure of the spread of variability of all previous guesses over a complete ensemble of true values, and therefore represents the a priori uncertainty about the next one, while the conditional entropy $H_x(y)$ represents the spread of the guesses in an ensemble in which the true value is fixed and only the noise-sample is different. This might seem to be the obvious approach, but it is not altogether satisfactory because it appears to contain subjectives elements. It appears to depend on the particular way in which the observer makes his guess. If he were no good at guessing, $Hx(y)$ would be large and the quantity of information small. The maximum quantity of information latent in the received waveform could only be evaluated by this method by giving rules for making the best possible guess.

This whole difficulty is avoided by starting from the alternative formula.

$$I = H(x) - H_y(x)$$

Here $H(x)$ is the entropy of the a priori distribution for the true value of τ, not the guessed value. The equivocation, $Hy(x)$, represents the observer's uncertainty about the true value of τ on any one occasion, i.e. when the received waveform is fixed. We have, then, to consider a fixed received waveform, arising from a true range τ_0 say, and find out from it, not simply the most probable value of τ it might represent, but a complete probability distribution for all possible values of τ. This is not subjective at all: it represents the matter-of-fact frequency distribution of those values of τ which could have given rise to this particular τ_0-waveform. The whole problem thus centres round two distributions, the a priori distribution, called $p_0(\tau)$, and the a posteriori distribution denoted $p_1(\tau)$. The difference of the two corresponding entropies is the quantity of information gained.

Without entering into too much mathematical analysis, we may indicate in outline how $p_1(\tau)$ is found, because this is really the heart of the problem. Let us suppose that the received waveform is observed for a duration of time D, and denote it by

$$y(t) = u(t - \tau_0) + n(t),$$

where $u(t)$ is what would have been received with no time-delay τ_0 and no noise $n(t)$, and is presumed known to the observer. If he supposes the true value of the range to be τ, he calculates that the noise would have to be

$$y(t) - u(t - \tau).$$

The probability of such noise then determines inversely(3) the probability that his hypothesis τ was correct. Now the probability density for the random noise fluctuation in its multi-dimensional waveform space is proportional to

$$e^{-W/N_0}$$

where W is the total energy of the noise over the interval D, and N_0 is the mean noise power per unit bandwidth, so the probability density in favour of the hypothesis τ is given by

$$p_1(\;) = \lambda \exp\left[-\frac{1}{N_0}\int_D \{y(t) - u(t - \;)\}^2 dt\right]$$

if the a priori probability distribution $p_0(\tau)$ is uniform. We shall in fact take $p_0(\tau)$ to be uniform over an arbitrary interval of range T, within which it is assumed that τ is known to lie. (This a priori interval T may be less than or equal to the repetition period of $u(t)$.) Consequently, $p_1(\tau)$, is only defined over the interval T and λ must be chosen to normalize it accordingly. We may note in passing that the most probable value of τ is that which gives the least mean square departure of y from $u(t - \tau)$.

Any factors in the expression for $p_1(\tau)$ which do not depend on τ may obviously be absorbed into the normalizing constant, and if D is an integral multiple of the repetition period, we are left only with

$$p_1(\tau) = \lambda \exp\left[\frac{2}{N_0}\int_D y(t)\, u(t - \tau)\, dt\right]$$

This expression is interesting because the integral in it has the familiar form of the output from a linear filter whose impulse response is $u(-t)$, the time-reverse of the transmitted waveform, and whose input is simply the received waveform $y(t)$. The limits of integration for such a filter, however, would go from $t-D$ to t, whereas the limits above are quite fixed. To follow up this question would take too long, but it does have an interesting bearing on the topic of optimum filtering, and especially on the result of Van Vleck and Middleton [4] that just such a filter would give the maximum peak signal-to-noise ratio.

Having decided on a form for $p_0(\tau)$ and derived an expression for $p_1(\tau)$, there would seem to be nothing to prevent us, in principle, from calculating the corresponding entropies H_0 and H_1 forthwith. Indeed, H_0 can be seen immediately to equal $\log T$, but the calculation of H_1 takes much longer and cannot be obtained without first investigating the properties of $p_1(\tau)$. This distribution itself is, in any case, more important both theoretically and practically than its entropy, for while the entropy enables us to determine the quantity of information for comparison with Shannon, it leaves the interpretation of the information - or lack of it - entirely out of account. We may begin examining $p_1(\tau)$ by writing the full expression for the received waveform in place of $y(t)$ in the integral. Then

$$p_1(\tau) = \lambda e^{g(\tau) + h(\tau)}$$

where

$$g(\tau) = \frac{2}{N_0}\int_D u(t - \tau_0)\, u(t-\tau)\, dt$$

$$h(\tau) = \frac{2}{N_0}\int_D n(t)\, u(t-\tau) dt$$

It will be seen that $g(\tau)$ is obtained from the signal actually received, and $h(\tau)$ from the noise. An observer, it must be remembered, could form $p_1(\tau)$ after any one observation, but he could not of course determine $g(\tau)$ and $h(\tau)$ separately as we are doing.

Consider first the "signal function" $g(\tau)$. The waveform $u(t)$ which generates it, is a high-frequency function of t and this makes $g(\tau)$ a high-frequency function of τ, but the envelope of $g(\tau)$ is slowly varying (by comparison with the carrier) because it is controlled by the bandwidth of $u(t)$. For the present, let us forget the carrier in $g(\tau)$, to which we shall return at the end, and concentrate on the envelope alone. It is almost obvious and not difficult to show mathematically, that this has a maximum at $\tau = \tau_0$, where its value is $2E/N_0$, E being the total received signal energy. In fact, the envelope of $g(\tau)$ will have a single peak at τ_0 and be negligible, if not precisely zero, elsewhere. This last is not a mathematical deduction, it is a statement applying to practical waveforms, whether amplitude or frequency modulated, and is the very feature by which the suitability of any waveform for radar may be judged. It ensures that the message-locus is well spread out in waveform space.

The "noise function" $h(\tau)$ has certain features in common with the signal function. It has, for example, a slowly varying envelope controlled by the bandwidth of $u(t)$, but it is a stationary random function of τ, and has all the characteristics of noise which has passed through a high-frequency pass band filter, except that the RMS value of its envelope is $2\sqrt{E/N_0}$, which happens to increase as the received signal energy increases.

We now have sufficient facts to discuss all that is of qualitative importance about the a posteriori distribution. It is clear, to start with, that $p_1(\tau)$ is partly a random function of τ, owing to the presence of $h(\tau)$. It may seem a confusing idea that a probability distribution should itself be random, but it is simply a matter of being clear about ensembles. The distribution $p_1(\tau)$ represents the frequency with which various ranges τ could give rise to the exact waveform which we have privately stated to be due on this occasion to a range τ_0. It will obviously depend, to some extent, on the particular way in which the noise happened to act on this occasion. With a fixed true range τ_0, therefore, $p_1(\tau)$ will be different from one occasion to another, and it is just this randomness which $h(\tau)$ represents.

It should be clear that if $E < N_0$, there will be no marked accumulation of probability near τ_0, because in terms of envelopes the RMS value $2\sqrt{E/N_0}$ of $h(\tau)$ will exceed the peak value $2E/N_0$ of $g(\tau)$. Throughout the remainder of the theory, we are in fact forced to assume that

$$E >> N_0$$

for purely mathematical reasons, but since it is a necessary condition for satisfactory observation, the assumption is not seriously embarrassing. This energy criterion is not in any way peculiar to radar, and is not connected with the threshold effect due to coding. When $E >> N_0$, the peak value of g will be so large that, after normalization, the probability distribution $p_1(\tau)$ will be almost unaffected by the presence of $h(\tau)$, at least for most values of τ. Indeed, the whole of the peak in $g(\tau)$ except a small region immediately surrounding its apex will be similarly reduced, and expansion of the exponent about τ_0 will yield a Gaussian distribution for $p_1(\tau)$. The standard deviation works out to be

$$\sigma = \frac{1}{\beta}\sqrt{\frac{N_0}{2E}}$$

where β^2 is the second moment of the power spectrum of the transmitted waveform about its centroid. The quantity β is the one really important parameter associated with the transmitted waveform, and is equal, apart from a constant factor, to the "effective bandwidth" adopted by Gabor (5). The standard deviation σ gives us a tentative measure of the accuracy with which τ may be determined. As would be expected, σ decreases with an increase in signal energy and also with an increase of transmitter bandwidth such as might arise from the use of shorter pulses.

Two things appear at first to be wrong with the above result: the first is that the a posteriori distribution is apparently centred exactly on τ_0. Since it is within the observer's power to determine $p_1(\tau)$ on any occasion, apparently he could look for the centre of the peak and so determine τ_0 exactly, which would contradict the very uncertainty the distribution is supposed (inversely) to describe. It is, of course, the noise function $h(\tau)$ which removes this paradox and it can be shown that its effect is to disturb the position of the peak in a random manner and by just the right amount. Its width is largely unaffected by the presence of h, however, so that σ remains a valid measure of range error. The second objection arises the moment comparison is made with general communication theory,(2) which sets a limit of E/N_0 natural units of information on any message of energy E. Yet if β is increased and E/N_0 kept fixed, it would appear that σ can be made as small as desired. This would make H_1 as small as desired and corresponds to increasing the quantity of received information without limit, which would contradict Shannon's fundamental theorem. Again it is the presence of $h(\tau)$ which prohibits this. As β is increased, the autocorrelation interval in $h(\tau)$ is proportionately reduced and more statistically independent opportunities exist for $h(\tau)$ to produce a spurious peak well in excess of its RMS value, and therefore big enough to show up on the normalized a posteriori distribution. Every time $h(\tau)$ succeeds in doing this, a spurious Gaussian distribution is introduced in $p_1(\tau)$ and there comes a stage when there are so many of these that $g(\tau)$ might almost as well not be present. These are conditions of completely ambiguous reception, in which the accuracy of measurement might be high if the observer only knew which peak to select. The ambiguity operates in such a way as to reduce the quantity of information by just the amount required to bring it within the fundamental limit of E/N_0 natural units.

The a posteriori distribution thus describes two quite different kinds of uncertainty of reception. First there is the small connected region of uncertainty in the neighbourhood of the true range. This must inevitably be present as long as the signal energy is finite, and is no cause for complaint. But in addition, when β is too large, there is a wild uncertainty, even though the received signal energy is large, which prevents an observer from knowing even approximately whereabouts in the interval T the true value of τ is to be found. We may call this effect non-topological error, because it arises from the non-topological mapping of a one-dimensional ensemble of messages into a multi-dimensional waveform space. This effect shows up one of the weaknesses of judging a communication system solely in terms of quantities of information. When non-topological error is present, the system is useless from a practical point of view and yet the mathematical quantity of information may be quite large. It is a question of intelligibility rather than information.

Intelligibility is a concept associated with meaning, and it is not to be expected that a general theory of it should be quite as straightforward as that of information-content. However, it may happen in particular problems that a quantitative assessment is possible. In radar, it is especially simple. We define unintelligibility by the non-topological ambiguity of reception, A, given by the area under $p_1(\tau)$ which lies nowhere near the true value. The Figure illustrates the dependence of A on log $T\beta$ and E/N_o. As remarked above, ambiguity increases with β, but it also depends on E/N_o and is responsible for a threshold of intelligibility as the total received energy increases, as it would with increasing time of obervation. The threshold extends over one or two units of E/N_o and occurs (very roughly) where

$$\log T\beta = E/N_o$$

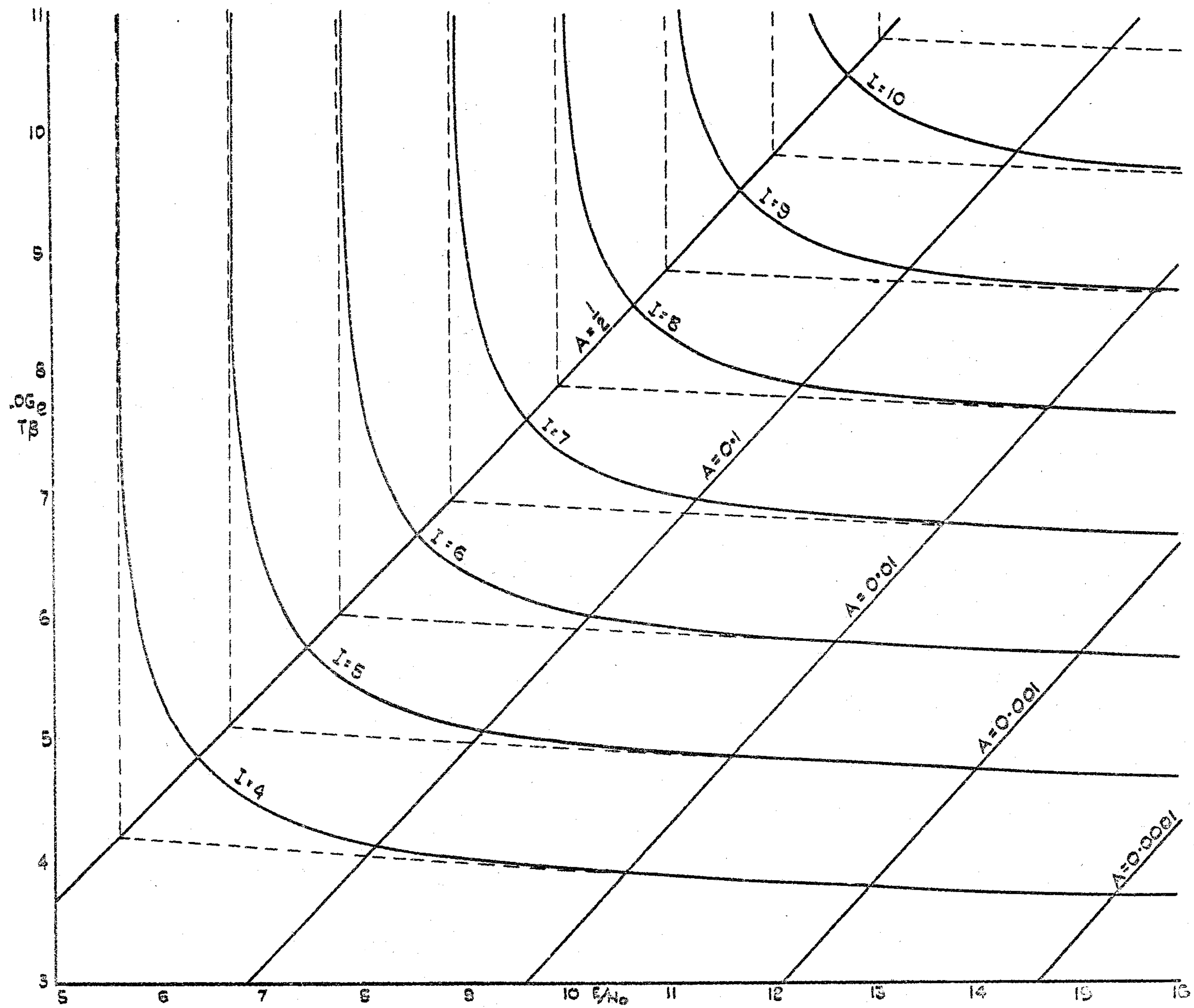
Contours of information, I, are also shown. It will be seen that they behave in a markedly different manner on either side of the threshold. In the ambiguous region they would be strictly asymptotic to the fundamental limit $I = E/N_o$ if the difference of H_o and H_1 had been evaluated in a straightforward manner. But in making a detailed calculation of I, we have in fact tampered with the a posteriori distribution by smoothing out the fine structure produced by the "carrier" in $g(\tau)$ and $h(\tau)$. The effect of the carrier is to make the signal peak in $p_1(\tau)$ consist, not of a single Gaussian distribution of standard deviation σ, mentioned before, but of a closely packed sequence of very narrow Gaussian distributions under a Gaussian envelope with this standard deviation. We have thought it best to remove the fine structure information by short-scale smoothing of $p_1(\tau)$ because it is of no value in practice. It arises from the comparison which the observer could make between the carrier phases of transmitted and received waveforms, and its removal increases the a posteriori entropy by a term in log (E/N_o), which is just what prevents the absolute limit being attained in the ambiguous region. In the unambiguous region, the information I is limited solely by the topological error in range and increases comparatively slowly with energy. This is because additional energy, once the threshold has been crossed, is not employed in the systematic removal of ambiguity but in the improvement of range accuracy by continued repetition of information already partly known.

To sum up, it is seen that there are in radar two quite different conditions of reception. There is ambiguous reception in which the information rate is high and the intelligibility low, and there is unambiguous reception in which the information rate is low but intelligibility is high. It would therefore appear that merely to evaluate quantities of information, and compare them with the ideal limits is not adequate guide to the behaviour of a communication system. Some figure for intelligibility may also be necessary in particular problems. The analysis of radar shows that loss of intelligibility can be measured, in at least this one problem, by non-topological error, but it is quite possible to imagine systems in which non-topological error would not imply complete loss of intelligibility. Even in radar, such a situation is not inconceivable. It really depends on the purpose for which range information is required and this is a question of meaning. In fact the whole question of defining a natural message dimensionality is one of meaning. So in suggesting that theree is some connection between intelligibility and non-topological error, which itself hinges on the specification of a natural message dimensionality, we may perhaps be merely replacing one vague term by another.

I must conclude by thanking my colleague, Mr. I.L. Davies, for his part in the work I have summarized (6). Between the present version and the more extended one (loc.cit.), there are certain inconsistencies of notation, deliberately introduced here for simplicity.

REFERENCES

1. Shannon C.E. "Communication in the Presence of Noise", Proc. I.R.E., 37, 10 (1949)
2. Shannon C.E. "A Mathematical Theory of Communication", Bell Sys.Tech.Jour., 27, 379 and 623 (1948)
3. Cherry E.C. Historical Introduction (see page 22)
4. Van Vleck J.H. & Middleton D. "A Theoretical Comparison of the Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise", Jour.App.Phys. 17, 940 (1946)
5. Gabor D. "Theory of Communication", J.I.E.E., 93 (Pt.III), 429 (1946)
6. Woodward P.M. & Davies I.L. "Theory of Radar Information", Phil. Mag., Ser.7, 41 1001 (1950)



INFORMATION AND AMBIGUITY CONTOURS.

T = A PRIORI INTERVAL OF TIME-DELAY

β = WAVEFORM BANDWIDTH (SPECIALLY DEFINED)

E = TOTAL RECEIVED SIGNAL ENERGY.

N_0 = NOISE POWER PER UNIT BANDWIDTH.

I = QUANTITY OF INFORMATION IN NATURAL UNITS.

A = AMBIGUITY BETWEEN SIGNAL AND NOISE.