THE THEORY OF HETERODYNE RECEIVERS

(A DISCUSSION ON "THE HETERODYNE RECEIVING SYSTEM"[1] BY JOHN L. HOGAN, JR.)

BY

BENJAMIN LIEBOWITZ

Certain misconceptions seem to be current regarding the mode of amplification in receivers of the heterodyne type, in which a local radio frequency current is made to produce beats in conjunction with the received current. It is usual to assume, for example, that the maximum energy present in the antenna due to both currents is proportional to $(i_1 + i_2)^2$, and the minimum to $(i_1 - i_2)^2$, giving an energy fluctuation of $4i_1 i_2$; whereas the energy due to the received current alone would be proportional to i^2_1; from which it is deduced that the ratio of amplification is $2\frac{i_2}{i_1}$. The incorrectness of this view will appear from the following discussion.

Suppose the received and local currents to be simple harmonic, the first expressed by

$$i_1 = A \sin pt,$$

and the second by

$$i_2 = B \sin qt.$$

Let L denote the effective inductance of the antenna, and W the instantaneous value of the energy present in L. Then

$$W = \frac{1}{2} L (i_1 + i_2)^2 = \frac{1}{2} L (A \sin pt + B \sin qt)^2$$

$$= \frac{1}{2} LA^2 \cdot \sin^2 pt + \frac{1}{2} LB^2 \cdot \sin^2 qt + ABL \cdot \sin pt \cdot \sin qt$$

$$= \frac{1}{2} L [A^2 \sin^2 pt + B^2 \sin^2 qt + AB \cos (p - q) t - AB \cos (p + q) t].$$

The instantaneous value of the energy has, therefore, four com-

[1] A paper printed in THE PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, 1913, Volume 1, Part 3, page 75, *et seq.*

ponents, which for convenience may be denoted by W_1, W_2, W_3, and W_4. In dealing with energy, however, it is necessary to consider not the instantaneous values, but the average values. Thus, it would be incorrect to assume that W_3 represents the energy available for producing signals because it fluctuates with audible frequency, without regard to its average value; for, if that were the case, the received current acting alone would have no energy available for producing signals. (It must be borne in mind that the energy present in the antenna is under consideration, without regard to the manner in which that energy is utilized.) It is, therefore, only the average value of W which is of importance, and to find this average value we have merely to find the average values of the four components and add them. The period of W_1 is $\frac{\pi}{p}$, and if we integrate W_1 from any instant, t, to an instant one period later, i. e., to $t+\frac{\pi}{p}$, and then divide by the period, we get:

Average of

$$W_1 = \frac{p}{\pi}\cdot\frac{LA^2}{2}\int_t^{t+\frac{\pi}{p}} \sin^2 pt\cdot dt = \frac{p}{\pi}\cdot\frac{LA^2}{4}\int_t^{t+\frac{\pi}{p}}(1-\cos 2pt)\,dt$$

$$= \frac{p}{\pi}\cdot\frac{LA^2}{4}\left(t-\frac{1}{2p}\sin 2pt\right)_t^{t+\frac{\pi}{p}}$$

$$= \frac{LA^2}{4} = \frac{1}{2}L\frac{A^2}{2} = \frac{1}{2}LI_1^2,$$

where I_1 is the effective value of i_1.

Similarly, it can be shown that

$$\text{Average of } W_2 = \frac{LB^2}{4} = \frac{1}{2}L\frac{B^2}{2} = \frac{1}{2}LI_2^2,$$

where I_2 is the effective value of i_2. Turning now to W_3, the period is $\frac{2\pi}{(p-q)}$ and if W_3 is integrated from any instant, t, to $t+\frac{2\pi}{(p-q)}$, and the result divided by $\frac{2\pi}{(p-q)}$, we get:

$$\text{Average of } W_3 = \frac{p-q}{2\pi}\cdot\frac{ABL}{2}\int_t^{t+\frac{2\pi}{p-q}}\cos(p-q)t\cdot dt$$

$$= \frac{1}{2\pi}\cdot\frac{ABL}{2}\Big[\sin(p-q)t\Big]_t^{t+\frac{2\pi}{p-q}} = 0.$$

Similarly, Average of $W_4 = O$.
Hence the average value of the energy present in the antenna is given by

$$\text{Average value of } W = \frac{1}{2} L (I_1^2 + I_2^2).$$

In other words, when currents of different frequencies are present in a circuit, the average value of the energy present is equal to the sum of the average values of the energy due to each current separately. In fact, the law of conservation of energy demands this; and furthermore, it is a well-knowm theorem in electrical theory, that if a number of currents of different frequencies and of effective values I_1, I_2, I_3 . . . are present in a resistance R, then the average rate of heat development is equal to

$$R (I_1^2 + I_2^2 + I_3^2 + \quad . \quad . \quad . \quad).$$

It is clear, therefore, that receivers of the heterodyne type do not amplify by increasing the energy component of the received currents in the antenna. Before considering the true mode of amplification in such receivers, it is necessary to distinguish between two types of amplification, namely: (1), by infusing new energy into the received oscillations, and (2), by increasing the efficiency of the receiving apparatus. As an example of amplification by the infusion of new energy into the incoming oscillations, consider receivers which employ an electron stream acted on by the currents to be amplified. In such receivers the resulting variations in the electron current can be made many times greater than the amplitude of the original current, so that here we have actually reproduced the original currents, but with greater energy. As an example of the other type of amplification, consider the ordinary telephone receiver. That the presence of the permanent magnet produces an enormous increase in the amplitude of the vibrations of the diaphragm is too well known to require mention, but it cannot be said that the permanent magnet puts new energy into the system. This is clearly amplification by increasing the efficiency of the receiving apparatus; the energy in the sound can never exceed the energy in the received current.

The theory of ordinary telephone receivers, as usually presented, is worthy of further scrutiny in this connection. The force of attraction between a magnet and a piece of iron is directly proportional to the square of the flux. If this flux has a constant component ϕ_1, and a variable component $\phi_2 \sin pt$, the force at any instant is proportional to

$$(\phi_1 + \phi_2 \sin pt)^2 = \phi_1^2 + 2\,\phi_1\,\phi_2 \sin pt + \phi_2^2 \sin^2 pt.$$

ϕ_1 is usually very large compared with ϕ_2; hence, neglecting the last term, the variable force is proportional to

$$\phi_2\,\phi_1 \sin pt,$$

and since ϕ_2 is proportional to A, the amplitude of the received current, the variable force is proportional to

$$A\,\phi_1 \sin pt.$$

Hence, the larger the permanent flux the larger the useful force.

This theory is correct, however, only so long as the motion of the diaphragm is very small; i. e., only so long as the efficiency of the receiver is very low. The telephone receiver is, after all, a synchronous motor, and the excursions of the diaphragm produce a back e. m. f. in the coils, just as in any other motor. This back e. m. f. is ordinarily negligible, because the efficiency of the telephone receiver is ordinarily very low. For higher efficiencies, however, this back e. m. f. would attain values of the same order of magnitude as the resistance reaction and inductance reaction, and the effect of this would be to diminish the incoming current. Hence, the useful force cannot be indefinitely increased by increasing the permanent flux; the best that can be attained is an increase in the efficiency. This phenomenon is analogous to the events in an ordinary motor; as the motor speeds up the back e. m. f. becomes increasingly important, and, at full speed, is the largest reaction in the circuit if the motor is efficient.

The theory of the electrostatic telephone receiver in which a constant difference of potential is maintained between the plates is entirely analogous. If V_2 represents the constant e. m. f. and $V_1 \sin pt$ a superimposed variable e. m. f., then, since the force between the plates varies as the square of the e. m. f., the force at any instant is proportional to

$$(V_2 + V_1 \sin pt)^2 = V_2^2 + 2\,V_1\,V_2 \sin pt + V_1^2 \sin^2 pt.$$

Again neglecting the last term, the variable component of the resulting force is proportional to

$$V_1\,V_2 \sin pt.$$

This could be indefinitely increased by increasing V_2 indefinitely; but here again a back e. m. f. is produced as soon as the device becomes appreciably efficient, and this back e. m. f. results in a decrease of V_1 when V_2 is increased. In this receiver, the battery which maintains the constant voltage does not supply any useful energy. It acts in a manner entirely analogous to the permanent magnet in the ordinary telephone receiver. The acoustic energy of such a device can never exceed the energy in the received currents.

Turning now to receivers of the heterodyne type, consider, for example, the form shown in Figure 11 of Mr. Hogan's paper, "The Heterodyne Receiving System." (See these PROCEEDINGS, July, 1913.) No attempt will be made to give a rigorous theory of the problem presented by these circuits, but an approximation of the facts sufficiently close for practical purposes will be presented.

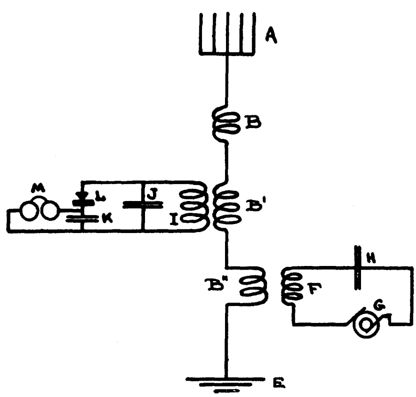

FIGURE 11
(OF ORIGINAL PAPER)

In the circuit IJ of Figure 11 of Mr. Hogan's paper, suppose that the two currents,

$$i_1 = A \sin pt \text{ and } i_2 = B \sin qt$$

are flowing. The voltage across the condenser J will be

$$v = \frac{1}{C}\int (i_1 + i_2) d t = a \cdot \cos pt + b \cdot \cos qt$$

where $a = -\frac{A}{p c}$, $b = -\frac{B}{q c}$, and C is the capacity of condenser J. It can be shown, in a manner entirely similar to the previous cases, that the average value of the energy present in this condenser is proportional to $(a^2 + b^2)$ and not to $(a + b)^2$. Suppose now that a is much smaller than b, as is the case in practice; then, denoting the difference in the amplitudes by h $(h = b - a)$, we may write v in the form

$$v = a(\cos pt + \cos qt) + h \cos qt$$
$$= 2a \cos\left(\frac{p-q}{2}\right) t \cdot \cos\left(\frac{p+q}{2}\right) t + h \cos qt.$$

The voltage across the condenser J can, therefore, be resolved into two components,

$$v_1 = 2a \cos\left(\frac{p-q}{2}\right) t \cdot \cos\left(\frac{p+q}{2}\right) t \text{ and } v_2 = (b - a) \cos qt.$$

The first may be called the "beat" component, the second the "sustained" component. The graph of v_1 is of the form shown by Mr. Hogan (loc. cit.) in Figure 4, curve C; the graph of v_2 is a simple sine curve. The effects of these components in the rectifying detector circuit KLM will now be separately considered.

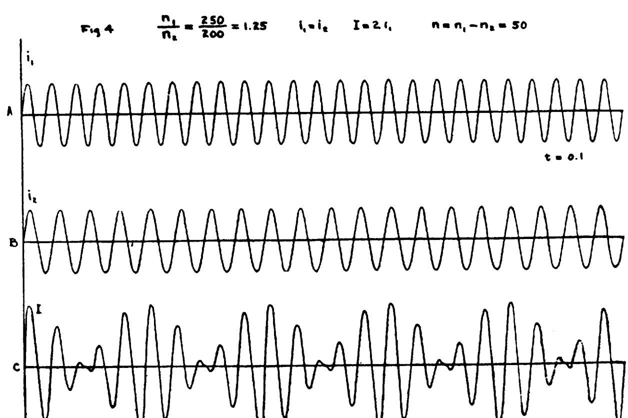

FIGURE 4
(OF ORIGINAL PAPER)

Owing to the rectifying and integrating action of the detector circuit, the rapidly varying voltages v_1 and v_2 give rise to constant or slowly varying unidirectional currents through the telephone receivers M. More specifically, the "beat" voltage component v_1 tends to produce a current in the detector circuit of the form shown in Figure 6, curve C of Mr. Hogan's paper, with the negative loops omitted, however. But owing to the high resistance of the detector and the large inductance of the telephone receivers, this series of unidirectional current loops is smoothed out into the form shown by Mr. Hogan's curve E of Figure 6. The maximum value of these smoothed

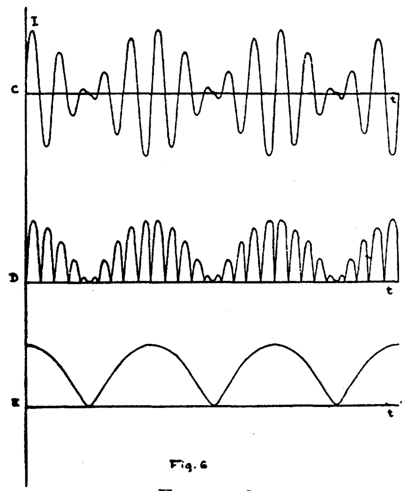

FIGURE 6
(OF ORIGINAL PAPER)

loops will obviously be proportional to the maximum amplitude of the "beat" component, i. e., to 2a, and since the minimum value is zero, the *amplitude* of the variable component of the telephone current will be directly proportional to a, which in turn is proportional to A, the amplitude of the received current. We see, therefore, that the amplitude of the variable telephone current is directly proportional to the amplitude of the received current.

Consider now the sustained voltage component (b − a) cos qt. It will obviously give rise to a practically constant, unidirectional current thru the telephone receivers. This might result in two improvements (1), a slight increase in the sensitiveness of the telephones resulting from a possible increase in the permanent flux, and (2) an increase in the detector sensitiveness resulting from working on a better part of its characteristic. Whether or not these improvements exist is immaterial from our present point of view, because if they do exist, the same results could be obtained by suitably placing a battery in the detector circuit. In any case it is clear that these last two improvements would be amplification by increase in efficiency, and not by infusion of new energy. Barring, therefore, possible improvements which could be obtained by the use of a battery, we see that no matter how large the amplitude of the local current may be, only that part is useful whose amplitude is equal to the amplitude of the received currents.

It will now be seen that the maximum true amplification, i. e., amplification by infusion of new energy, which the heterodyne receiving system can produce is four. To prove this, suppose that the local current is absent, that the same system of circuits is employed, and that a "chopper" in series with the telephones is used to break up the sustained received oscillations into trains of audible frequency. The maximum value of the voltage across condenser J will now be equal to $-\frac{A}{p\,c}$, assuming the current in I J to be expressible by A sin pt, as before. The resulting pulsating current thru the telephone receivers will, therefore, vary between O and a maximum value proportional to $\frac{A}{p\,c}$, i. e., proportional to A. Hence the amplitude of the telephone current will be proportional to $\frac{A}{2}$. But we have seen that when the local current is present, the amplitude of the telephone current is proportional to A, the factor of proportionality being the same in both cases; and since the acoustic energy is proportional to the square of the telephone current, it follows that the useful energy is four times as great when the local current is present as it is when the local current is not present. At the very most, therefore, the maximum true amplification which the heterodyne receiving system can produce is four. Any additional amplification which has been observed must be regarded as due to an improvement in the efficiency of the receiving system, and not to any particular virtue of the heterodyne principle. Such additional amplification is obtained, for example, by making the beat frequency equal to the natural frequency of the telephone receivers.

The form of the heterodyne receiver which we have been discussing, i. e., the form shown in Figure 11, is the most efficient of all those described in Mr. Hogan's paper. The other forms shown offer considerable mathematical difficulties when the energy relations are analysed, altho the principal forces acting may be readily found. Thus, for example, in Figure 10, suppose that the voltage v across the electrostatic telephone receiver D is expressible by

$$v = a \cos pt + b \cos qt,$$

the first term being due to the incoming, the second to the local oscillations. Since the force between the plates of the receiver

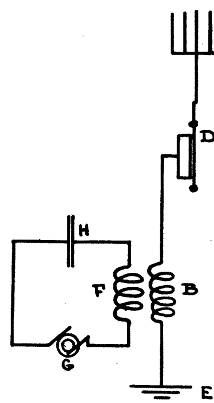

Fig. 10.

FIGURE 10
(OF ORIGINAL PAPER)

is proportional to $(v)^2$, we have, denoting the force by F and a proportionality factor by K,

$$\begin{aligned} F &= K (a \cos pt + b \cos qt)^2 \\ &= K (a^2 \cos^2 pt + b^2 \cos^2 qt + 2 ab \cos pt \cos qt) \\ &= K [a^2 \cos^2 pt + b^2 \cos^2 qt + ab \cos (p - q) t + ab \cos (p + q) t]. \end{aligned}$$

The force, therefore, has four components, only the third of which is useful in producing acoustic energy. Hence the useful force is given by

$$f = K a b \cos (p - q) t.$$

This shows that the greater the amplitude of the local oscillations, the greater is the useful force; but here again we should have back e. m. f.'s produced which, as soon as the device became efficient, would limit any further increase in the force due to a further increase in the local amplitude, by reducing the incoming amplitude proportionately. It is clear that the local oscillations perform the same function in this system that the permanent magnet does in the ordinary telephone receiver and that the constant impressed voltage does in the electrostatic telephone receiver. It is possible, however, that the local oscillations may give, besides, a limited amount of true amplification, as they do in the form discussed above. An investigation to determine whether or not this is the case would be difficult; the value, moreover, of such an investigation would be doubtful, since this form is not the most efficient and since it has been just shown that the maximum true amplification obtainable in the most efficient form of the heterodyne receiver is four.

SUMMARY: The necessity of viewing the energy relations in the heterodyne receiver from the standpoint of average, not of instantaneous values, is pointed out; and the average energy present due to two currents of different frequencies is studied.

A distinction is made between two general types of amplification; (1) by infusion of new energy into the received currents, and (2) by increase of efficiency of the receiving apparatus. Only the first may be regarded as true amplification. As examples typical of the second, the theory of the electromagnetic and of the electrostatic telephone receivers is sketched. Finally, it is shown that the maximum true heterodyne amplification is four.