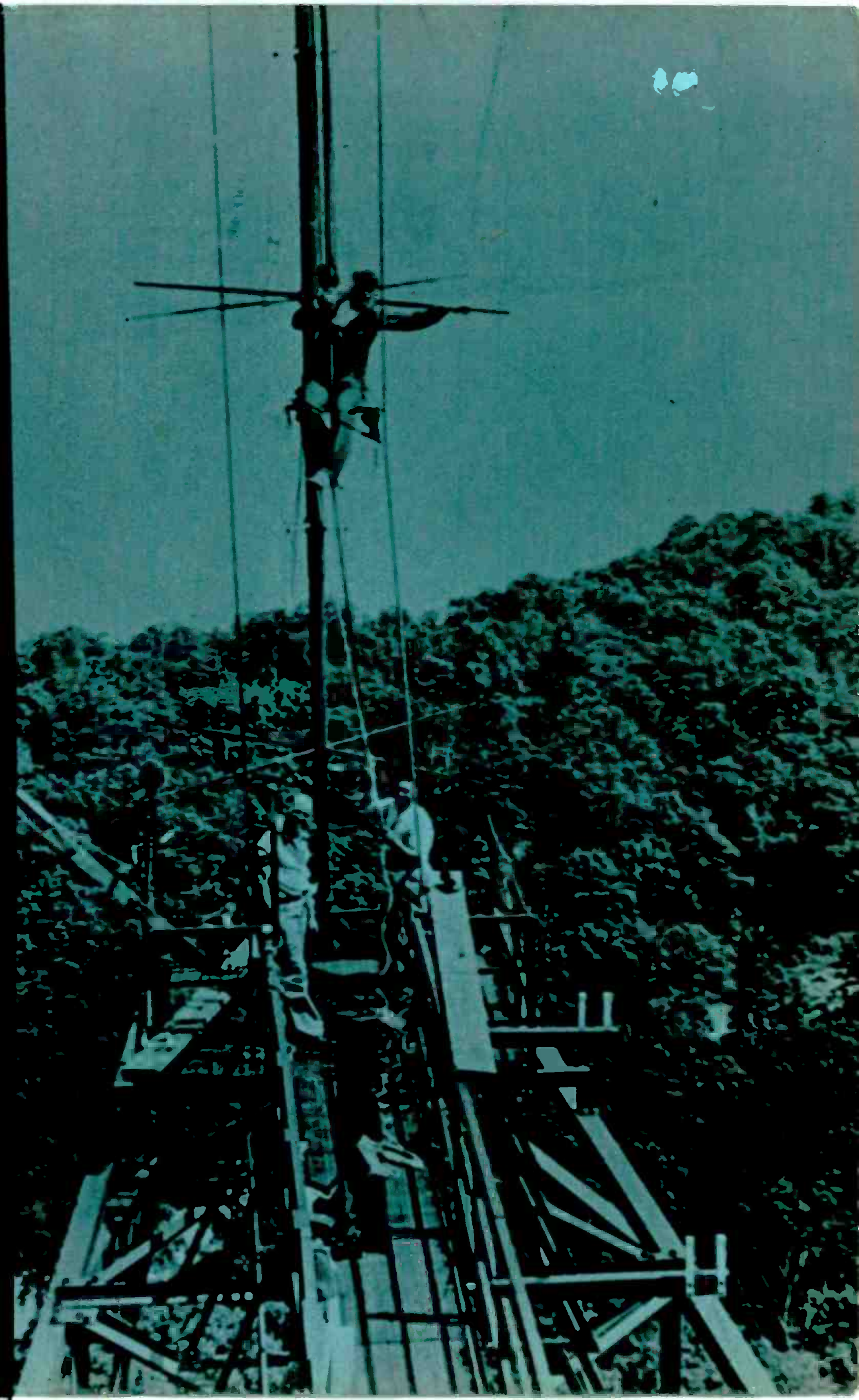


EDWIN H. ARMSTRONG

A compendium of five basic engineering papers from the **Proceedings of the Institute of Radio Engineers** reproduced in cooperation with the Smithsonian Institution for the Annual Conference of the Antique Wireless Association, Washington, D.C., September 22-24, 1972.

An exhibit of Armstrong's inventions may be seen in the Hall of Electricity, National Museum of History and Technology.

* * * *

Pictures courtesy Smithsonian Institution and Columbia University. Portrait by Fabian Bachrach. Monographs printed permission Institute of Electrical and Electronic Engineers. Printing of brochure under a grant from Harry W. Houck.

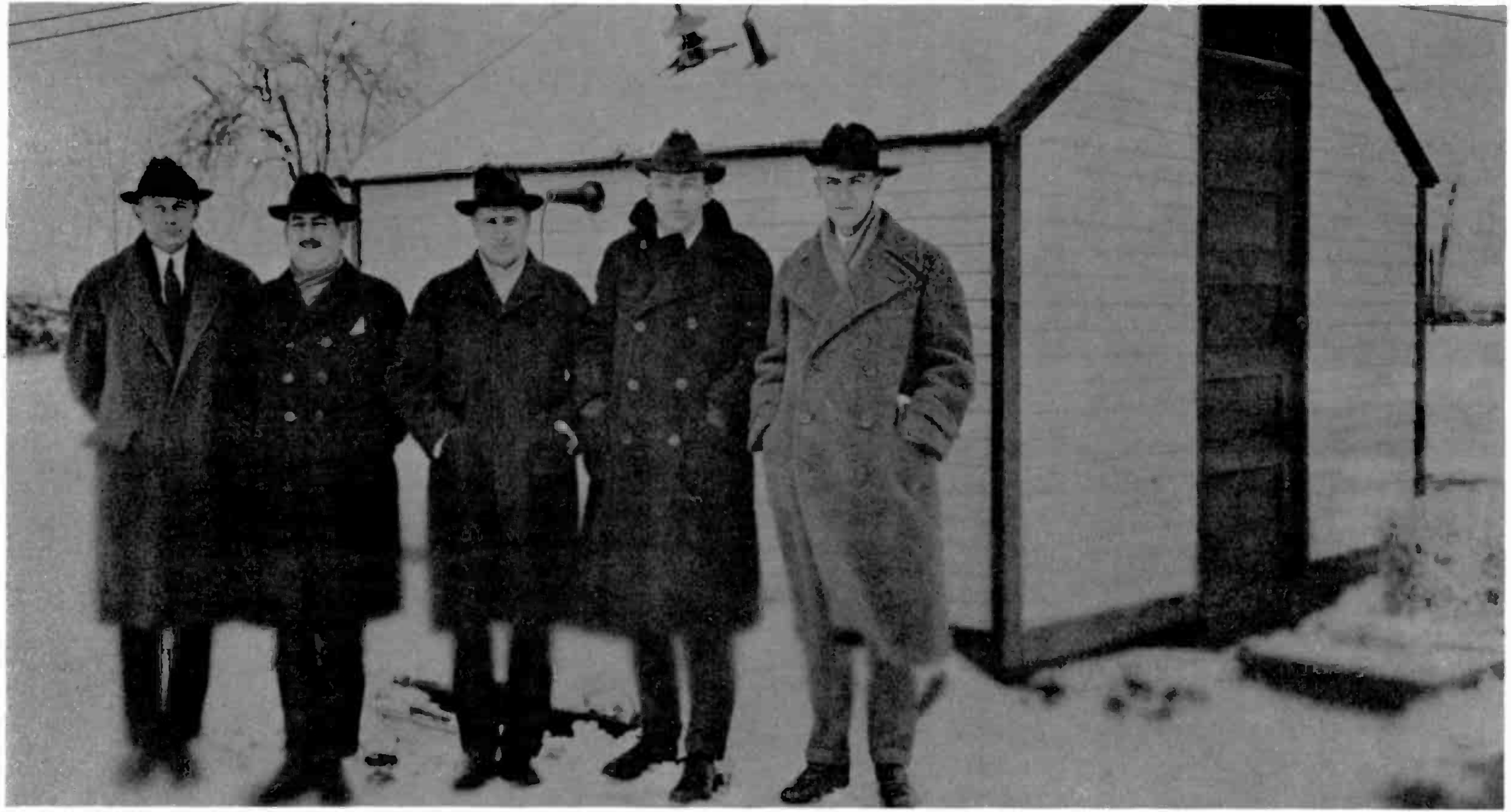
COVER: Armstrong and assistants atop Alpine, N.J. tower, ca. 1939

Division of Electricity & Nuclear Energy
SMITHSONIAN INSTITUTION
Washington, D.C.
1972

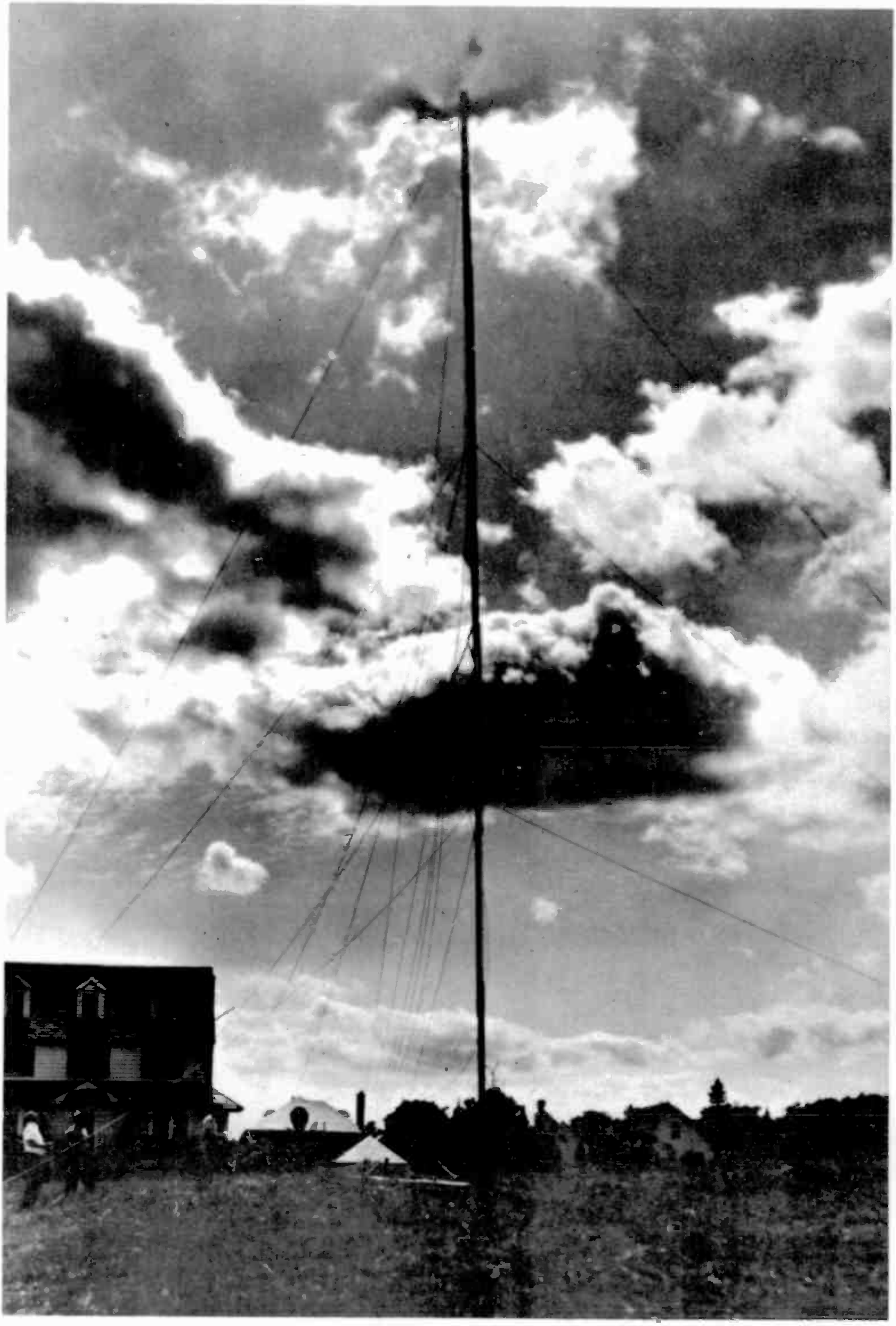
Of all contributors to the field of telecommunications Edwin Armstrong (1890-1954) perhaps is unique in fundamental contributions to science and technology. He graduated Columbia University in 1913, with electronics barely underway and the vacuum tube (or "audion") just past the stage of laboratory curiosity. Yet, in the period between the Wars, he was to take the budding radio industry through concepts and techniques that would have far reaching consequences not only for consumer products but for communication methodology and the science of information theory.

Four basic circuit innovations relate to Armstrong's work: the principles of regeneration, superregeneration, superheterodyne detection and frequency modulation. Although patent litigation was keen, and court battles not always won, the engineering fraternity overwhelmingly agreed that the inventor's instinctive understanding of circuit theory was crucial for the growth of radio.

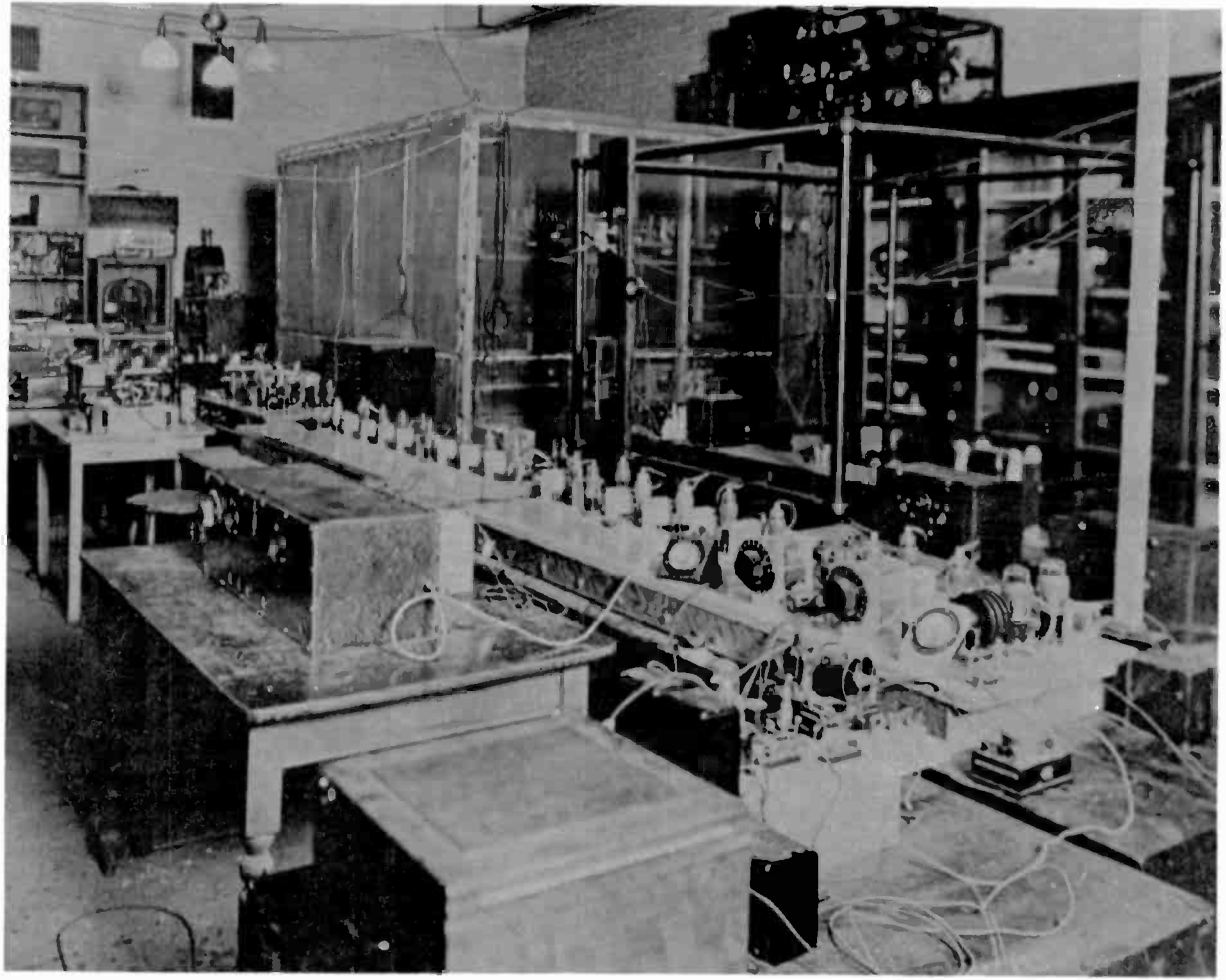
Portrait by Fabian Bachrach



Personnel at 1BCG, star station of the Transatlantic tests of 1921. Left to right: Amy, Grinan, Burghard, Armstrong and Cronkhite.



Sixty-five foot mast, Westhampton Beach, Long Island. Used in reception of signals from Empire State Building transmitter, 1933-34.



Experimental FM radio apparatus Columbia University, early 1930's

SOME RECENT DEVELOPMENTS IN THE AUDION RECEIVER[1]

By

EDWIN H. ARMSTRONG

THE AUDION AS DETECTOR AND AMPLIFIER

The fundamental operating characteristic of the audion is the relation between the wing current and the potential of the grid with respect to the filament—say the negative terminal of the filament. Such a characteristic is shown in Figure 1, and

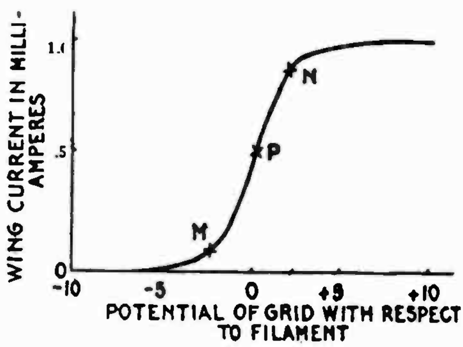


FIGURE 1

from it we see that a positive charge placed on the grid produces an increase in the wing current, and that a negative charge placed on the grid produces a decrease in the wing current. When the audion is used as an amplifier, and an alternating e. m. f. is impressed between the grid and the filament, the continuous current of the wing circuit will be varied in accordance

[1] Delivered before The Institute of Radio Engineers, New York, March 3, 1915, and before the Boston Section, April 29, 1915.

(The introductory material of this paper was originally submitted as a discussion by letter on Haraden Pratt's paper, "Long Range Reception with Combined Crystal Rectifier and Audion Amplifier." The first six figures have been kindly lent by the "Electrical World"; the remaining figures and text are herewith published for the first time.)

with the characteristic of Figure 1, producing on the continuous current a superimposed a. c. wave in phase with and of the same frequency as the impressed e. m. f. Diagrammatically this action is shown in Figure 2.

The action of the audion as a detector of radio frequency oscillations is very different from its action as a simple amplifier. Some form of connection must be used, such that the effect of a

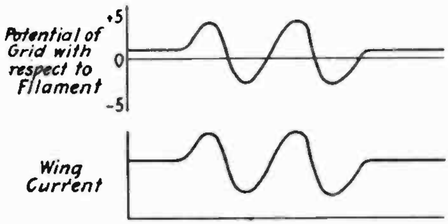


FIGURE 2

group of radio frequency oscillations in the grid circuit of the audion is translated into a single audio frequency variation of the current in the telephones. The usual method is to make use of the valve action between the hot and cold electrodes at low pressures, and the connection used to do this is shown in Figure 3. In this method of connection there are two distinct

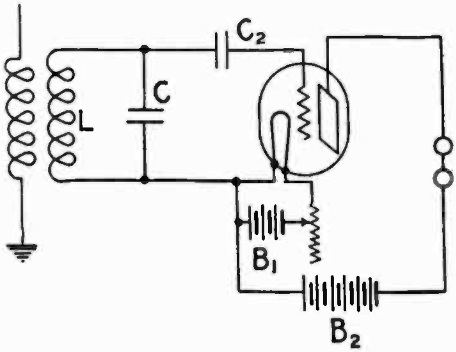


FIGURE 3

actions; one rectifying and the other amplifying. The closed oscillation circuit: LC, filament, grid, and condenser C_2, behaves exactly as a Fleming valve receiver, the incoming oscillations being rectified between the grid and filament and the rectified current being used to charge the condenser C_2 (the side connected to the grid being of course negative). The negatively charged grid then exerts a relay action on the wing current, decreasing it; the wing current returning to its normal value as the charge in the grid condenser leaks off by way of the grid and the grid resumes its normal potential. If the audion is properly constructed, the relay action results in an amplification of the energy available for use in the telephones over that which would be available in a simple rectifier. Figure 4 indicates the features of the valve method of detection.

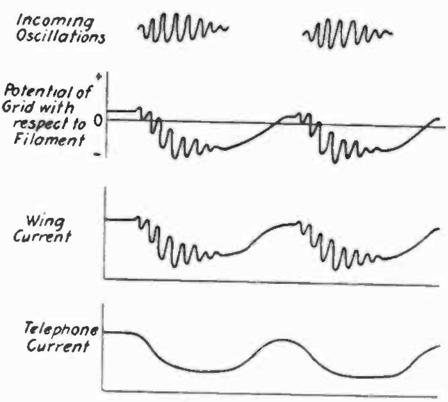


FIGURE 4

Working in conjunction with Professor Morecroft, I have recently secured oscillograms which confirm the explanations already advanced and these oscillograms and the means by which they were obtained are herewith shown in Figures 5, 6 and 7.

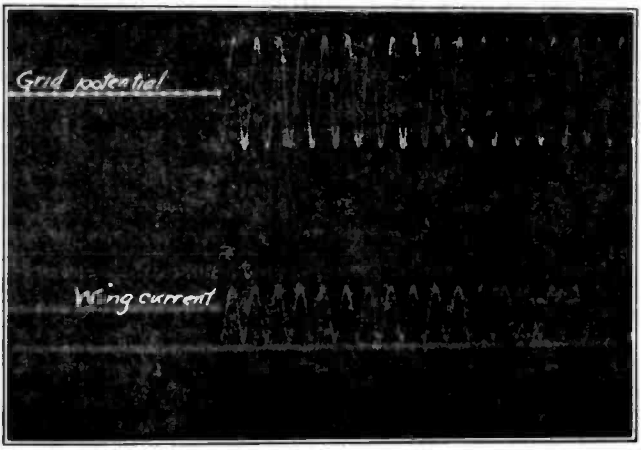


FIGURE 5

It will be seen, therefore, that using the audion as a detector of radio frequency oscillations, it has been shown that in addition to operating as a rectifier it simultaneously acts as a

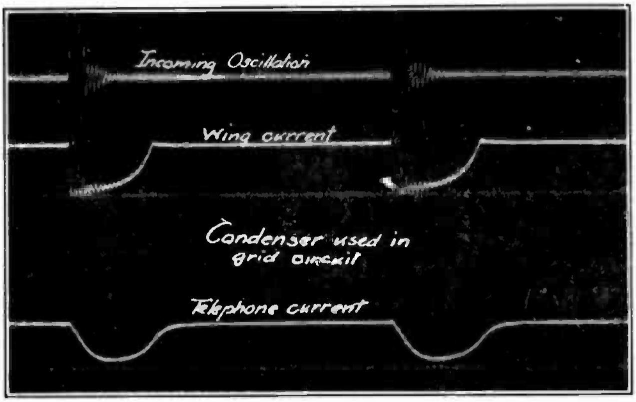


FIGURE 6

repeater of the radio frequencies; so that oscillations in the grid circuit set up oscillations of similar character in the wing circuit of the audion. In the ordinary detector system no use

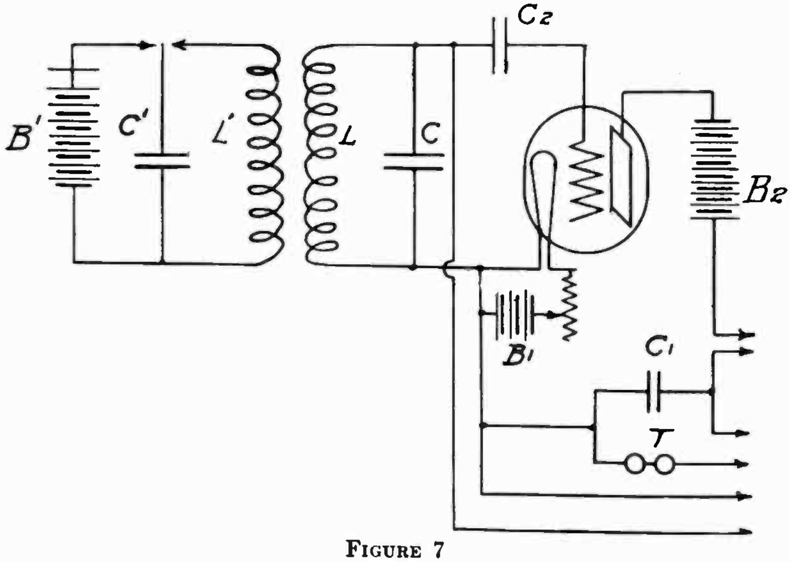


FIGURE 7

is made of the repeating action, and it is the purpose of the present paper to show that it may be turned to account to produce improvements in the reception of signals which completely overshadow any of the particular advantages of the audion when used as a simple detector. The ordinary detector circuit is illustrated by Figure 3 and the phenomena present therein may be summed up diagrammatically by the curves of Figure 4. It will be seen from these that the radio frequency oscillations present in the wing circuit of Figure 3 with the ordinary audion are necessarily small and also that they are of no value in producing a response in the telephones; but by providing means for increasing their amplitude and means for utilizing them to reinforce the oscillations of the grid circuit, it becomes possible to produce some very remarkable results.

REINFORCEMENT OF RADIO FREQUENCY OSCILLATIONS BY THE AUDION

There are two ways of reinforcing the oscillations of the grid circuit by means of those in the wing circuit. The simplest way perhaps is to couple the two circuits together in the manner shown in Figure 8. This is essentially the same as Figure 3, but

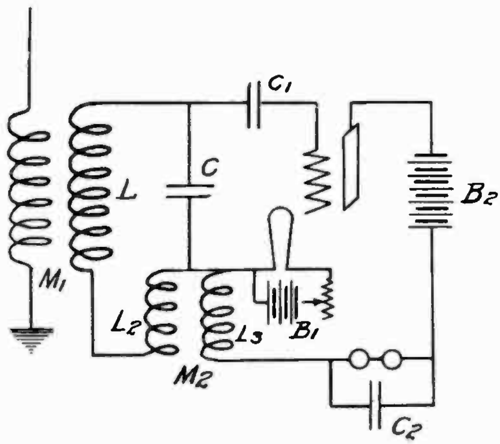


FIGURE 8

modified by the introduction of the inductively coupled coils L_2 and L_3 in the grid and wing circuits respectively and by the condenser C_2 which forms a path of low impedance across the telephones for the radio frequencies. In such a system, incoming signals set up oscillations in the grid circuit which repeat into the wing circuit producing variations in the continuous current, the energy of which is supplied by the battery B_2. By means of the coupling M_2, some of this energy of the wing oscillations is transferred back to the grid circuit, and the amplitude of the grid oscillations thereby increased. The amplified grid oscillations then react on the wing circuit by means of the grid to produce larger variations in the wing current, thus still further reinforcing the oscillations of the system. Simultaneously with this procedure the regular detecting action goes on; the condenser C_1 is charged in the usual way, but accumulates a charge which is proportional, not to the original signal strength but to the final amplitude of the oscillations in the grid circuit. The result is an increased response in the telephone proportional to the energy amplification of the original oscillations in the grid circuit. It will be observed from the operating characteristic (the relation between grid potential and wing current), that the amplitude of the variation in the wing current is directly dependent on the variation of the grid potential. This indicates that the grid circuit should be made up of large inductance and small capacity to obtain the maximum voltage which it is possible to impress on the grid. For moderate wave lengths the tuning condenser C of the grid circuit may be omitted altogether and the capacity of the audion alone used to tune the circuit. For long wave lengths, the distributed capacity of the grid circuit inductance becomes so high with respect to the capacity of the audion that better results are obtained by the use of a tuning condenser to fix definitely the points of maximum potential difference across the grid and filament of the audion.

In the second method of reinforcing the oscillations of the grid circuit the wing circuit of the audion is tuned by means of an inductance introduced as shown by Figure 9. This differs

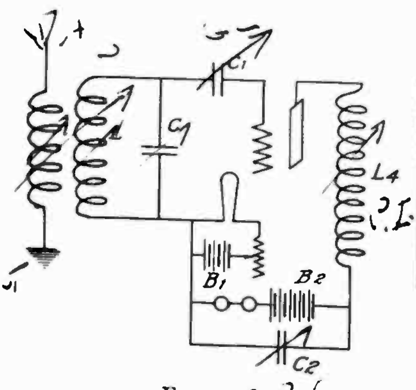


FIGURE 9

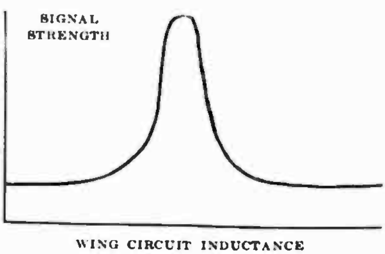


FIGURE 10

from the ordinary detector circuit of Figure 3 by the addition of the coil L_4 and the condenser C_2. The manner in which the grid oscillations are amplified may best be understood by the following analysis. With no oscillations in the system, the potential difference between filament and wing will be approximately the voltage of the battery B_2, but when oscillations are set up in the grid circuit, causing radio frequency variations of the wing current, the potential of the wing with respect to the filament varies as the reactance voltage of the wing inductance alternately adds to and subtracts from the voltage of the battery. When a negative capacity charge is placed on the grid, the wing current will be reduced and the direction of the reactance voltage of the wing inductance will therefore be the same as the voltage of the battery B_2. The reactance voltage will therefore add to the battery voltage and the difference of potential between wing and filament and also between wing and grid will be increased. Similarly when a positive charge is placed on the grid the wing current is increased and the reactance voltage of the wing inductance opposes the battery voltage, producing a decrease in the potential difference between grid and wing. Hence, supposing a negative capacity charge is placed on the grid, the tendency of the corresponding increase in the potential of the wing with respect to the grid will be to draw more electrons out on the grid, thereby increasing the charge in the condenser formed by the wing and grid, the energy for supplying this charge being drawn from the wing inductance as the wing current decreases. The increased negative charge on the grid tends to produce a still further decrease in the wing current and a further discharge of energy from the wing inductance into the grid circuit. On the other hand, when a positive charge is placed on the grid, the potential difference between grid and wing is reduced and some of the energy stored in the capacity formed by them is given back to the wing inductance. During this part of the cycle, electrons are being drawn into the grid from the surrounding space to charge the grid condenser in accordance with the well known valve action, and this, in effect, is a conduction current, so that a withdrawal of energy from the circuit takes place. In spite of this withdrawal of energy, however, a well defined resonance phenomena between the audion capacity and the wing inductance is to be expected and in the reception of signals such is found to be the case. When the wing inductance is properly adjusted at the resonance frequency, energy from the wing circuit is transferred freely to the grid circuit and oscillations build up therein and are rectified in the usual way.

A curve showing the general relation between signal strength and value of wing inductance is shown in Figure 10, the circuits used being those of Figure 9. As the capacity of the audion is the main means of transferring energy from the wing to the grid circuit, best results are obtained when the condenser C is very small. On account of the very small capacity of the audion, the effectiveness of this method of tuning is more pronounced at the higher frequencies, but by the use of a shunt condenser across the inductance of the wing circuit very good amplification is secured on frequencies as low as 30,000 cycles (10,000 meters wave length). The best results, however, are obtained with some combination of coupling and wing circuit tuning, as il-

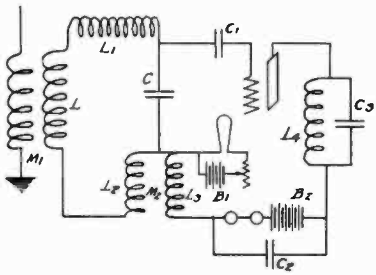


FIGURE 11

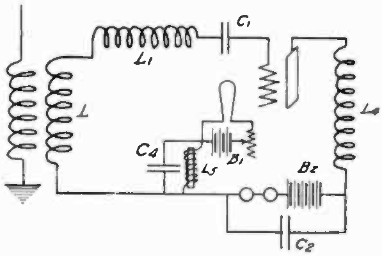


FIGURE 12

lustrated in Figure 11. Other methods of coupling may be employed between the grid and wing circuits, electrostatic and direct magnetic couplings being illustrated in Figures 12 and 13. The arrangement of Figure 13 operates in the same way as the system with the two coil coupling; but the electrostatic coupling

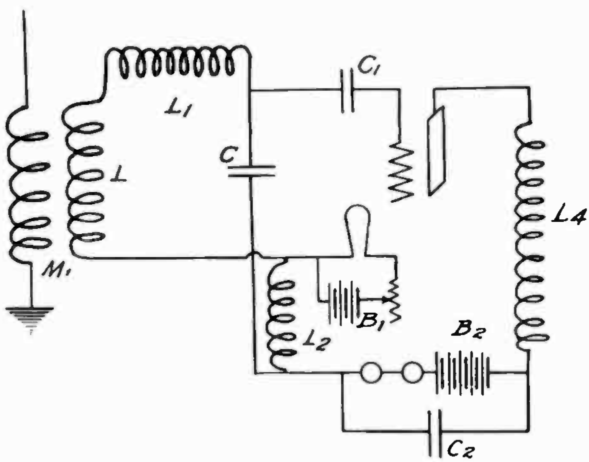


FIGURE 13

of Figure 12 works in an odd way. It is necessary, in this connection, to complete the wing circuit for the continuous current of the battery and this is done by shunting the coupling condenser C_4 by a coil of high inductance. The continuous current of the wing circuit flows thru this coil and C_4 provides a path of low impedance around this coil for the radio frequency oscillations of both grid and wing circuits. When a positive charge is placed on the grid, an increase in the wing current results, the alternating component of the wing current charging the condenser C_4 and the sum of the currents passing thru C_4 and L_4 equalling the current thru the audion. When a negative charge is placed on the grid the current thru the audion is reduced and the inductance L_5 discharged into the condenser shunted across it, charging it in the opposite way to that caused by the increase in the wing current. In both cases, C_4 then discharges thru the grid circuit reinforcing the oscillations therein.

AUDIO FREQUENCY AMPLIFICATION

It is possible to combine with any of these systems a system of audio frequency circuits which amplify the telephone current in exactly the same manner as the radio frequency oscillations are amplified, and such a system is shown in Figure 14. Here M_2

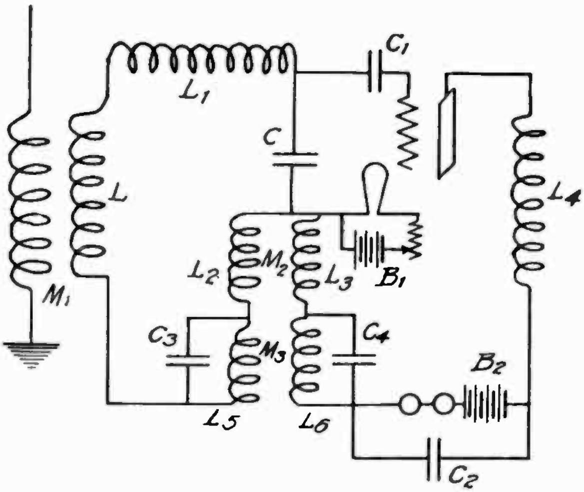


FIGURE 14

represents the coupling for the radio frequencies and the coils are of relatively small inductance. M_3 is the coupling for the audio frequencies, and the transformer is made up of coils having an inductance of the order of a henry or more. The condensers C_3 and

C_4 have the double purpose of tuning M_3 to the audio frequency, and of by-passing the radio frequencies. The total amplification of weak signals by this combination is about 100 times, with the ordinary audion bulb. On stronger signals, the amplification becomes smaller as the limit of the audion's response is reached.

THE AUDION AS A GENERATOR AND BEAT RECEIVER

Any repeater, which is also an energy amplifier, may be used to produce continuous oscillations by transferring part of the energy in the circuit containing the battery back to the controlling circuit to keep the latter continuously excited. By providing a close enough coupling between the grid and wing circuits, sufficient energy is supplied to the grid circuit to keep it in continuous oscillation, and as a consequence thereof oscillations of similar frequency exist in all parts of the system. The frequency of these oscillations is approximately that of the closed grid circuit if the tuning condenser of that circuit is large with respect to the capacity of the audion. If this capacity is small, then the wing circuit will exert a greater influence on the frequency of the system, and it will not approach that of the grid circuit so closely. When such a system of circuits is in oscillation, it has been found possible not only to receive continu-

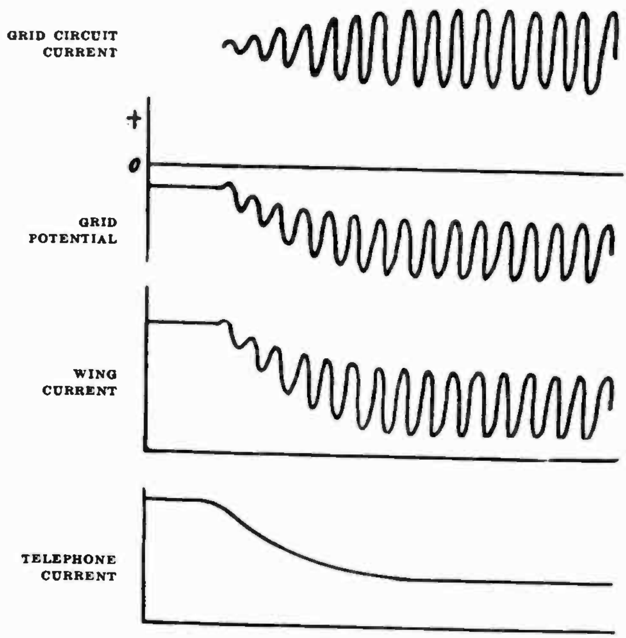


FIGURE 15

ous waves by means of the beat method but also very greatly to amplify them as well.

The phenomena involved may best be understood by reference to Figures 15 and 16, which show the relation between wing

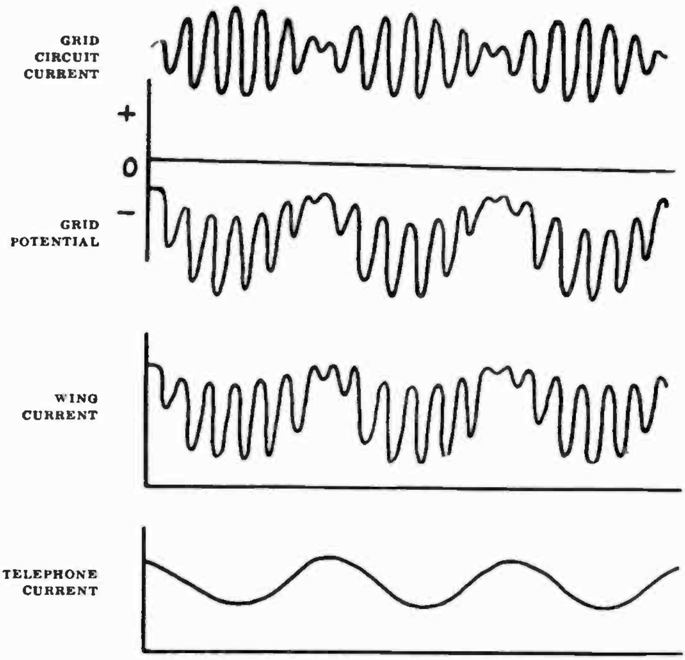


FIGURE 16

current and time at the beginning of oscillation. When the audion begins generating, the grid oscillations are continuously rectified to charge the grid condenser, and this charge continuously leaks off either by way of the grid or by means of a special high resistance placed in shunt with the condenser. As the negative charge builds up in the grid condenser, it decreases the average value of the continuous current component of the wing current and therefore limits the amplitude of the oscillations of the grid circuit until a point is finally reached where the rate at which electricity is supplied to the grid condenser is just equal to the rate at which it leaks off. Consider now the effect on the system of an incoming continuous wave having a frequency slightly different from the frequency of the local oscillations. The presence of the local oscillations will not in any way interfere with the amplifying powers of the system and the incoming oscillations will build up in exactly the same manner as for the non-oscillating state but to a greater degree because of the closer grid and wing coupling. Simultaneously with the amplifying of the incoming wave, beats are produced between the local and the signalling currents, the effect being alternately to increase and decrease the amplitude of the oscillations in the system. From Figure 15 it will be apparent that when this steady state is reached an increase in the amplitude of the grid oscillations by any means whatever will increase the negative charge in the grid condenser, producing a decrease in the average value of the wing current and hence a decrease in the telephone current. On the other hand, a decrease in the amplitude of the oscillations will allow some of the negative charge in the grid condenser to leak off and thereby permit an increase in the telephone current. Hence, when incoming and local oscillations add up, the negative charge in the grid condenser is increased and a decrease in the telephone current results. When the two frequencies are opposed, some of the charge in the grid condenser leaks off and an increase in the telephone current occurs. The result is the production in the telephones of an alternating current having a frequency equal to the difference in the frequencies of the local and incoming oscillations and having the very important property of being almost simple harmonic. Figure 16 illustrates the characteristics of this method of reception. The complete phenomena may be summed up as follows. Incoming oscillations are simultaneously amplified and combined in the system to produce beats with a local oscillation continuously maintained by the audion. The radio frequency beats are then rectified by the audion to charge the grid and the grid condenser, and this charge varies the electron current to produce an amplifying action on the current in the telephones.

When the grid condenser is omitted, the beat phenomenon is slightly modified, and the audio frequency variation of the telephone current is produced according to the asymmetric action outlined in a previous publication dealing with the operating features of the audion. The system is more sensitive with the grid condenser, but the same general result is obtained by either method of reception.

PECULIAR FEATURES OF OSCILLATION

Some very interesting features of operation accompany the production of oscillations in the system. Suppose the audion is not oscillating, and the grid and wing coupling is fairly weak. As this coupling is increased, the point at which oscillations begin is indicated by a faint click in the telephones accompanied by a slight change in the character of the static. The oscillations produced are usually so high in frequency and constant in amplitude that they are entirely inaudible. As the coupling is still further increased, a rough note is heard in the telephones the pitch decreasing with increase of coupling. This note is produced by the breaking up of the oscillations into groups, and it occurs whenever electricity is supplied to the grid condenser at a greater rate than that at which it can leak off. The result is that the grid is periodically charged to a negative potential sufficient to cut off entirely the wing current, causing a stoppage of the local oscillations until the grid charge leaks off and the wing current re-establishes itself. The frequency of this interruption depends largely on the capacity of the grid condenser, the resistance of its leakage path, and the amplitude of the local oscillations; and it may be varied from several hundred down to one or less per second. This effect is sometimes troublesome in the reception of signals, especially with high vacuum tubes. It may be eliminated, however, by increasing the leak of the grid condenser by means of a high resistance shunt. The best coupling for receiving continuous waves lies somewhere between the point at which oscillations start and the point at which interruption begins, and can only be determined by trial. In this region, trouble is sometimes experienced by the appearance of a smooth musical note in the telephones. This occurs under certain critical conditions of coupling with the antenna when the grid circuit oscillates with two degrees of freedom. Two slightly different frequencies are therefore set up, producing beats which are rectified by the audion in the usual way. This effect is quite critical, and when it causes

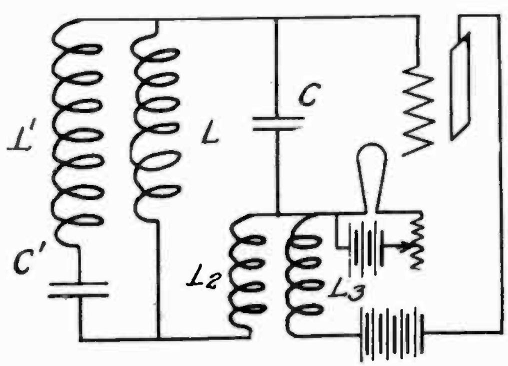


Figure 17

interference with signals, a slight readjustment of the circuits will usually make it disappear. It may, however, be made perfectly steady and reproduced at will by the system shown in Figure 17, where two grid circuits of different periods are provided. Two frequencies are therefore generated one having the frequency of the circuit LCL_2, and the other the frequency of the circuit $L'C'L_2C$. This arrangement may replace to advantage the ordinary buzzer for producing groups of oscillations. The foregoing explanations refer to the audion only when it is used as an electron relay.* When there is an appreciable amount of gas, in the tube in the ionized state, disturbances of an entirely different character occur.

AUDIO FREQUENCY TUNING

One of the very important advantages of the receiver when used for continuous waves is that the alternating current produced in the telephones is almost a pure sine wave. Only when the audio frequency is simple harmonic can selectivity be obtained by tuning the telephone circuit. A distorted wave such as that produced by spark signals possesses many harmonics and as each may be picked out by the tuned telephone circuit there is little chance of separating two spark signals by audio frequency tuning. With continuous waves, however, the pure wave produced by the beat method of reception makes it possible to obtain selectivity by the audio frequency tuning, resonance

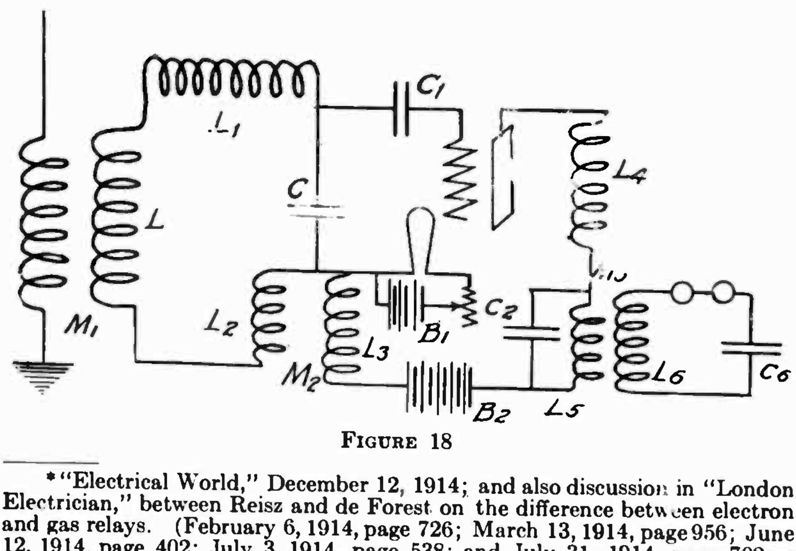


FIGURE 18

*"Electrical World," December 12, 1914; and also discussion in "London Electrician," between Reisz and de Forest on the difference between electron and gas relays. (February 6, 1914, page 726; March 13, 1914, page 956; June 12, 1914, page 402; July 3, 1914, page 538; and July 31, 1914, page 702.—EDITOR.)

being fully as sharp as in the radio frequency circuits. Two methods of audio frequency tuning are shown in Figures 18 and 19. In Figure 18, the telephone is inductively connected to the wing circuit of the audion by means of a transformer the secondary of which includes besides the telephone a tuning condenser.

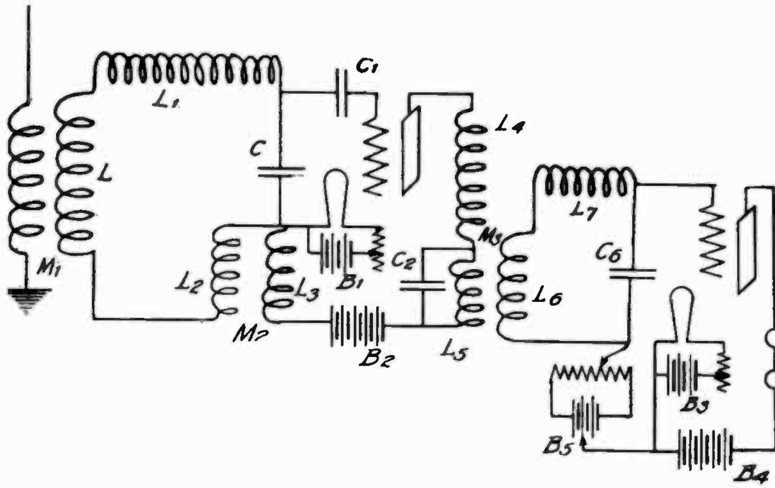


FIGURE 19

In this connection, the telephone, with a resistance of many thousand ohms, is placed directly in the tuned audio frequency circuit, and hence for good tuning the inductance of the coil L_6 must be made extremely large to secure the necessary ratio of the reactance of L_5 to the resistance of the circuit. This disadvantage is overcome in the system of Figure 19 by removing the telephones from the audio frequency circuit, and using the latter to operate a second audion. The telephones may then be placed in the wing circuit of this audion without adding appreciably to the damping of the circuit. The tuning of the circuit L_6C_6 may therefore be made very sharp with reasonable values of inductance simply by keeping the resistance low. In this case considerable amplification is obtained by the use of resonance in the transformer M_3 to increase the voltage impressed on the grid of the second audion. The great advantage of this kind of tuning is shown by the following example. Suppose the incoming signal has a frequency of 50,000 cycles, and the local frequency is 49,000 cycles. The differential frequency is 1,000, and the audio frequency circuit is tuned accordingly. An interfering wave 1 per cent. shorter than the signalling wave, or 49,500 cycles, will produce an audio frequency of 500 cycles per second, which will not appear at all in the wing circuit of the second audion unless it is many times stronger than the 1,000 cycle signal. This combination of radio and audio frequency tuning is too selective for use at the present time even when the sending station is equipped with an alternator, as the slight changes in frequency of the radiated wave produce changes in the beat frequency of the receiver which carry it out of range for the sharply tuned audio frequency circuit. A disadvantage of this method of tuning is that atmospheric disturbances produce a musical note due to shock excitation of the audio frequency system. Very loose coupling with the wing circuit of the first audion is a partial remedy for this. There are times, however, when interference is more troublesome than static and in such cases the method may be used to great advantages. If desired, both radio and audio frequency tuning can be carried out in the same audion as indicated in Figure 14. This combination is apt to be somewhat troublesome to operate as a cumulative amplification is obtained in the audio frequency as well as in the radio frequency system.

CASCADE SYSTEMS

Where a greater amplification than can be obtained with one audion is required, cascade working of the radio frequency systems may be resorted to by coupling together two or more audion systems, each connected as already described, in the manner indicated in Figure 19. The incoming oscillations in the first audion system are amplified in the usual manner and

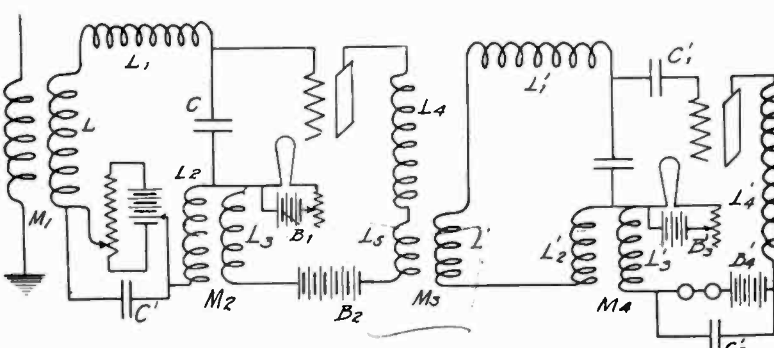


FIGURE 20

set up oscillations in the second system by means of the coupling M_3 (See Figure 20). The oscillations initially set up in the second system are again amplified, and then rectified in the second audion to produce audible response in the telephones.

For the reception of spark signals, considerable adjustment is required to get the best results without causing one or the other or both of the systems to generate oscillations. It will be found that after the first circuit is adjusted to the point of oscillation and the second is coupled with it, the strength of signal in the first system will be reduced owing to the withdrawal of energy from it by the second system. The signals may then be again brought up in strength by increasing the coupling between the grid and wing circuits of the first audion until the appearance of the local oscillations indicates that the limit of amplification has been reached. By careful adjustment about a thousand times amplification and very sharp tuning can be obtained with two steps.

For continuous wave reception, there are several methods of operating cascade systems. It is possible to have either system generate oscillations, the other system acting simply as an amplifier or both systems may be made to generate in synchronism. It will generally be found that when both systems produce oscillations, beats will be produced, so that a continuous note is heard in the telephones; but by adjusting the frequency of one of the systems the pitch of this note will be reduced as the two systems approach synchronism, until finally at one or two hundred beats per second the two systems pull into step in much the same way as two alternators. The ability of the two systems to keep in step depends mainly on the value of the coupling between them, and the closer this is the better the two hold together. There is still another way of working this combination, and that is asynchronously. In this case beats are continuously produced in the system so that a continuous note is heard in the telephone, but the circuits may be so adjusted that the note is not loud enough to be troublesome or it may be tuned out of the telephone in the manner previously described. Incoming oscillations are combined in the system to produce beats with the beats already present so that a rather curious note is heard. Very good amplification is secured by this method though naturally the system is troublesome to operate.

It may be noted here that whenever a signal is too weak to read with one audion system and cascade operation becomes necessary, it is always better practice to use the cascade circuits for the radio frequencies, even if the regenerative circuits are not employed with each individual audion system. The frequency of the oscillations set up in the circuits by static are, under normal conditions, the same as those of the incoming signal; and the static is therefore never amplified more than the signal. Usually it is amplified to a somewhat lesser extent, especially if regenerative circuits are employed. In the cascade systems used for audio frequencies, a different condition exists. It is ordinary practice to connect the different stages by means of transformers, and this leads to conditions which cause the system to produce greater amplifications of the higher frequencies. The rate of change of the wing current of the detecting audion produced by static corresponds to a very high frequency, and as such is invariably amplified to a greater extent than the signal.

There is a second method of receiving continuous oscillations which makes use of the generating feature of the audion, but does not employ the beat phenomena. The amplifying ratio of the audion depends more or less directly on the value of the wing current, and by varying this current periodically there will be a corresponding periodic change in the amplifying power of the audion. Hence an audion arranged to repeat a continuous wave under such conditions will produce in its wing circuit oscillations which vary periodically in amplitude, and which may therefore be received by a simple audion system. The first audion may be arranged to produce the necessary variation in its amplifying power in the manner indicated in Figure 21, which also

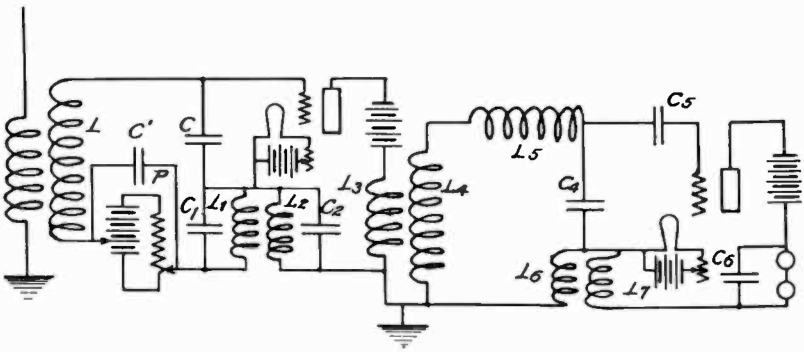


FIGURE 21

shows the complete circuits for carrying out this method of reception. Here $C_1L_1L_2C_2$ is an audio frequency system designed to produce audio frequency oscillations; and P is a potentiometer for adjusting the potential of the grid so that on the negative part of the oscillation in the wing circuit, the wing current is reduced practically to zero. The radio frequency circuit $C'L\ C\ C_1$ is tuned to the oscillation frequency of the incoming wave. The radio frequency oscillations cannot be detected in the first audion system as the strong audio frequency current circulating in this system would produce a continuous note in the telephone receivers of such strength as to render inaudible all save very strong signals. By arranging to detect the oscillations in a second audion system coupled to the wing circuit of the first, interference of this sort is avoided; as the circuit $L_4\ C_4$ has a very high impedance for the audio frequency currents and the effect produced thru the magnetic coupling of L_3 and L_4 on the second system is negligible. The capacity current between these two coils thru the telephones to ground is, however, appreciable; and to avoid it it is advisable to ground their two adjacent ends as shown. The action of the system may be summed up as follows. The first audion system varies the amplitude of the incoming radio frequency oscillations at an audio frequency, and the second audion system amplifies and detects the radio frequency oscillations supplied to it by the first system. Diagrammatically, the phenomena occurring are as illustrated in Figure 22. The system gives about the same

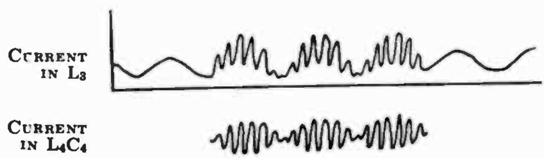


FIGURE 22

response as can be obtained with a single audion working with the beat method of reception. The advantages derived from the heterodyne method of amplification and the dependence of the audio frequency note in the receivers on the wave length are, of course, lacking; but for the reception of waves having a frequency higher than that at which beat reception is practicable, this method is of value.

EFFECTS OF ATMOSPHERIC DISTURBANCES

A very interesting feature of these receiving systems is their behavior under conditions of severe atmospheric disturbances, particularly when used for receiving continuous waves. Their success under such conditions is due to the fact that they combine in addition to their inherent property of responding more readily to a sustained wave than to a strongly damped one, the characteristics of the two most effective static eliminators known; the balanced valve and the heterodyne receiver. The function of the balanced valve is a physiological one, as it simply provides a means to shield the ear from the loud crashes which temporarily impair its sensitiveness for the relatively weak signals. In effect, it puts a limit on the noise which can be produced in the telephone by a stray, regardless of its amplitude. Now the effect of the static on an audion is to build up a negative charge on the grid, reducing the wing current, and the limit of the response which can be produced in the telephones is reached when the wing current is reduced to zero. Under ordinary conditions, this limit is too great to do much good; but when the audion is generating it is possible, by proper adjustment of the amplitude of the local oscillations, to reduce the wing current to a point just above the lower bend in the operating characteristic so that the audion is rendered insensitive to a further increase in the negative charge on the grid. The strays which cause serious interference are of a much greater amplitude than the local frequency, so that no appreciable interaction between the two takes place, and the wing current is invariably decreased. Since the decrease in the wing current is not in proportion to the change in the grid potential, the response in the telephone and the effect on the ear of the operator are correspondingly reduced. Static of smaller amplitude than the local oscillations may interact with them to produce either an increase or a decrease in amplitude of the oscillations in the grid circuit and may therefore cause either a decrease or an increase in the wing current. The wing current can, of course, increase to a relatively large value, but as it is impossible for the wing current to increase faster than the charge in the grid condenser can leak off, the rate of increase is necessarily slow. The response in the telephones is therefore not so disturbing as would be caused by a decrease of similar value where the rate of change of current is usually large.

When the system is operated without an auxiliary leak around the grid condenser, a peculiar paralysis of the audion is frequently caused by heavy static, no sound of any kind being heard in the telephones for a considerable length of time. If the apparatus is not touched, the paralysis may last for many minutes, and then suddenly disappear and the former sensitiveness be restored. The effect is primarily caused by the charging of the grid condenser to a sufficient potential to cut off entirely the flow of electrons to the wing, thereby decreasing the wing current to zero. Now the way in which the negative charge in the grid condenser leaks off is chiefly by means of the positive ions in the tube, which are drawn into contact with the grid when it becomes negatively charged. These positive ions are the result of ionization by impact, and when the voltage of the wing battery is properly adjusted, they can be produced only in the region between the grid and the wing, since the velocity attained by the electrons between the filament and grid is very low. When the grid is charged to a high negative potential it keeps all the electrons between the grid and filament, thereby barring them from the region between grid and wing. Hence the production of positive ions must cease and the usual means of removing the negative charge from the grid vanishes. The resistance of the leakage path of the grid condenser must then be almost infinite, as is shown by the very long time taken for the charge to leak from a condenser of approximately 0.0001 microfarads capacity. The effect is naturally the more pronounced the higher the vacuum, as the number of positive ions present is correspondingly reduced. A resistance of several hundred thousand ohms placed across the grid condenser gives a leak which is independent of the value of the wing current and which effectually prevents trouble of this kind. With the very high vacua now obtainable by the use of a molecular pump, there are practically no positive ions present so that the auxiliary leak is always necessary. Under these conditions, it not only prevents paralysis by the static but it also removes from the grid condenser the excess of negative electricity which accumulates in it, thereby increasing the sensitiveness of the audion and the sharpness of the signals in the telephones. The very high potentials to which the grid condenser may be charged by the static when it is not provided with an auxiliary leak are surprising. These potentials may be measured in a very simple and accurate way, here described. After a stray has cut off the wing current, if we continuously increase the capacity of the grid condenser the potential across it, and hence the potential of the grid, with respect to the filament, will be decreased inversely as the capacity. A point will finally be reached where the grid potential is sufficiently reduced to allow the wing current to flow. When this occurs it indicates that the potential of the grid condenser is slightly less than that shown by the operating characteristic as necessary to reduce the wing current to zero. The potential to which the grid condenser was originally charged is equal to th's voltage times the ratio of the capacity of the condenser at which the wing current began to flow to the original capacity. Voltages of over a hundred are not uncommonly reached by the grid; and as one volt represents a very strong signal, the difficulties of the static problem are very forcibly presented.

The fact that static of large amplitude produces almost invariably a decrease in the wing current while a signal (with beat reception) produces alternately an increase and decrease in the wing current is a circumstance of which it should be possible to take advantage. The circuits can be arranged to rectify the wing current in such a way that only the increases in this current are available to produce a response in the telephones, but in carrying this method out, trouble is experienced from a shifting zero. A better way of making use of the difference in response is the following one. Suppose that we arrange two complete receiving systems oscillating in step with each other, but so related to the antenna that the beat currents in the two systems are 180 degrees apart. The result of this will be that at the instant when the incoming signal is producing an increase of current thru the telephones in one receiver, it will be producing a decrease of current thru the telephones of the other receiver; so that the two telephone currents are 180 degrees out of phase. Static of large amplitude does not interact with the local frequencies, and will produce simultaneously in each receiver a decrease in the telephone current. These two currents are therefore in phase with each other. On replacing each telephone by the primary of a transformer, and connecting their secondaries thru a telephone in the proper phase, it is possible to balance out the static and at the same time secure an additive response of the signals from each receiver.

An arrangement of circuits by means of which this method can be carried out is shown in Figure 23. Here two oscillating receiving systems are kept in step by means of the circuits L_1 C_1 C_1' L_1'. L_1 C_1 and L_1' C_1' are identical, and each is tuned separately to the frequency to be received. When both audions are oscillating in step, the flow of current in these circuits as indicated by the vectors of Figure 23 will be alternately up on one side and down on the other. The point between the condenser C_1 and C_1' will be a node; and the antenna may be connected to this point without disturbing the conditions appreciably if a resistance R placed as indicated is included in the antenna. This resistance need not be large enough to interfere seriously with the signal strength; it need only be large with respect to the resistance of the circuit L_1 C_1 C_1' L_1', which circuit has a very low resistance.

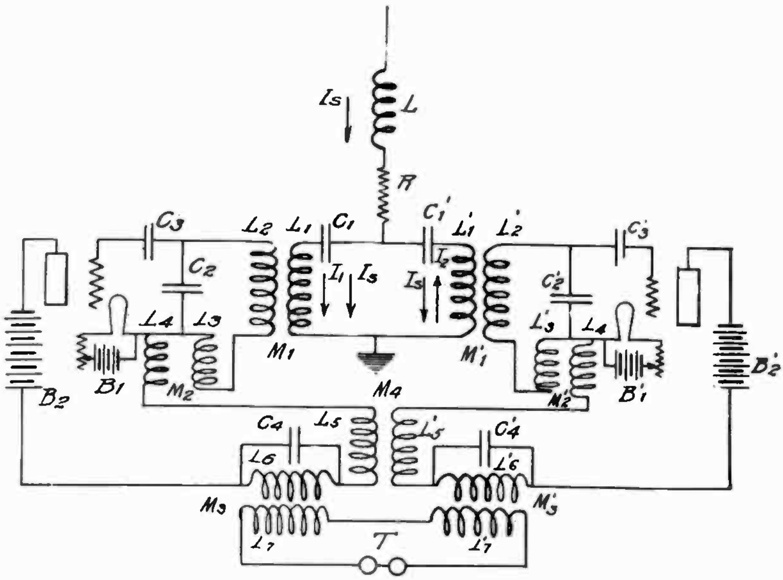


FIGURE 23

Incoming oscillations pass thru the divided circuit as indicated in the diagram, and therefore are in phase with the local oscillations of one receiver and 180 degrees out of phase with the local oscillations of the other. This produces the desired result in the currents thru the transformers of the circuit T which act in the manner already described.

It is found in practice that the oscillations set up in each system by the incoming signals tend to neutralize each other thru the circuit L_1 C_1 C_1' L_1'. This effect is avoided by introducing in the wing circuits a differential coupling arranged to neutralize the coupling between the two grid circuits. It is possible to do this, as it does not affect the coupling of either receiver with the antenna, and does not interfere with the local operation until the effective coupling between the two systems is reduced to a point below which they will no longer remain in step. There are other ways of securing the same result, but the system shown will illustrate the general procedure in carrying out this method of balancing.

The practical results obtainable with these receivers may perhaps be of interest. At the present time, signals from all high power stations from Eilvese (Germany) to Honolulu are heard day and night at Columbia University with a single audion receiver. Cascade systems give correspondingly better results, two stages being sufficient to make the night signals of Honolulu audible thruout the operating room. Interference with the signals from Nauen by the arc station at Newcastle, New Brunswick (Canada), is very easily eliminated by means of an audio frequency tuning circuit; and this is the most severe interference we have yet experienced, the two frequencies sometimes differing by less than 1 per cent. and the arc signals being much the stronger.

These receivers have been developed in the Research Laboratory of Electro-Mechanics, Columbia University; and are mainly the result of a proper understanding and interpretation of the key to the action of the audion; the grid potential-wing current curve. In conclusion, I want to point out that none of the methods of producing amplification or oscillation depend on a critical gas action; they depend solely on the relay action of the tube employed (electron or gas relay) and the proper arrangement of its controlling circuits.

SUMMARY: The action of the audion as a detector and simple amplifier is explained, with the method of verification of the theory by means of oscillograms. To reinforce the oscillations in the grid circuit two methods are employed: first, to couple the grid circuit to the wing circuit and arrange the latter to permit radio frequency currents to pass freely in it; and second, to use a large inductance in the wing circuit, thereby tuning it to the incoming frequency (in conjunction with the capacity between the filament and wing in the audion itself). Both methods may be used together. Various methods of coupling grid and wing circuits are shown. Methods of combined audio and radio frequency amplification are described.

The audion, being a generator of alternating current of any desired frequency, can be used as a beat receiver. A steady audion generator of regular groups of radio frequency oscillations is illustrated. Various methods of audio frequency tuning permitting high selectivity are possible. By the use of two audions in cascade, amplifications as high as 1,000 are attainable. The cascade systems can be arranged so as to operate both audions either synchronously or non-synchronously.

As an alternative to beat reception of sustained wave signals, an arrangement is explained wherein the amplifying ratio of a repeating audion is varied periodically at an audio frequency. Coupled to this system is a simple audion detector. Musical signals of any desired pitch are thus obtained.

It is found that static of large amplitude nearly always decreases the wing current, while a signal (with beat reception) alternately increases and decreases it. A system of circuits is described whereby this fact is taken advantage of in balancing out static while retaining an additive response to signals, thus effecting an elimination of static to a considerable extent.

Finally, instances of long distance stations received and interference overcome in practice are given.

DISCUSSION

Lee de Forest (by letter): Absence from New York and stress of business prevents my giving to Mr. Armstrong's paper the thoro discussion it merits from me.

Briefly, I must state that my investigation of the simple audion detector, the audion amplifier, and the "ultraudion" detector for undamped waves do not bear out completely the results and conclusions announced by that writer.

In the first place, anyone who has had considerable experience with numerous audion bulbs must admit that the behavior of different bulbs varies in many particulars, and to an astonishing degree. The wing potential-wing current curves for different bulbs, or even for the same bulb at different times, under differing conditions (filament temperature, etc.) vary widely.

What may appear to be a fixed law for one bulb may not hold for another.

Mr. Armstrong makes no mention of this well-known fact; nor does he even state that his grid potential-wing current curve may be quite otherwise than he has shown it with different applied "B" battery voltage, or filament temperatures.

He makes no mention of the fact, often demonstrated, that a continuous current indicating instrument, e.g., a micro-ammeter, may show a decrease in deflection, or practically no change in deflection either way when fairly strong radio frequency (or audio frequency) impulses are delivered to the grid even when the telephone receiver in the wing circuit gives strong response.

I have frequently proven that a *positive* charge applied to the grid, may decrease, rather than increase the "wing current." If I may say so, he treats the entire subject in much too cursory and cavalier a manner, even as he appears to be quite oblivious of the work of any other investigator or discoverer.

As I stated in an article in the "*Electrical World*," February 20th, the *oscillating* quality of the audion was discovered by me several years ago.

I found that the complicated circuits Mr. Armstrong illustrates were quite unnecessary for producing the effects mentioned. In fact, the combination of oscillating and amplifying functions in the same bulb are obtained almost, if not quite, as efficiently, and far more simply by much simpler circuits.

The second method he shows for a combination tuning to radio and audio frequencies is ingenious and highly creditable. Unfortunately, as he truly points out, there is to-day no continuous wave generator of sufficiently constant frequency to permit full advantage being taken of this elegant method.

Edwin H. Armstrong: The condition in which a positive potential applied to the grid produces a decrease in the wing current is a remarkable one, in that it has been the cause of that mysterious atmosphere with which the audion has long been surrounded. The effect occurs under certain conditions which are very easily explained. Suppose there is an appreciable amount of gas in the tube and the difference of potential between the wing and filament is adjusted so that a considerable number of positive ions are produced. In such a state it frequently happens that the number of positive ions coming in contact with the grid is in excess of the number of negative ions. As a consequence of this the grid assumes a positive charge with respect to the filament. Suppose the potential to which the grid becomes charged is three volts positive with respect to the negative terminal of the filament. Under these conditions a battery of say one or two volts connected as shown in Figure 1 with its

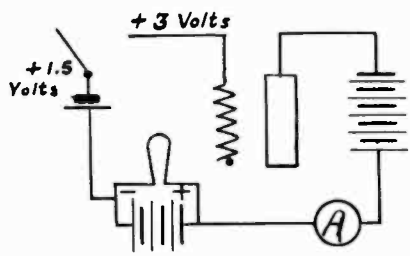


FIGURE 1

positive terminal connected to the grid will really change the potential of the grid in the negative sense. The natural result is a decrease in the wing current. The converse of this effect: the condition in which a negative potential applied to the grid produces an increase in the wing current, is invariably met with in high vacuum audions where the potential assumed by the grid is invariably negative. Both cases, however, can be explained on the same grounds. Figure 2 shows the potential assumed by the grid when a large number of positive ions are present.

Edwin H. Armstrong (by letter): In replying to Dr. de Forest's communication, I want to point out that the paper was intended to deal with the application of circuits of a new type to the actuation of the audion. The fundamental operating features of the audion itself were outlined purely as a basis on which to explain the action of the circuits. A detailed explanation of the various phenomena involved in the audion as a detector and as a relay, radically different from that previously advanced by Dr. de Forest, was published by me some time ago in the "*Electrical*

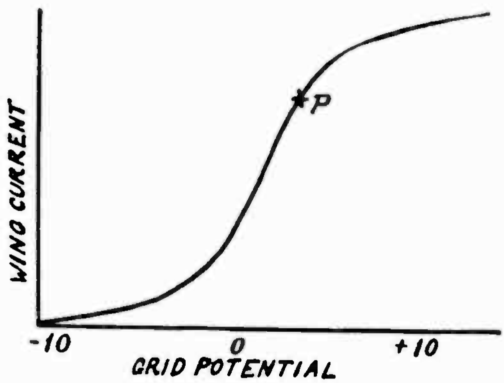


FIGURE 2

World," December 12th, 1914, and the columns of that paper, are, no doubt, still open to discussion of these elementary matters.

Dr. de Forest speaks of the great differences existing between the wing potential-wing current curves. It will be readily understood by those familiar with the laws of the conduction of electricity thru gases that such is bound to be the case where any considerable amount of gas is present in the bulb. The potential at which progressive ionization of the gas begins, is dependent, among other things, on the pressure; and hence the upper parts of the wing potential-wing current curves vary, but the lower parts, *the only place where the electron relay can be operated*, are invariably of the same general shape. With the modern methods now available, for producing very high vacua, it is a simple matter to construct audions whose characteristics are for all practical purposes identical. With these high vacuum bulbs, the astonishing differences of which Dr. de Forest speaks disappear to an astonishing extent.

The great differences which sometimes exist between the grid potential-wing current curves of different audions or for the same audion under different conditions of wing potential or filament temperature are again due to the residual gas, and are eliminated as before by the use of very high vacua. It will be evident, of course, that for each value of wing potential and filament temperature there will be a different grid potential-wing current curve; but for high vacuum bulbs these curves lie one above the other in an orderly manner and, barring minor differences, are of the same general shape.

For an explanation of the fact that a continuous current instrument in the wing circuit shows no change in deflection when an alternating e. m. f. of *audio* frequency is impressed on the grid even when a telephone in circuit with the meter gives a strong response, I want to call attention to Figures 2 and 5, of the original paper, together with a suggestion that a telephone perhaps is apt to respond somewhat more strongly to an alternating current than does a continuous current instrument! An explanation of the decrease of wing current which may occur will be found in the publication in the "*Electrical World,*" December 12th, 1914, with an accompanying oscillogram which shows the asymmetric effect in question. The circumstance stated by Dr. de Forest in which a *radio* frequency e. m. f. impressed on the grid produces a response in a telephone but not in a continuous current instrument is an impossible one. If the telephone responded, and there were no changes in the reading of the instrument, it would be an indication of an alternate and equal increase and decrease of the wing current at an audio frequency rate. This is an effect which *radio* frequency oscillations applied to the grid cannot produce. When a condenser is used in connection with the grid, radio frequency oscillations invariably produce a net decrease in the wing current and hence a decrease in the telephone current. Where use is made of the asymmetric relaying, which is possible because of the bends in the operating characteristic, either a net increase or net decrease may be produced in the wing current by radio frequencies applied to the grid, depending at which bend the audion is worked, but an increase and decrease can never be produced at the same time.

Dr. de Forest attempts to throw doubt on the validity of the operating characteristic, and hence on all explanations depending thereon, by stating that he has frequently proven that a positive charge applied to the grid may decrease rather than increase the wing current, a contention originally advanced by him in explanation of the relay and detecting action of the audion. In the discussion, I have pointed out the fallacy in this view and explained the seeming paradox which is found in low vacuum bulbs on the working part of the grid potential-wing current curve. There is another effect which may lead to incorrect conclusions concerning the action of the electron relay, which is due to effects found above the working part of the curve. As the potential of the grid is increased, it is possible that the wing current may reach a maximum and then fall off. This is due to the fact that a conduction current flows to the grid when it is positive with respect to the filament, and that under certain conditions, this current is subtracted from the wing current. The maximum current which can flow from filament to wing is limited to the number of electrons emitted by the filament, and if the condition of maximum current flow in the wing circuit is established before the grid potential becomes highly positive, then a further increase in the grid potential will increase the number of electrons absorbed by the grid and the result is a decrease in the wing current. The impossibility of working an electron relay on this part of the curve will be evident from the accompanying diagrams (Figure 3) which show how the effective resistance of the input side of the audion increases as

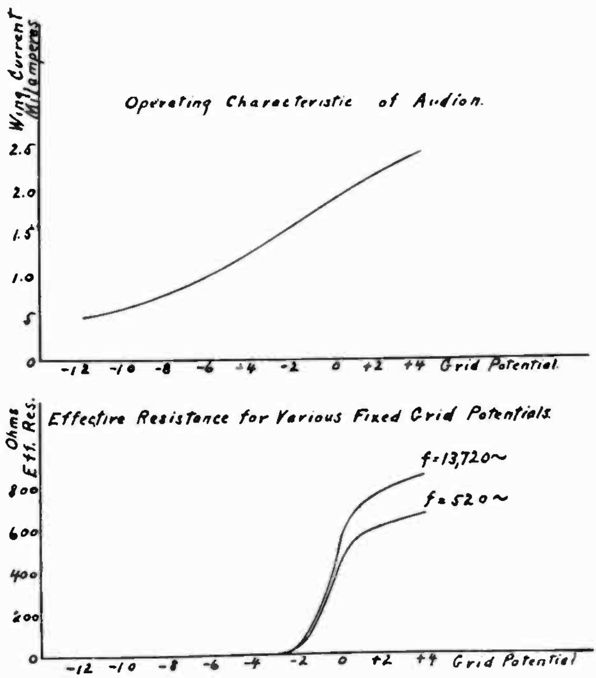


Figure 3

the potential of the grid is varied. Only when the grid is negative with respect to the filament can the full amplifying power of the audion be realized, as the input side consumes no energy. Herein lies the explanation of the great differences which exist in the amplifying powers of different bulbs when used in the customary fashion. It is usual to operate the audio frequency amplifier with the grid insulated from the filament for conduction currents so that the potential of the grid is determined solely by the characteristics of the audion. If it should chance to be sufficiently negative, the input side consumes no energy and the result is a good one; if it should be positive, then the input side consumes energy and the amplification is seriously impaired, the degree depending on the value of the positive charge. All this is clearly supported by the fact that when the potential of the grid of a good bulb is arbitrarily made positive, the amplification falls off. The curves shown in Figure 3 are additional confirmation, and in this connection it is interesting to note the agreement between the radio and audio frequency curves.

The statement by Dr. de Forest that he originally discovered the oscillating phenomenon and applied it to producing the effects described several years ago cannot be here discussed, because his priority in this matter will be contested shortly in another way.

Lee de Forest (by letter): While I cannot accept Mr. Armstrong's explanations of my observations as to the action of a positively charged grid on the wing current, they have at least more to recommend them than has his previous flat contradiction that such an effect as I have described existed at all.

What Mr. Armstrong states are "elementary matters" have not appeared so to associates and collaborators of Drs. Rutherford and Soddy with whom I have discussed them. These discussions, however, were prior to the appearance of Mr. Armstrong's paper.

In spite of Mr. Armstrong's explanations, we are left quite in the dark as to how high these consistent vacua are, and just what operating voltages he refers to. More quantitative explicitness and citations of the exact performances of scores of bulb would be more convincing than are the theories proposed as a basis for description of sundry complicated circuits.

If he is dealing with a type of tube which is quite distinct from the audion (on account of the degree of vacuum, the applied potentials, etc.), this should have been explicitly stated at the outset. This is my chief complaint. No essential data are given, but only general laws with attempted axioms. I assumed that we were dealing with phenomena in the audion as popularly known, operating on from 20 to 50 volts. With such, at least, there still remain some unexplained problems.

If he be unable to explain my observation that, using audio frequencies, certain bulbs show a decrease, others no perceptible change in deflection of a direct current micro-ammeter while a telephone receiver gives responses many times audibility—this fact should be frankly stated. I should also like to have his explanation as to why certain audions are distinctly more sensitive to low than to high spark frequencies while others show the exact reverse. Tho I have theories on this point, I have not yet proven them.

In connection with Mr. Armstrong's insistence on the value of his oscillograms which were taken at audio frequencies because audio and radio frequency phenomena are identical in nature, I should like to call attention to his statement that "This is an effect which radio (as distinguished from audio) frequency oscillations applied to the grid cannot produce."

Is it not perhaps possible that where successive strongly damped wave trains, of radio frequencies, have alternately positive and negative initial wave fronts, an alternating increase and decrease of wing current may occur which would, while giving loud signals in the telephone receiver, produce practically no change in deflection in a direct current micro-ammeter in series therewith?

As to Mr. Armstrong's closing remark, I had not before realized that he actually claimed broadly the discovery of the oscillating property of the audion. I think it can and will be established that this was discovered some time before his first work in this field. If any are still of the opinion that the oscillating quality of the audion awaited the discovery of the complicated circuits he describes, I would refer them to the article on "The Double Audion Type of Receiver," by Professor A. H. Taylor, in the "*Electrical World*" of March 13th, 1915.

Edwin H. Armstrong (by letter): Replying to Dr. de Forest's latest communication in regard to the effect of a charged grid on the wing current, I cannot but assume, from his failure to produce evidence to the contrary, that his observations may be explained by the residual positive charge on the grid. This applies to that type of tube in which so many "unexplained" phenomena are observed; "the audion, as popularly known, operating on from 20 to 50 volts."

Dr. de Forest's misapprehension as to the type of tube referred to in the paper rests entirely with himself. It was definitely stated in the article in the "*Electrical World*," and on the occasion of the presentation of this paper before the Institute of Radio Engineers that the vacuum of the bulbs was such that only thermionic currents existed. The methods used to obtain these vacua were those recently described by Dr. Irving Langmuir in a paper presented before the American Physical Society, and also in another paper presented before the Institute of Radio Engineers (See this issue of the PROCEEDINGS, together with the discussion on Dr. Langmuir's paper).

In explanation of Dr. de Forest's observation that audio frequencies applied to the grid may produce either a decrease or no change in the reading of a *direct* current micro-ammeter, while a telephone responds strongly, I have pointed out the oscillograms which fully explain both cases. It seems necessary to add that a *direct* current instrument of the type mentioned measures *average* values!

The question of the relative sensitiveness of an audion as a detector to high and low spark frequencies is entirely irrelevant to the present discussion. It has, however, some points which are of interest. The effect occurs only when the valve action of the audion is used to rectify the oscillations and a condenser is necessarily used in series with the grid. When there is a scarcity of positive ions, the rate of leak of the charge accumulated in the grid condenser from one group of oscillations may be so slow that the condenser fails to clear itself before the arrival of another group of oscillations. Under these conditions, a residual negative charge is continuously maintained in the grid condenser during the periods of signaling, and this charge interferes with the rectifying action between grid and filament. Obviously, this effect will be more pronounced at the higher spark frequencies, and the sensitiveness of the audion will be less impaired on the low spark frequencies. The phenomenon is an interesting one, but on the whole it is quite simple and elementary in character.

Dr. de Forest attempts to explain the circumstance which I have shown is impossible—the circumstance in which radio frequencies applied to the grid produce response in a telephone in the wing circuit but no change in the deflection of a continuous current instrument in series with the telephone. The explanation advanced is impossible. The effect described could be produced only by wave trains that were practically aperiodic. Needless to say, nothing remotely approaching this is in use in radio telegraphy at the present time.

In conclusion, I wish to point out that this discussion was originally begun by Dr. de Forest in an attempt to invalidate the explanations advanced to account for the various detecting, repeating, and oscillating phenomena. It is my opinion that the explanations given stand as correct.

Robert H. Marriott: It has been frequently charged that there has been a lack of research in radio engineering carried out in physical research laboratories. Mr. Armstrong deserves much praise in carrying out his highly interesting investigation, and it is to be hoped that further valuable results will be obtained under similar auspices.

(This discussion is herewith closed.—EDITOR.)

SOME RECENT DEVELOPMENTS OF REGENERATIVE CIRCUITS*

By

EDWIN H. ARMSTRONG

(MARCELLUS HARTLEY RESEARCH LABORATORY, COLUMBIA UNIVERSITY, NEW YORK)

It is the purpose of this paper to describe a method of amplification which is based fundamentally on regeneration, but which involves the application of a principle and the attainment of a result which it is believed is new. This new result is obtained by the extension of regeneration into a field which lies beyond that hitherto considered its theoretical limit, and the process of amplification is therefore termed *super-regeneration.*

Before proceeding with a description of this method it is in order to consider a few fundamental facts about regenerative circuits. It is well known that the effect of regeneration (that is, the supplying of energy to a circuit to reinforce the oscillations existing therein) is equivalent to introducing a negative resistance reaction in the circuit, which neutralizes positive resistance reaction, and thereby reduces the effective resistance of the circuit. There are three conceivable relations between the negative and positive resistances: namely—the negative resistance introduced may be less than the positive resistance, it may be equal to the positive resistance, or it may be greater than the positive resistance of the circuit.

We will consider what occurs in a regenerative circuit containing inductance and capacity when an alternating electromotive force of the resonant frequency is suddenly impressed for each of the three cases. In the first case (when the negative resistance is less than the positive), the free and forced oscillations have a maximum amplitude equal to the impressed electomotive force over the effective resistance, and the free oscillation has a damping determined by this effective resistance. The steady state is attained after the initial free oscillation dies out and continues until the impressed emf. is removed, when the current dies out

*Received by the Editor, June 3, 1922. Presented before THE INSTITUTE OF RADIO ENGINEERS, New York, June 7, 1922.

in accordance with a second free oscillation. The maximum amplitude of current in this case is always finite; it reaches this maximum amplitude in a finite time, and when the impressed emf. is removed the current dies away to zero. This is the action of the circuits which are now in every-day practical use.

In the second case the negative resistance is equal to the positive resistance, and the resultant effective resistance of the circuit is therefore zero. When an emf. is suddenly impressed in this case, the current in the circuit starts to increase at a rate which is directly proportional to the impressed electromotive force and to the square root of the ratio of the capacity to the inductance of the circuit (for a given impressed frequency). If the force is impressed for an infinite time, then the current in the circuit reaches infinity. If the emf. is impressed for a finite time, then the current reaches some finite value. When the impressed emf. is removed, the current in the circuit at that instant continues indefinitely with unchanged amplitude as a free oscillation. Theoretically, this is the limiting case for regeneration; practically, it is always necessary to operate at some point slightly below this state at which the circuits have a definite resistance.

It is important to note here that altho the circuit of this case has zero resistance, oscillations will not start unless an emf. is impressed upon the circuit; furthermore, that oscillations once started continue with undiminished amplitude indefinitely. This state cannot be attained in practice, because the negative resistance furnished by the tube is dependent on the amplitude of the current and for stable operation decreases with increasing amplitude.[1]

In the third case the negative resistance introduced into the circuit is greater than the positive resistance, and the effective resistance of the circuit is therefore negative. When an emf. is impressed upon a circuit in this condition, a free and a forced oscillation are set up which have some interesting properties. The amplitude of the forced oscillation is determined by the value of the impressed emf. divided by the resultant resistance of the circuit. The free oscillation starts with an amplitude equal to the forced oscillation, and builds up to infinity regardless of whether

[1] It is very important at this point to distinguish between this purely theoretical state and the state which exists in oscillating tube circuits. In the various forms of self-heterodyne circuits a free oscillation of constant amplitude is maintained in the system and the circuit may be considered as having zero resistance, *but only for that particular amplitude of current.* An external emf. impressed on the circuit always encounters a positive resultant resistance, assuming, of course, that the existing oscillation is stable. This is due to the non-linear characteristic of the tube.

or not the external emf. is removed. This free oscillation starts with an amplitude which is proportional to the impressed force, and this proportionality is maintained thruout any finite time interval (with constant impressed electromotive force).

It is important to note that altho the negative resistance of the circuit exceeds the positive, and the effective resistance of the circuit is negative, oscillations will not occur until some emf. is impressed. *Once an emf. is impressed, however, no matter how small it may be, the current in the circuit builds up to infinity regardless of whether or not the external emf. is removed.*

The fundamental difference between the case in which the resistance of the circuit is positive and the case in which the resistance of the circuit is negative may be summed up as follows: in the first, the forced oscillation contains the greatest amount of energy and the free oscillation is of very minor importance[2] (after a short interval of time), in the second, it is the free oscillation which contains the greatest amount of energy and the forced oscillation which is of negligible importance.

It is, of course, impossible with present-day instrumentalities to set up a system in which the negative resistance exceeds the positive without the production of oscillations in the system, since any irregularity in filament emission or impulse produced by atmospheric disturbances is sufficient to initiate an oscillation which builds up to the carrying capacity of the tube. It is, however, possible, by means of various expedients, to set up systems which avoid the production of such a paralyzing oscillation and which approximate the theoretical case in the use of a free oscillation to produce amplification.

The first use of the free oscillation in a regenerative system for the amplification of signals appears to have been made by Turner[3] in his valve relay system. Briefly, Turner prevented the regenerative circuit from producing oscillations when no signals were being received by placing a negative potential on the grid of sufficient value to hold it just below that point on the characteristic curve at which self-oscillation would start. The impressing of a small electromotive force of sufficient value would carry the potential of the grid over the "threshold" value and a free oscillation would start which would build up to the limiting

[2] This is strictly true when dealing with continuous waves which we have been considering. It is not true in the regenerative reception of spark signals, particularly of short wave length, large damping, and low spark frequency. In this case the energy in the free oscillation exceeds the energy in the forced oscillation.

[3] British Patent, 130,408.

value of the tube. The system was returned to its initial sensitive state by means of a relay operated by the increase in the plate current of the tube. This relay short-circuited the feed-back coil, thereby cutting off the supply of energy and permitting the potential of the grid to drop back below the "threshold" value. As Turner explains, the device is a relay with a low limit (as distinguished from an amplifier), but it appears to be the first device in which the free oscillation set up by an impressed electromotive force produced the magnified result.

Bolitho[4] contributed an important improvement by replacing the mechanical relay of Turner which operated only upon the receipt of a signal by a valve relay which was continuously operated by independent means. Briefly, this was accomplished by connecting a second valve to the oscillating circuit of the Turner arrangement with a reversed feed-back connection and supplying the plate circuit of this second valve with alternating current. When the "threshold" value of the first tube was overcome and a free oscillation started in the system, the reversed feed-back of the second tube comes into action and at that time when the voltage supplied to the plate is positive, damps out the free oscillation and permits the grid of the first tube to return below the "threshold" value. This represents the second step in the utilization of the free oscillation for the production of amplification.

It is the purpose of this paper to describe a principle of operation based on the free oscillation which is quantitative and without a lower limit. This new method is based on the discovery that if a periodic variation be introduced in the relation between the negative and positive resistance of a circuit containing inductance and capacity, in such manner that the neagtive resistance is alternately greater and less than the positive resistance, but that the average value of resistance is positive, than the circuit will not of itself produce oscillations, but during those intervals when the negative resistance is greater than the positive will produce great amplification of an impressed emf. The free oscillations which are set up during the periods of negative resistance are directly proportional in amplitude to the amplitude of the impressed emf. The variation in the relation between the negative and positive resistance may be carried out by varying the negative resistance with respect to the positive, by varying the positive resistance with respect to the negative, or by varying both simultaneously at some frequency which is generally rela-

[4]British Patent, 156,330.

tively low compared to the frequency of the current to be amplified.

These three methods of producing the super-regenerative state are illustrated respectively by Figures 1, 2, and 3, which figures indicate the general scheme of the system and the methods of varying the relation between the negative and positive resistance. Figure 1 shows a method of varying the negative resistance produced by the regenerative system by varying the voltage of the plate of the amplifying tube by means of a second tube, the grid of the second tube being excited by an emf. of suitable frequency.

Figure 2 illustrates a method of varying the positive resistance of the circuit with respect to the negative. This is accomplished by connecting the plate circuit of a vacuum tube in parallel to the tuned circuit of the regenerative system and exciting the grid by an emf. of suitable frequency. Figure 3 illustrates a combination of these two systems in which simultaneous varia-

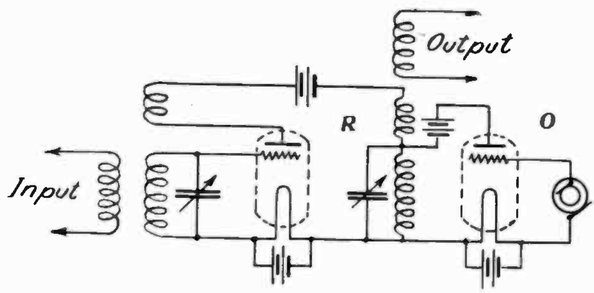


FIGURE 1

tions are produced in both the negative and positive resistances and provision made for adjusting the relative phases of these two variations.

A general idea of the phenomena occurring in these systems when an emf. is applied to the input circuit will be obtained from the diagram of Figure 4 which applies specifically to the circuit of Figure 1. This figure illustrates the principle relations existing in the system in which the positive resistance is constant and the variation is introduced into the negative resistance. It will be observed that the frequency of variation appears as a modulation of the amplified current so that the output circuit contains currents of the impressed frequency plus two side frequencies differing from the fundamental by the frequency of the variation.

Oscillograms of the essential current and voltage relations existing in the systems of the type illustrated by Figures 1 and 2

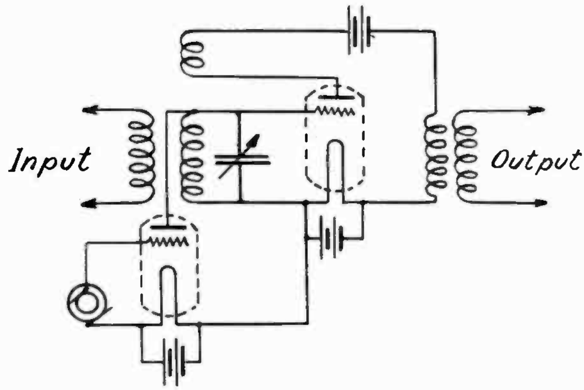


FIGURE 2

were obtained with the set up of apparatus illustrated in Figures 5 and 6, respectively. In the arrangement of Figure 6, in order to produce sufficient variation in the positive resistance of the tuned circuit, which was of large capacity and low inductance, it

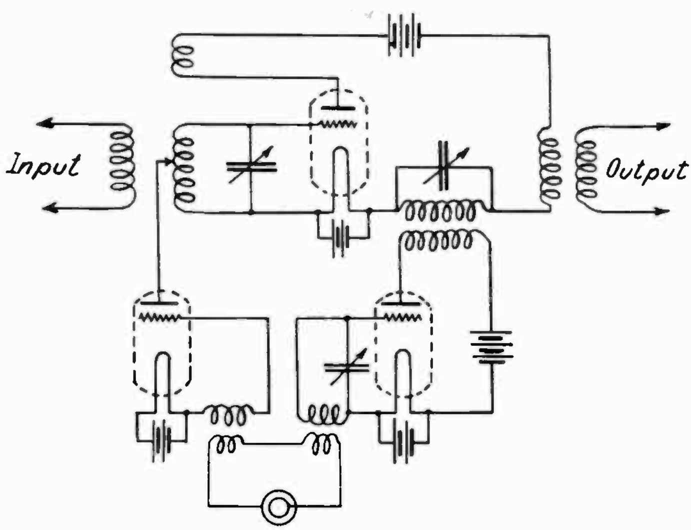


FIGURE 3

was necessary to use a two-electrode tube in series with the auxiliary emf.

Figures 7 and 8 are oscillograms respectively for a negative resistance variation and a positive resistance variation. The signaling emf. was impressed about half way along the film, the

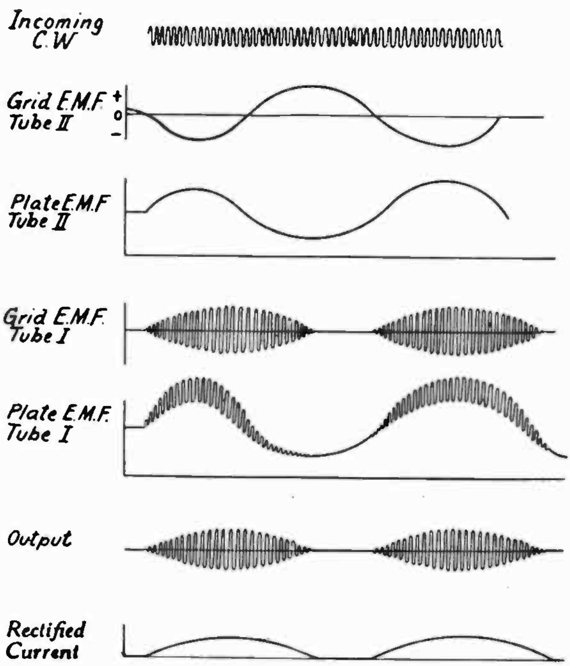


FIGURE 4
Tube I refers to R in Figure 1
Tube II refers to O in Figure 1

exact point at which the key was closed being indicated by the arrow. These oscillograms show phenomena which are in accordance with the explanations already given, but, in addition, show evidence of self excitation. It has been stated in the preceding pages of this paper that the basis of super-regeneration was the discovery that a variation in the relation between the

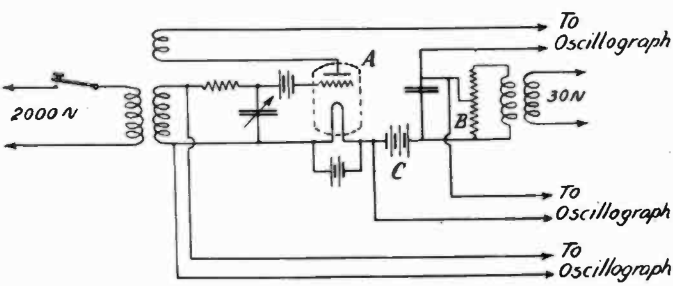


FIGURE 5
A-4 Western Electric Type L Tubes in parallel
B-AC Voltage = 100 Volts
C-DC Voltage = 160 Volts

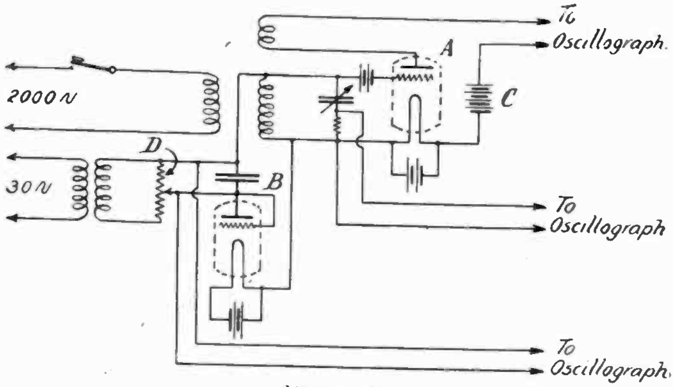


FIGURE 6

A-4 Western Electric Type L Tubes in parallel
B-1 Western Electric Type L Tubes with grid and plate in parallel
C-DC Voltage = 160 Volts
D-AC Voltage = 30 Volts

negative and positive resistances prevented a system which would normally oscillate violently from becoming self-exciting. An

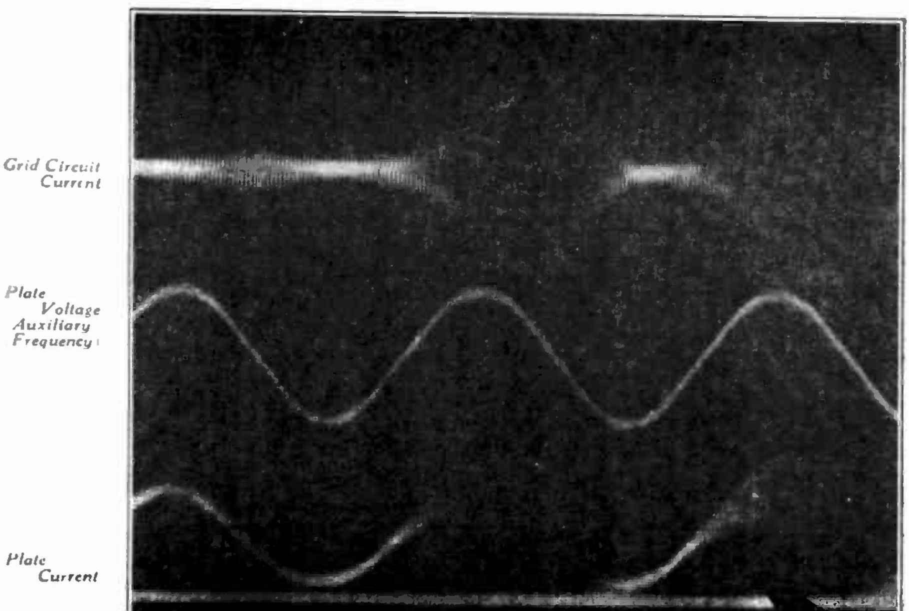


FIGURE 7

examination of the oscillograms will show that this is not strictly true, as a free oscillation starts every time the resistance of the circuit becomes negative. It will be observed however, that this free oscillation is small compared to that produced by the signal, and therein lies the complete explanation of the operation of the system. The free oscillations produced in the system when no signaling emf. is impressed, must be initiated by some irregularity of operation of the vacuum tubes, and must start at an amplitude equal to the amplitude of this disturbance. This initial value is of infinitesimal order, and hence, in the limited time interval in which it can build up the locally excited oscilla-

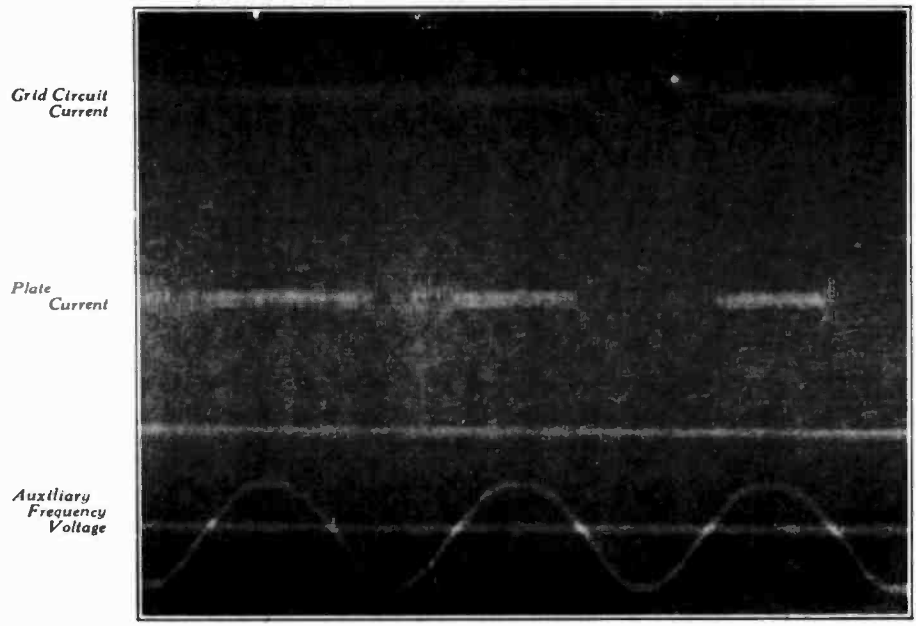


FIGURE 8

tion, never reaches an amplitude comparable to the oscillation set up by a signal of any ordinary working strength.

There is a second point of interest which is most evident from the curves of Figure 7. It will be observed that there is a decided lag in the maximum value attained by the free oscillation set up by a signal and the maximum value of plate voltage (negative resistance) of the amplifying tube. This is most evident from the plate current curve. It is a point of considerable interest, and the phenomena involved will be analyzed in a later part of the paper.

The rate of variation in the relation between the negative and positive resistance is a matter of great importance. It may be at sub-audible, audible, or super-audible frequencies. In radio signaling, for the reception of telephony, the variation should be at a super-audible frequency. For modulated continuous wave telegraphy and spark telegraphy, to retain the tone characteristics of the signals, it must be well above audibility; for maximum amplification a lower and audible rate of variation should be used. In continuous wave telegraphy, where an audible tone is required, the variation is at an audible rate; where the operation of an indicating device is required, a sub-audible frequency may be best. The choice of frequency is a compromise, particularly in telephony, since obviously the lower the frequency the greater the amplification, and the higher the frequency the better the quality.

Some practical forms of circuits are illustrated by Figures 9, 10, and 11, which illustrate respectively the three types of variation. Figure 9 shows a method of varying the plate voltage of the amplifying tube *R* by means of the vacuum tube oscillator *O* coupled into the plate circuit. In this arrangement a third tube *D* acts as a detector. This is essential when an audible frequency is employed; when a super-audible frequency is used the telephones can be placed directly in the plate circuit of the amplifying tube.

Figure 10 shows the second case in which the variation is

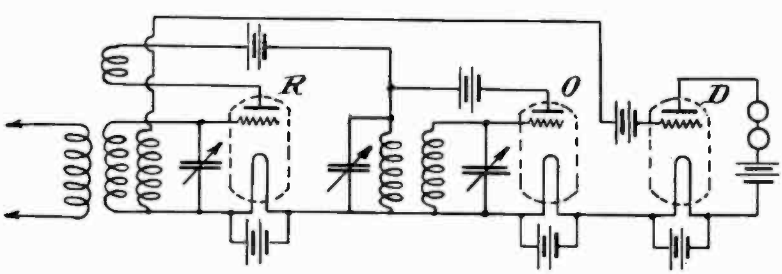


FIGURE 9

introduced into the positive resistance of the tuned circuit. This is done by means of an oscillating tube *O*, the grid circuit of which is connected thru the tuned circuit *LC* of the amplifying tube *R*. The variation in the resistance of the circuit is effected thru the variation in potential of the grid of the oscillating tube. During that half of the cycle, when the grid of the oscillating tube is positive, energy is withdrawn from the tuned circuit in the form of a conduction current from the grid to the filament of the oscillating tube, thereby increasing the effective resistance of the circuit. During the other half of the cycle, when the grid of the oscillating tube is negative, no conduction current can flow thru the grid circuit of the oscillating tube, and hence no resistance is introduced into the tuned circuit of the amplifying tube. In this case the amplifying tube serves also as the detector for any frequency of variation, as the tuned circuit forms a sufficiently good filter even for an audible frequency to prevent a disturbing audible tone in the telephones.

Figure 11 illustrates the case of a simultaneous variation in both positive and negative resistances. This is accomplished by providing the amplifying tube R with a second feed-back circuit L_1C_1 and L_2C_2 adjusted to oscillate at some lower frequency, thereby introducing a variation in the negative resistance thru the variation of the plate potential of the amplifier and a varia-

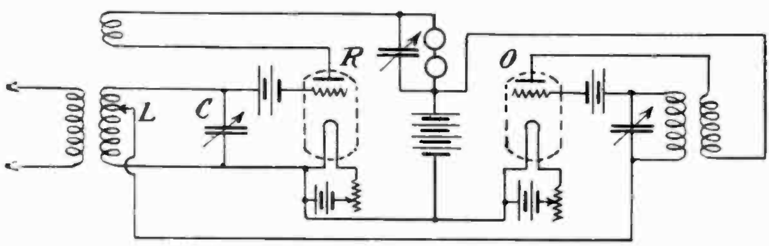


FIGURE 10

tion in the positive resistance by means of the variation of the grid of the amplifier. The proper phase relations between the negative and positive resistance are obtained by adjustment of the capacity of condensers C_1 and C_2 and the coupling between L_1 and L_2. In operation this system is very critical, and extreme care is necessary in order to obtain the super-regenerative state.

In each of the preceding cases the detecting function has been carried out either by a separate tube or by means of the amplifying tube. When a super-audible frequency of variation is employed, it is sometimes of advantage to perform the detecting function in the oscillating tube, and an arrangement for carrying this out is illustrated in Figure 12. The operation of this system is as follows: incoming signals are amplified by means of the regenerative action of the amplifier tube R and the variations of potential across the tuned wave frequency circuit LC impressed upon the grid of the oscillating tube O. These oscillations are

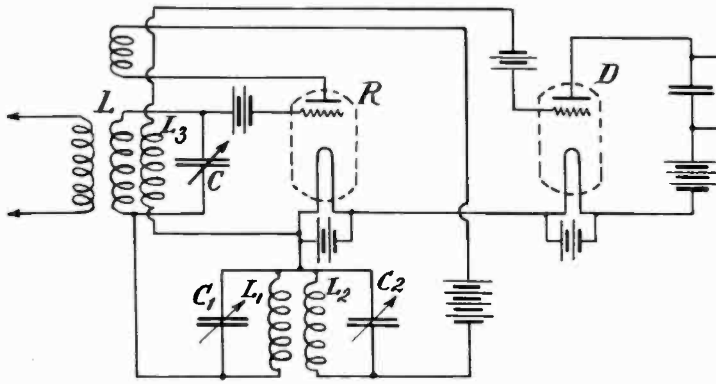


FIGURE 11

then rectified, and two frequencies are produced in the circuits of the amplifier tube. One of these frequencies corresponds to the frequency of modulation of the signaling wave. The other corresponds to the frequency of the variation and contains a modulation in amplitude corresponding to the modulation of the transmitted wave. This second frequency is then impressed upon the circuits of the oscillating tube with which it is in tune, amplified by the regenerative action of the system $L_1C_1L_2O$, and then rectified. The amplification obtainable with this form of

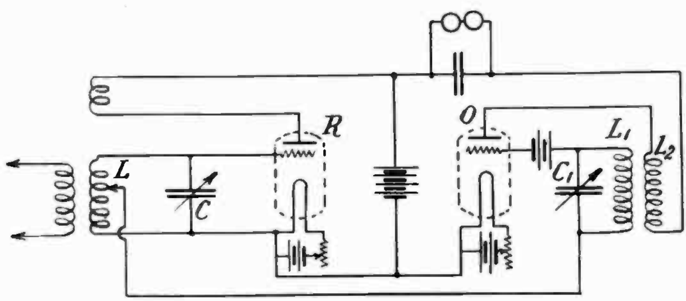


FIGURE 12

system is considerably greater than that of the single amplification circuits, but is naturally more complicated to operate.

When a super-audible variation is employed in a system such as illustrated in Figure 1, it is generally necessary to introduce a certain amount of resistance in the tuned circuit to insure the dying out of the free oscillation during the interval when the resistance of the circuit is positive. This is most effectively carried out by means of the arrangement illustrated in Figure

13, in which a secondary coil L_1 of large inductance and high resistance is coupled to the tuned circuit LC and the energy withdrawn thereby from the oscillating circuit stepped up and applied to the grid of the tube. In the operation of this system, a curious phenomena is encountered. This is the manifestation of an inductive reaction by the plate circuit of the amplifying tube to the auxiliary frequency emf. supplied the plate circuit

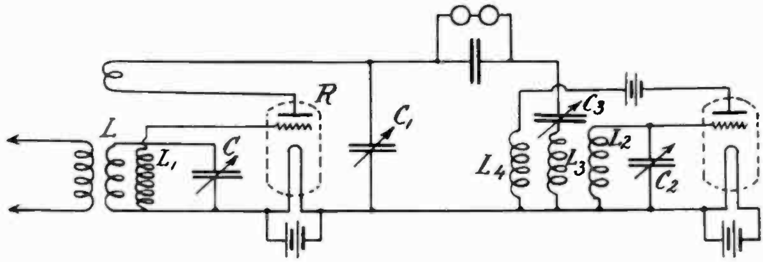


FIGURE 13

by the oscillating tube, which comes about in the following way. When the auxiliary emf. is impressed upon the plate of the amplifying tube, a current is produced in this tube in phase with the emf. across the tube. Now suppose the plate voltage is at its maximum positive value. This means that the negative resistance of the circuit is a maximum in amplitude. This in turn means that the average value of the grid is becoming more positive and the current in the plate circuit is likewise increasing. Since the free oscillation in the system will increase in amplitude as long as the resistance of the circuit is negative, it will reach its maximum amplitude after the maximum positive voltage is applied to the plate. Hence the component of current corresponding to the frequency of the variation set up in the plate circuit by the rectification of the radio frequency oscillations lags in phase behind the auxiliary emf. impressed on the plate. Hence the plate circuit of the tube manifests an inductive reaction to the auxiliary emf. It was found that this inductive reaction could be tuned out by means of the parallel condenser C_1 with a great improvement in the stability of the operation of the system and increase in the signal strength. The resonance point is pronounced, and once the other adjustments of the system have been correctly made is as readily found as any ordinary tuning adjustment.

The problem of cascade amplification with these systems is a rather involved one on account of a great number of effects which are not encountered in ordinary methods of cascade amplification. The principal trouble is the reaction of the second amplifying system on the first, and the difficulty of preventing it in any simple way on account of the high amplification per stage. While this difficulty is not insuperable, a simple expedient may be employed which avoids it. On account of the large values of radio frequency energy in these amplifying systems, the second harmonic is very strong in the plate circuit of the amplifying tube and is of the same order of magnitude as the fundamental if the tube is operated with a large negative voltage on the grid. Hence by arranging the second stage of a cascade system to operate at double the frequency and to amplify this harmonic, the difficulty is avoided. The general arrangement of such a system is illustrated by Figure 14, in which the positive resistance of the circuits LC and L_1C_1 of a two-stage amplifier are varied synchronously by a single oscillator. The circuit L_1C_1 in this case is tuned to the second harmonic of the circuit LC, but the combinations of circuits which may be arranged on this principle are very numerous.

One of the curious phenomena encountered with the super-regenerative system is found when it is attempted to secure

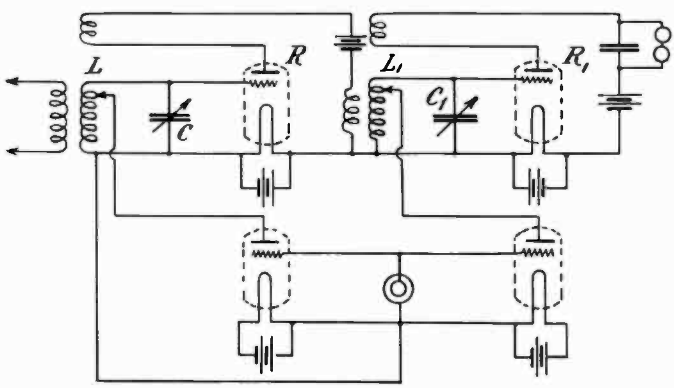


FIGURE 14

sharp tuning by the use of tuned circuits placed between the antenna and the amplifying system. The free oscillations set up in these circuits by the reaction of the amplifying system continue in these circuits during the interval when the resistance of the amplifier circuit is positive, re-excite the amplifier when the resistance becomes negative, and hence the entire system is kept in a continuous state of oscillation. The effect is most critical, and may be produced with most extremely weak couplings between the amplifier circuit and the second tuned circuit. The simplest solution of the difficulty is to perform the function of tuning at one frequency and amplification at another, and this is best accomplished by means of the super-heterodyne method illustrated by Figure 15. This may be adapted to work on either the sum or difference frequencies, but when the higher frequency is used, care should be taken that it is not near the second harmonic of the local heterodyning current. In the particular arrangement illustrated, *LCD* represents, together with the heterodyne, the usual agency for changing the incoming frequency, and

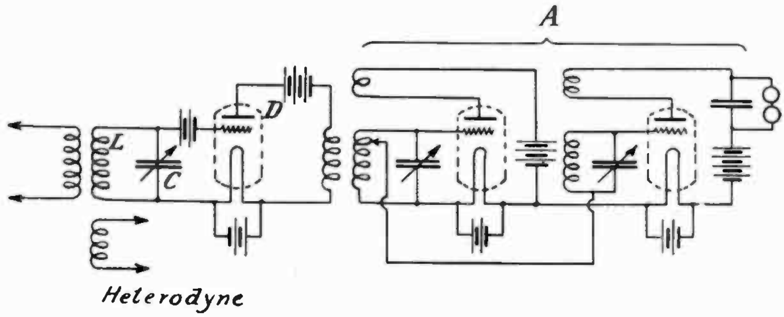


FIGURE 15

A represents the super-regenerative amplifier which may be of any suitable type.

Some of the results obtained in practice with super-regenerative systems compared to simple regenerative systems may perhaps be of interest. In general, it may be stated that the amplification which can be obtained varies with the frequency of the incoming signal and with the radio of the wave frequency to the auxiliary frequency. The higher the signaling frequency and the greater the ratio of this frequency to the auxiliary frequency, the greater the amplification. Other things being equal, it appears that the energy amplification varies as the square of the ratio of the signaling frequency to the auxiliary frequency. Hence, it follows that for telegraphic signals where an audible auxiliary frequency is used, much greater amplification can be obtained than in the case of telephony, where a super-audible auxiliary frequency must be employed. Using the arrangement of Figure 11 for a signaling frequency of five hundred thousand cycles, an energy amplification several million times greater than that obtainable with a simple self-heterodyne circuit is readily secured. Where a super-audible frequency is used for the reception of telephone signals, amplification of fifty thousand to one hundred thousand times energy can be obtained.

In a practical way the relative amplification of the new system with respect to the standard regenerative system for reception of telephone signals may be visualized as follows: With a signal so extremely weak that only the faintest of beat notes can be heard in the ordinary regenerative receiver, the super-regenerative receiver will give clearly understandable speech. For signals of sufficient strength to be understandable with the ordinary regenerative system with zero beat adjustment but not audible without local oscillations, the super-regenerative receiver will produce signals loud enough to be heard thruout the room.

Perhaps the most surprising characteristic of the system, apart from the amplification, is its selectivity with respect to spark interference when a super-audible frequency of variation is used. The explanation of this selectivity with respect, for example, to the ordinary regenerative receiver, lies in the periodic suppression of all free vibrations in the system. In the ordinary regenerative system spark interference approximates a form of shock excitation setting up a free vibration in the system which, because of the low damping existing therein, continues for a long period of time. An examination of the character of the oscillation set up will show that the energy existing in the free vibration after the initial impressed electromotive force is removed, is far greater than the forced vibration. In the ordinary system this free vibration may exist for a thousandth of a second or more. In the super-regenerative system this free vibration is damped out before it has proceeded more than one twenty thousandth of a second as a maximum. Hence, the interference from spark signals is greatly reduced. This phenomenon opens up a new field for the suppression of interference produced by shock excitation.

At the present time, on a three-foot loop antenna located twenty-five miles from the station WJZ at Newark, New Jersey, and a system of the type illustrated in Figure 12 with one stage of audio frequency amplification (three tubes in all) the announcements and musical selections are clearly audible five hundred yards from the receiver. With the same loop at the same distance, using the arrangement of Figure 11 without the separate detector tube, that is, with the telephones directly in the plate circuit of the amplifier tube, it is possible to operate a loudspeaking telephone so that the program from the Newark station is plainly heard thru a large size room. The signals with the arrangements of either Figures 11 or 12 are still heard loudly if the loop is discontinued from the receiver, the coils and wires of the receiver itself collecting sufficient energy to produce response.

While the new system does not amplify the ordinary spark signal with anything approaching its efficiency on continuous wave signals. one example of spark reception may be of interest. During the past winter an amateur spark station located at Cleveland, Ohio, and operating on a wave length of about 340 meters was received nightly at Yonkers, New York, on a three-foot (1 meter) loop and the arrangement of Figure 13 with sufficient intensity to enable the signals to be read thruout the room.

In conclusion, I wish to express my very great indebtedness to Professor L. A. Hazeltine for much valuable aid in connection with the theoretical side and to Mr. W. T. Russell for his assistance thruout the experimental side of this development.

June 2, 1922. Marcellus Hartley Research Laboratory,
Columbia University, New York.

SUMMARY: A system of circuits is described whereby the effective resistance of a regenerative circuit is periodically made positive and negative, tho predominantly positive. Such a circuit will respond to impressed electromotive forces by setting up free oscillations during the negative resistance period, which oscillations are proportional to the exciting emf. The forced oscillations produced by the exciting emf. will be comparatively small. The free oscillations caused by any normal variations in tube operation will also be small.

Means of carrying regenerative circuits thru the requisite resistance cycles are shown, and the practical operation of the system, including the case when extreme amplifications are desired, is discussed.

A NEW SYSTEM OF SHORT WAVE AMPLIFICATION*

By

Edwin H. Armstrong

(Columbia University, New York)

The problem of receiving weak signals of short wave length in a practical manner has become of great importance in recent years. This is especially true in connection with direction finding work where the receiver must respond to a very small fraction of the energy which can be picked up by a loop antenna.

The problem may be summed up in the following words:—to construct a receiver for undamped, modulated continuous, and damped oscillations which is substantially equally sensitive over a range of wave lengths from 50 to 600 meters, which is capable of rapid adjustment from one wave to another, and which does not distort or lose any characteristic note or tone inherent in the transmitter.

It is, of course, obvious that some form of amplification must be used, but a study of the various known methods soon convinces one that a satisfactory solution cannot be obtained by any direct method. In the interests of completeness, we will consider the three well-known direct means which might possibly be employed, and examine the limitations which apply to each. These three methods are:—

(1) Amplification of the audio frequency current after rectification;

(2) Amplification of the radio frequency current before rectification; and

(3) Application of the heterodyne principle to increase the efficiency of rectification.

Consider first the method of rectifying the radio frequency current and amplifying the resulting audio frequency current. Two limitations at once present themselves, one inherent in audio frequency amplifiers, and the other inherent in all known rectifiers. The limitation in the amplifier is the residual noise

* Presented before The Institute of Radio Engineers, New York, December 3. 1919. Received by the Editor, January 12, 1920.

which makes it impractical to use effectively more than two stages of amplification. The second limitation lies in the characteristic of the detector or rectifier. All rectifiers have a characteristic such that the rectified or audio frequency current is roughly proportional to the square of the impressed radio frequency emf. Hence the efficiency of rectification becomes increasingly poorer the weaker the signal until a point is reached below which the detector practically ceases to respond.

The second method of attack on the problem is the amplification of the received radio frequency currents before rectification to a point where they can be efficiently dealt with by the detector. This method is ideal on long waves, and various methods of inductance, resistance, and capacity couplings have been successfully used, but when the attempt is made to use the same methods of coupling on wave lengths below 600 meters, it results in complete failure. This is because the low capacity reactance existing between the various elements of the tubes causes them, in effect, to act as a short circuit around the coupling means and thereby prevents the establishment of a difference of potential in the external plate circuit. It is, of course, possible to eliminate the short-circuiting by tuning with a parallel inductance but this introduces a complication of adjustment which is highly objectionable and the tuning of all circuits also leads to difficulty with undesirable internal oscillations.

The third method which might be used is the heterodyne method to increase the efficiency of rectification. Great increase in signal strength is possible by means of this method, particularly where the signal is very weak, but there are certain reasons why it cannot be effectively used in practice at the present time. The chief reason in receiving continuous waves of short wave length is the instability of the beat tone which makes operations below 600 meters unsatisfactory. This disadvantage does not apply to the reception of spark signals but here the loss of the clear tone and its individuality offsets much of the gain due to increased signal strength. In the case of telephony the distortion which always results likewise offsets the gain in strength. It is, of course, undeniable that there are many special cases where the use of the heterodyne on short wave lengths is of the greatest advantage but the foregoing remarks apply to the broad field of commercial working where the practical aspects of the case greatly reduce the value of the amplification obtained by this method.

In spite of the great difficulties involved in a direct solution, great success was obtained by Round in England and Latour in France in the production of radio frequency amplifiers to cover effectively a range from 300 to 800 meters. This result was accomplished only by the most painstaking and careful experiment and it represents some of the very finest radio work carried out during the war. Round secured his solution by constructing tubes having an extremely small capacity without increase in internal resistance above normal values and coupling the tubes by means of transformers wound with very fine wire to keep down the capacity and very high resistance to prevent oscillation at the resonant frequency of the system. The effect of the high ratio of inductance to capacity and the high resistance of the winding is to flatten the resonance curve of the system and widen the range of response. Latour solved the problem by the use of iron core transformers wound with very fine wire, the iron serving the double purpose of increasing the ratio of inductance to capacity and introducing resistance into the system. Both these factors widen the range of response.

It is the purpose of this paper to describe a method of reception evolved at the Division of Research and Inspection of the Signal Corps, American Expeditionary Force, which solves the problem by means of an expedient. This expedient consists in reducing the frequency of the incoming signal to some predetermined super-audible frequency which can be readily amplified, passing this current thru an amplifier, and then detecting or rectifying the amplified current. The transform tion of the original radio frequency to the pre-determined value is best accomplished by means of the heterodyne and rectification, and the fundamental phenomena involved will be understood by reference to the diagram of Figure 1. Here *LC* represents the

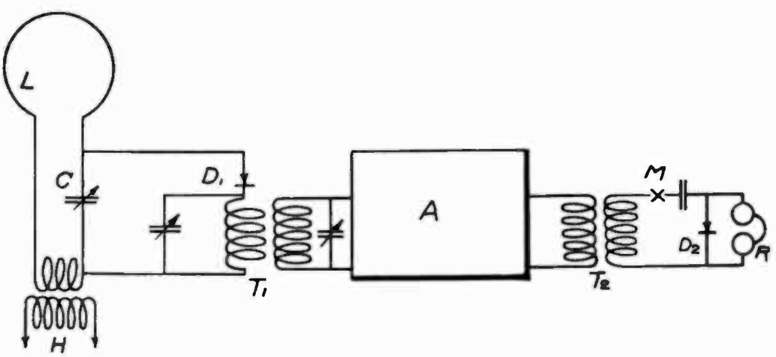


FIGURE 1

usual tuned receiving circuit, loop or otherwise, H a separate heterodyne, and D_1 a rectifier. A is a radio frequency amplifier designed to operate on some pre-determined frequency. This frequency may be any convenient frequency which is substantially above audibility. The amplifier is connected on its input side to the rectifier D_1, and on its output side to a second rectifier D_2 and a telephone or other receiver.

Suppose now that the frequency to be received is 3,000,000 cycles per second corresponding to a wave length of 100 meters and, for the sake of simplicity, that the incoming waves are undamped. Also, assume that the amplifier A has been designed for maximum efficiency at 100,000 cycles per second. The circuit LC is tuned to 3,000,000 cycles, and the heterodyne H is adjusted to either 3,100,000 or 2,900,000 cycles either of which will produce a beat frequency of 100,000 cycles per second. The combined currents of 3,000,000 and 3,100,000 (or 2,900,000) cycles are then rectified by the rectifier D_1 to produce in the primary of the transformer T_1 a direct current with a superimposed 100,000-cycle component. This 100,000-cycle current is then amplified to any desired degree by the amplifier A and detected or rectified by D_2. In order to get an audible tone where telephone reception is used some form of modulation or interruption must, of course, be employed in connection with this second rectification as the current in the output circuit of the amplifier is of a frequency above audibility. While this frequency is only 100,000 cycles and while it is therefore well within the range of practical heterodyning, its steadiness depends on the beats between 3,000,000 and 3,100,000 cycles per second and hence in any attempt to heterodyne it to audibility the same difficulties due to fluctuation would be encountered as in heterodyning the original radio frequency to audibility. However, the inability to use the heterodyne on the second rectification is not of great importance because the amplitude of the signal to be rectified is large and hence the difference (as far as signal strength in the telephone is concerned) between heterodyne and modulated reception is not great.

It is important to note here that the value of the heterodyne current in the first rectifier should always be kept at the optimum value in order to ensure the carrying out of the first rectification at the point of maximum efficiency. This adjustment, however, is not a critical one, and, once made, it is seldom necessary to change it. The amplifier A may be made selective and highly regenerative if so desired, and some very great increases in the selectivity of the system as a whole can be secured. Figure 2 illustrates the principle involved. This arrangement is substantially the same as Figure 1 except that the primary and secondary coils of the transformer T_1 are tuned by means of condensers as shown and the coupling between them is reduced to the proper value to insure sharp tuning. This system of

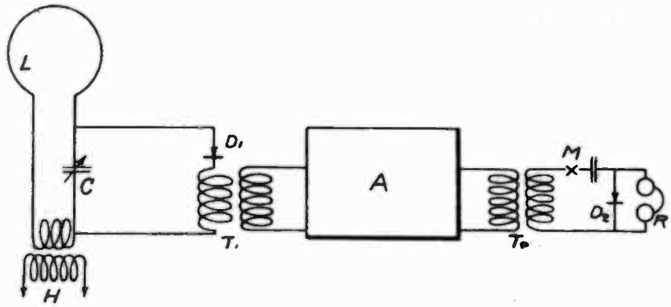


FIGURE 2

connection has all the advantages of tuning to the differential frequency in the manner well known in the art and an additional one due to the fact that since it is above audibility the musical character of atmospheric disturbances so troublesome in audio frequency tuning, does not appear.

So far, the reception of undamped waves only has been considered, but this method of amplification is applicable also to the reception of damped wave telegraphy and to telephony with practically equal efficiency and without distortion of any characteristics of tone. It is somewhat difficult to understand this, particularly in the case of the reception of spark signals as in all previous experience the heterodyning of a spark signal has resulted in the loss of the note, whereas in the present case the individuality between stations is more marked even than on a crystal rectifier.

This is the most interesting point in the operation of the system and the reason will be understood from the following analysis:

In heterodyning, the efficiency of rectification of the signaling current depends on its phase relation with the local current. If the two currents are either in phase or 180° out of phase the efficiency of rectification is a maximum; if 90° out of phase a minimum. In ordinary heterodyning, the initial phase difference depends on the time of sparking at the transmitter and hence this initial phase difference will be different for each wave train. As the frequency of the two currents are substantially the same, and as the duration of a wave train is short compared to the time necessary to produce a complete beat at an audible frequency, this initial phase difference is maintained thruout the wave train. Hence, the different wave trains are rectified with varying efficiency, the telephone current becomes irregular, and a rough or hissing tone results.

In the present method of heterodyning, the beat frequency is high so that several beats per wave train are produced. As a consequence, the phase angle between the signaling and local currents varies thru several cycles and the initial phase difference becomes a matter of minor importance. The number of beats which actually occur in practice depends on the beat frequency, the damping of the incoming wave, and the damping of the receiving circuit. As the damping of the receiving circuit is almost invariably much less than the damping of the incoming wave, it is the determining factor. In any practical case, however, where the beat frequency is kept above 20,000 cycles per second there is a sufficient number of beats to minimize the initial phase differences and maintain the characteristic tone.

The phenomena which occur in the reception of modulated continuous wave telegraphy and telephony are substantially a combination of those explained in the cases of undamped and damped wave reception. The adjustments are made in the same manner as for damped waves and the only precaution necessary in the reception of telephony is to damp the amplifier circuits somewhat to prevent distortion of the speech by excessive resonance.

The general arrangement found most suitable for practical working is shown in Figure 3. Both rectifications are carried out by three-element vacuum tubes. The amplifier here shown is resistance coupled, although any form of coupling may be used. The tuned circuits LC and L_2C_2 are preferably adjusted to some frequency between 50,000 and 100,000 cycles. The circuit LC may be made regenerative, if so desired, by any form of reactive coupling, but the practicability of this depends largely on the amount of time which is available for making adjustments.

In the diagram of Figure 3, only two stages of radio frequency amplification are shown, but at least four and preferably six should be used to get the maximum advantage of this method.

This is because the transformation of frequency is accomplished only by a certain loss so that something between one and two stages of amplification is required before this is overcome and it is possible to realize a gain. In this figure a separate heterodyne is shown, and it will generally be necessary to use it on account of the mistuning which is involved in the use of the self heterodyne This mistuning is considerable on 600 meters but on the shorter waves it is possible to use the self heterodyne method with equal efficiency as far as signal strength is concerned and a great gain in simplicity, as adjustments have been reduced to the minimum of a single one.

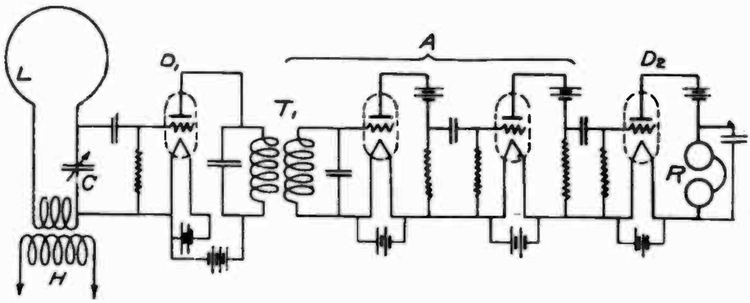


FIGURE 3

It may be observed here that this method is not limited to one transformation of frequency with one subsequent amplification. If the frequency to be received is 5,000,000 cycles this may be stepped down to 500,000 cycles, amplified, stepped down again to 50,000 cycles, re-amplified and detected. The great advantage of this method of amplification is that the tendency to oscillate due to the reaction between the output of the amplifier and the input is eliminated as the frequencies are widely different. The only reaction which can take place is in each individual amplifier. Hence, the process of extreme amplification is best carried out in stages of several frequencies, the amplification on each frequency being carried as far as possible without loss of stability. As soon as the limit of stable operation is approached, no further amplification should be attempted until the frequency has been changed.

The foregoing descriptions and explanations do not pretend to any save a most superficial treatment of the phenomena present in this method of reception. Lack of time has prevented a careful study and quantitative date only of the roughest sort has been obtained. Sufficient work has been done, however, to demonstrate the value of the method particularly in the case of modulated continuous wave telegraphy and telephony. In this field neither the amplification nor the selectivity can be equalled by any direct method.

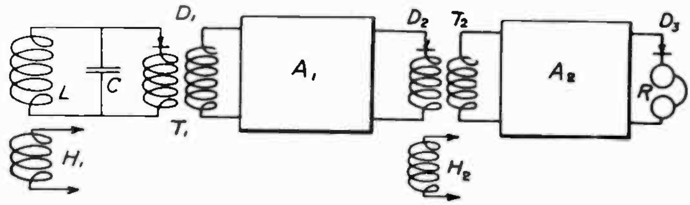


FIGURE 4

The practical results which have been obtained may perhaps be of interest. With a ten-turn, three-foot (1 meter) loop antenna and an amplifier consisting of six stages, resistence coupled, making a total of eight tubes, the night signals of ships working with the Florida and Gulf stations are loudly received. The night signals of amateur stations in the Middle West are regularly received as are also the signals of stations in the Gulf States. The general arrangement of the apparatus used is shown in Figures 5 and 6 which illustrate the scheme of connections of the frequency transformer and amplifier respectively. Four stages of amplification only are shown but six were actually used.

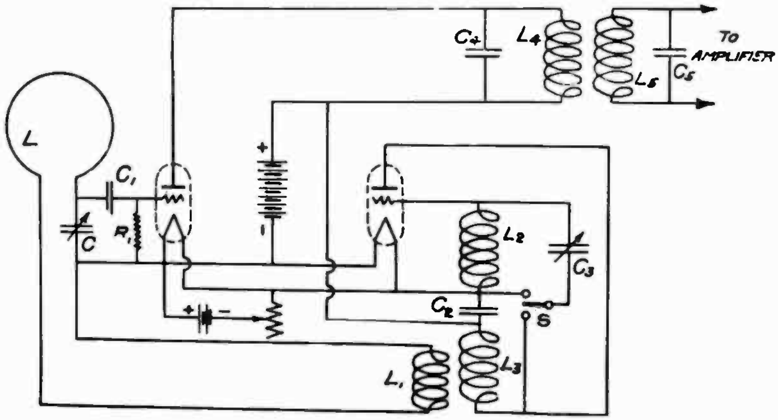


FIGURE 5

It is beyond question much more efficient to use some form of inductive coupling since the amplifier is intended to operate on only one frequency and the use of a resistance coupled amplifier is not recommended where one of the former type is available.

The new practice of this method involves the use of many known inventions, but in connection with the production of a superaudible frequency by heterodyning I wish to make due acknowledgment to the work of Meissner, Round, and Levy, which is now of record. The application of the principle to the reception of short waves is, I believe, new and it is for this reason that this paper is presented.

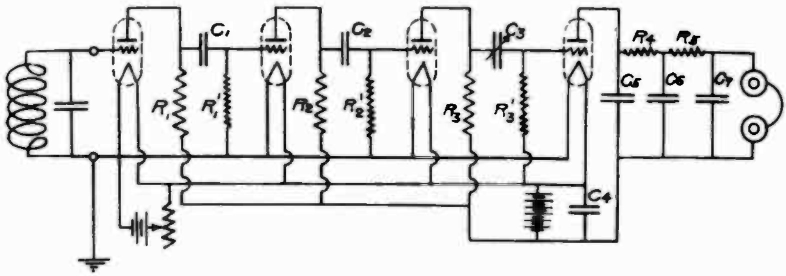


FIGURE 6

While the fundamental idea of this method of reception is relatively simple, the production of the present form of the apparatus was a task of the greatest difficulty for reasons known only too well to those familiar with multi-stage amplifiers; and to Lieutenant W. A. MacDonald, Master Signal Electricians J. Pressby and H. W. Lewis, and Sergeant H. Houck, all of the Division of Research and Inspection Signal Corps A. E. F., I wish to give full credit for its accomplishment.

Hartley Research Laboratory,
Columbia University, New York City.

SUMMARY: The various possible known methods of amplifying incoming signals of very short wave length (below 600 meters) are described and their limitations considered.

The new method then described consists (for continuous wave reception) of the following steps:—

1. Heterodyning, with the production of a beat frequency which is itself a *radio* frequency (for example, 100,000 cycles per second).
2. Rectification of the beat current.
3. Amplification at the beat radio frequency, preferably by a tuned amplifier.
4. Audio frequency modulation of the amplified current.
5. Rectification of the modulated current.

For reception of damped wave or radiophone signals, step 4 is omitted. It is shown that in this case the quality (characteristic tone) of the incoming signals is preserved.

DISCUSSION

A. S. Blatterman (by letter received December 17, 1919): Up to the present time it has been found very difficult to amplify radio signals having oscillation frequencies of the order of 1,000,000 cycles and practically impossible to do so when the frequency reaches 3,000,000 cycles or greater. The difficulties are attributable chiefly to capacity effects in the vacuum tubes as well as in the wiring, and also because it is a fairly difficult problem to build a really satisfactory coupling impedance or transformer to connect up the output of one tube with the input side of the tube next in the series when the frequency is very high. Movements of the hands of the operator or of his body near the apparatus in such cases cause extremely minute changes in capacity which are, nevertheless, sufficient to cause changes in tuning that seriously reduce the received signals.

Moreover, it is seldom, if ever, that a radio receiver can be designed for a single frequency. Both at the extremely high frequencies just mentioned as well as for the lower frequencies corresponding to long wave lengths, it is practically always necessary to arrange for reception over a more or less limited range of wave lengths and this requirement has also been a very serious factor in the design of all radio frequency amplifiers up to the present time. At radio frequencies it is possible, convenient and desirable to use tuned transformers for the couplings between successive stages; but because of the necessity for making the amplifier responsive over a large number of wave lengths the tuning of the transformers must be relatively broad. This involves the arbitrary introduction of resistance into the circuits and the loss in efficiency that results seriously reduces the overall amplification.

Major Armstrong has met the above difficulties in the way of radio frequency amplification by a method which in principle is as simple as it is highly ingenious, and, at least for the amplification of excessively short wave lengths, appears to be a satisfactory solution of the problem in hand. The principle can, as is stated by Major Armstrong, be applied to damped wave and continuous wave telegraphy and to telephony. For receiving continuous waves a second heterodyne, either self or separate, must be brought to act on the second detector or else some form of chopper must be used. For very short waves of the order of 50 meters, it is possible to make a self-heterodyne of the first tube and thus avoid the extra adjustments and apparatus required by a separate local oscillator. In this case it is advisable to use as low a beat frequency as possible in order not to necessitate too much mistuning, and to design the amplifier circuits accordingly. The question, however, of selecting the proper super-audible beat frequency and the actions involved in the performance of these circuits are not as simple perhaps as Major Armstrong may have led some of us to believe. Upon closer inspection it is found that certain limitations must be imposed upon the design, especially in application to the reception of spark and telephone signals, and it appears likely that the system cannot be used to advantage at all radio frequencies.

The following paragraphs may be of particular interest in connection with the opinion held by some that the present amplifier will tend toward returning spark radio systems to the favor accorded them before the advantages of continuous waves were so fully appreciated and utilized.

General Theoretical Considerations

In the reception of *continuous waves* by the method under consideration the actions involved are relatively simple. The interference of the incoming signal oscillation with that produced locally results in a beat frequency which is almost truly sinusoidal and makes the design of the coupling transformers a very satisfactory proposition with the possibility of securing maximum amplification through sharp tuning and accurate resonance adjustments. In this case also, it is quite immaterial, as far as the operation of the amplifier is concerned, whether the super-audible beat frequency used is adjusted to something of the order of 100,000 or 200,000 cycles or whether it is set at a low value of say 15,000 cycles.

For receiving spark signals, however, and for telephony the situation is somewhat different. Special precautions must be taken in order to avoid distortion effects, and the selection of proper value of the super-audible beat frequency is important.

Figure 1 is supposed to represent trains of damped voltage oscillations such as are produced at the detector of a receiving circuit by a spark transmitter. The successive groups of oscillations recur at tonal frequencies, each group being the result of a discharge at the spark gap of the transmitter. The mathematical expression for such a train of oscillations may be written as follows:

$$[V+V_1 sin(pt+\phi_1)+V_2 sin(2pt+\phi_2)+V_3 sin(3pt+\phi_3)+ \cdots V_n sin(npt+\phi_n)] sin\, \omega_1 t \qquad (1)$$

wherein the bracketed expression is the equation of the envelope curve bounding the amplitude of the radio frequency oscillations, expressed in the form of a Fourier's series, and the last term, $sin\, \omega_1 t$, refers to the radio frequency oscillation of periodicity ω_1 which is to be considered as an oscillation modulated at audible frequency according to the envelope curve just mentioned.

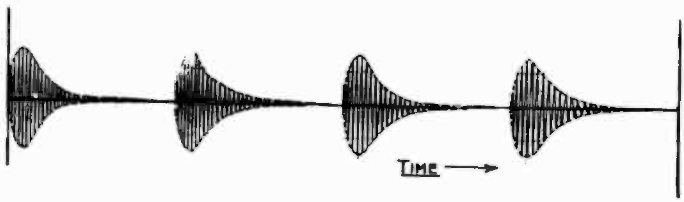


FIGURE 1

The envelope contains a fundamental frequency corresponding to p and all the harmonics $2p$, $3p$, $4p$, $\cdots$ np characteristic of the spark frequency and of the decrements of the transmitter and receiver. Thus, ordinarily the periodicity p would correspond to a 500- or 1000-cycle spark and the harmonics may run to the 10th or 20th before their amplitudes are small enough to make them negligible. V_1, V_2, V_3, $\cdots$ V_n designate respectively the amplitudes of the fundamental and the various harmonics. ϕ_1, ϕ_2, $\cdots$ ϕ_n and so on represent their phases.

The voltage produced by the local oscillation for heterodyning is

$$V' cos\,(\omega_2 t+\theta) \qquad (2)$$

The total or resultant voltage acting on the first detector at every instant is therefore given by the sum of expressions (1) and (2). This can be written in the following form

$$
\begin{aligned}
&V_1\left[\cos\frac{(\omega_1-\omega_2-p)\,t+(\phi_1-\theta)}{2}\cos\frac{(\omega_1+\omega_2-p)\,t+(\phi_1+\theta)}{2}\right.\\
&\qquad\left.-\sin\frac{(\omega_1-\omega_2+p)\,t+(\phi_1-\theta)}{2}\sin\frac{(\omega_1+\omega_2+p)\,t+(\phi_1+\theta)}{2}\right]\\
&+V_2\left[\cos\frac{(\omega_1-\omega_2-2p)t+(\phi_2-\theta)}{2}\cos\frac{(\omega_1+\omega_2-2p)t+(\phi_2+\theta)}{2}\right.\\
&\qquad\left.-\sin\frac{(\omega_1-\omega_2+2p)t+(\phi_2-\theta)}{2}\sin\frac{(\omega_1+\omega_2+2p)t+(\phi_2+\theta)}{2}\right]\\
&+V_3\left[\cos\frac{(\omega_1-\omega_2-3p)t+(\phi_3-\theta)}{2}\cos\frac{(\omega_1+\omega_2-3p)t+(\phi_3+\theta)}{2}\right.\\
&\qquad\left.-\sin\frac{(\omega_1-\omega_2+3p)t+(\phi_3-\theta)}{2}\sin\frac{(\omega_1+\omega_2+3p)t+(\phi_3+\theta)}{2}\right]\\
&+\\
&+V_n\left[\cos\frac{(\omega_1-\omega_2-np)t+(\phi_n-\theta)}{2}\cos\frac{(\omega_1+\omega_2-np)t+(\phi_n+\theta)}{2}\right.\\
&\qquad\left.-\sin\frac{(\omega_1-\omega_2+np)\,t+(\theta_n-\theta)}{2}\sin\frac{(\omega_1+\omega_2+np)t+(\phi_n+\theta)}{2}\right]\\
&+V\sin\omega_1 t \qquad\qquad (3)
\end{aligned}
$$

In each of the bracketed terms four different frequencies appear, namely,

$$\frac{\omega_1-\omega_2-k\,p}{4\pi}$$

$$\frac{\omega_1-\omega_2+k\,p}{4\pi}$$

$$\frac{\omega_1+\omega_2-k\,p}{4\pi}$$

$$\frac{\omega_1+\omega_2+k\,p}{4\pi}$$

k having the different values 1, 2, 3, 4, $\cdots n$ corresponding to the 1st, 2nd, 3rd, 4th, or nth bracket involving the 1st, 2nd, 3rd, or nth harmonic.

The explicit values of these frequencies depend principally upon the values ω_1 and ω_2 of the incoming and local radio frequencies and also to an increasing extent upon the periodicities $k\,p$, of the audio harmonic spark frequencies, for the higher harmonics. Relatively the four frequencies concerned may be of the same or very different orders of magnitude, and the two cases presented hereby involve important practical considerations in the design and use of the amplifier. The two different conditions may be treated separately under the headings (1) Short Wave Reception and (2) Long Wave Reception.

The wave lengths to be considered here are of the order of 50 or 100 meters, or shorter. In this case ω_1 and ω_2 are both very large and of the four frequencies mentioned above the two involving the differences $\omega_1 - \omega_2$ are considerably smaller than the two comprising the sums $\omega_1 + \omega_2$. Thus, the two trigonometric products which appear in each of the bracket terms of (3) indicate a radio frequency voltage of frequency

$$\frac{\omega_1 + \omega_2 \pm k\,p}{4\,\pi}$$

modulated by a considerably lower, tho still super-audible, frequency, in the present amplifier, of value

$$\frac{\omega_1 - \omega_2 \pm k\,p}{4\,\pi}$$

The form of such a voltage wave for one of the trigonometric products is shown in Figure 2.

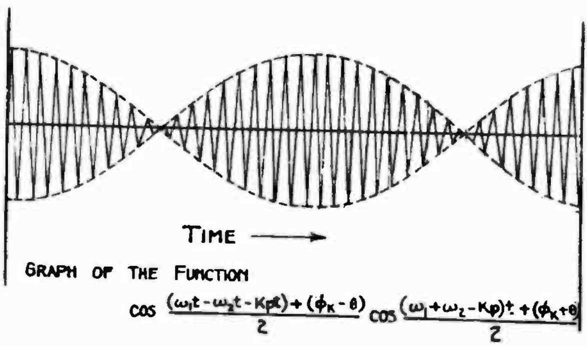


FIGURE 2

After rectification at the first detector tube the above frequencies are still essentially present and are impressed upon the amplifier proper. The frequencies $\frac{\omega_1 - \omega_2 \pm k\,p}{4\,\pi}$ are the heterodyne beat frequencies produced by interference of the local and signal voltages. The transformers of the amplifier are designed for frequencies of their order of magnitude and are not, therefore, affected by the radio frequencies $\frac{\omega_1 + \omega_2 \pm k\,p}{4\,\pi}$. No energy of these latter frequencies passes thru the amplifier. Neither does energy of the incoming signalling (radio) frequency ω_1 represented by the last term of (3), particularly if the transformers between stages of the amplifier are not broadly tuned. This is the normal way in which the amplifier works and is that described by Major Armstrong.

It is only the beat or difference frequencies

$$\frac{\omega_1-\omega_2-k\,p}{4\,\pi} \quad \text{and} \quad \frac{\omega_1-\omega_2+k\,p}{4\,\pi}$$

that have to be considered in designing the transformers and circuits. All of these frequencies lie in the neighborhood of the value

$$\frac{\omega_1-\omega_2}{4\,\pi}$$

which is the fundamental or basic beat frequency produced by the signal and local oscillations. They are greater and less than this value by the amounts

$$\pm\frac{p}{4\,\pi},\ \pm\frac{2\,p}{4\,\pi},\ \pm\frac{3\,p}{4\,\pi},\ \cdot\ \cdot\ \cdot\ \pm\frac{n\,p}{4\,\pi}$$

The transformers are fundamentally designed for the basic or mean frequency $(\omega_1-\omega_2)/4\,\pi$. This can be adjusted by regulating the local oscillation *but its proper value is by no means immaterial.* It is limited in the lower ranges by the fact that it must be above audibility, and thus about 20,000 cycles is as low as is permissible. The limitations in the other direction are those usually encountered in amplification of extremely high frequencies and a value of 5×10^5 cycles is about as high as can be used effectively.

The transformers should be as sharply tuned as possible to permit the building up of high voltages and avoid losses in resistance. A second requirement is that there shall be no distortion in the tonal quality of the received signal as it passes thru the transformers. This means that essentially all of the harmonics contained in the envelope curve of the arriving modulated oscillations must appear in the telephone current of the last detector. Thus, it is necessary to transmit equally thru the coupling transformers of the amplifier all of the frequencies.

$$\frac{\omega_1-\omega_2+k\,p}{4\,\pi} \quad \text{and} \quad \frac{\omega_1-\omega_2-k\,p}{4\,\pi}$$

and while designing the transformers for the basic frequency $(\omega_1-\omega_2)/4\,\pi$ the tuning must be broad enough so that the response is practically uniform over all the frequencies up to $n\,p/4\pi$ on either side of this basic value. A spark signal may contain appreciable harmonics up to the 10th or 20th which in a 500 cycle transmission of the usual type would mean that the amplifier transformers at the receiver would have to pass side frequencies up to 10,000 or 20,000 cycles above and below the basic frequency on which the design is based.

Laboratory experience has shown that it is difficult to build high frequency transformers tuned flatly enough to pass frequencies more than about 40 per cent above and below their best frequency. Even this value is accompanied by a marked loss of over-all efficiency because of the resistance effect, that must be introduced to broaden the tuning. It is obviously impracticable, therefore, to use transformers designed for a heterodyne frequency of 20,000 or 30,000 cycles, because a great many of the harmonic side frequencies that have to be transmitted to preserve the quality would be lost, and in order to get even a few of them the flat tuning required and the resistance inserted to secure it would mean low efficiency. It is much better in this case to work at a beat frequency of 100,000 cycles. The 10th harmonic in the spark signal under consideration, that is, 10,000 cycles, is then only off tune by 10 per cent which allows fairly good efficiency to be realized in the transformers. A beat frequency of 200,000 cycles would be even better.

There is another circumstance which favors the use of high beat frequencies, at least for the reception of short wave lengths, and that is that small changes in either the signal or the local oscillator frequencies such as might be caused by movements of the operator's hand or body in the neighborhood of one of the circuits, cause a much smaller percentage change in the beat frequency when this is high than when it is low, and the apparatus thereby becomes more nearly immune to such variations. At longer wave lengths, however, conditions are altered somewhat and there is an upper limit to the usable beat frequency.

The beat frequency can be produced with the local frequency (ω_2) either less or greater than the incoming frequency (ω_1). It is usually best, with short waves, to make ω_2 less than ω_1, because it is then more easily controlled and freer from variations of the type just mentioned.

Long Wave Reception

In the reception of long wave lengths a condition arises in which the incoming signal frequency is of the same order of magnitude as the heterodyne frequency for which the transformers are designed. Such is the case, for instance, when receiving a wave length of 3,000 meters with an amplifier tuned to the beat frequency of 100,000 cycles. When this condition exists, the incoming frequency, $\omega_1/2\pi$, represented by the last term of (3), passes thru the amplifier together with all the heterodyne frequencies

$$\frac{\omega_1-\omega_2-k\,p}{4\,\pi} \quad \text{and} \quad \frac{\omega_1-\omega_2+k\,p}{4\,\pi}$$

and interfering with all of them in their different amplitudes and phases produces a conglomeration of resultants which will be heard in the telephones, after rectification at the last detector, as a badly distorted, mushy signal like that usually heard when receiving spark signals on an ordinary oscillating receiver. This will always happen if the incoming signal frequency passes thru the amplifier. In order to avoid the effect, therefore, it is necessary to design the amplifier for heterodyne frequencies that lie wholly outside the range of wave lengths to be received. It is easy to accomplish this, as will readily be seen, when short wave lengths are involved but when waves of one or several thousand meters are to be handled the proper selection of the value of the heterodyne frequency requires careful consideration.

As an example, consider the case of a receiver to function on all wave lengths from 1,000 meters to 5,000 meters; that is, 300,000 cycles to 60,000 cycles. In order to avoid distortion of the kind just mentioned on certain wave lengths this whole band of frequencies is at once eliminated from use as heterodyne frequencies in the amplifier, and the range ought to be extended at least 10,000 cycles beyond this at both ends because the spark signal may contain appreciable harmonics up to this value and certain of the side frequencies of the incoming oscillation might therefore get directly through the amplifier and produce distortion. In the case under consideration, therefore, the amplifier ought to be designed for a frequency either less than 50,000 cycles or greater than 310,000 cycles.

The disadvantages in using low heterodyne frequencies (on the 50,000 cycle end in this case) have been pointed out above in discussing the reception of short waves. Broad transformer tuning with comparatively low efficiency is required to avoid the other kind of distortion due to elimination or at least the reduction of the higher harmonics. But in addition to this there must be considered the fact that static is always more pronounced at long wave lengths and an amplifier designed for low frequencies might therefore be expected to be more affected by these disturbances than one using higher frequencies.

For these reasons it appears very desirable to design the amplifier transformers for a beat frequency of the order of 350,000 or 400,000 cycles, that is, about 750 meters, in the case under consideration.

If spark or telephone signals were to be received on extremely long wave lengths such, for instance, as 15,000 meters (20,000 cycles) there is another consideration that would come in to limit the upper value of heterodyne frequency that could be used. This may best be explained by reference to the formula (3) above. High heterodyne frequencies of the order of 500,000 cycles cannot be used in this case because the sum of the signal and local frequencies $(\omega_1+\omega_2 \pm k\,p)/4\pi$ (carrier frequencies) would come thru almost as well as the difference or desired beat frequencies, namely, $(\omega_1-\omega_2 \pm k\,p)/4\pi$ (modulating frequencies) and very bad distortion would result. To take the figures given, f_1, would be 20,000 cycles and f_2 520,000 cycles. Their sum would be 540,000 and their difference 500,000, a variation of less than 10 per cent and both therefore conceivably within the working range of an amplifier transformer.

The type of distortion discussed above which is caused by the passage of the incoming frequency directly thru the amplifier and which results in a mushy, harsh signal can be confined to a rather narrow range of wave lengths by making the tuning of the amplifier transformers sharp. But this cannot be carried to extremes or, as has already been explained, it will then not be possible to pass the side frequencies. These will, in telephone transmissions, probably not exceed 2,000 cycles either side of the basic frequency but in spark signals may run to 10,000 cycles or so in extreme cases.

Sharpness of Transformer Tuning

In order to get an idea of the sharpness of tuning desirable in the transformers under different conditions the curves of Figure 3 are given showing the variation of secondary transformer potential as function of the ratio f_2/f_1 that is the ratio of the frequency to which the transformer secondary is tuned to the varying impressed frequency. Curve "a" is for a broadly tuned transformer of decrement 0.8; curve "b" represents sharper tuning with a decrement of 0.2. It will be seen that in the first case a frequency change of 10 per cent from the best value will cause a reduction in signal of about 5 per cent. In the second case, a difference in frequency from the best value of only 2 per cent causes the same change in signal.

If the 5 per cent reduction in potential for the side frequencies is assumed to be as much as is allowable in order to avoid distortion, and if it is further assumed that as sharp tuning as is represented by the curve "b" with 0.2 decrement is to be usable and the harmonics or side frequencies to be passed are to run to 5,000 cycles then the basic heterodyne frequency for which the transformers must be set will have to be at least 250,000 cycles; and if 10,000 cycles either side of the basic frequency are to be passed the latter cannot be less than 500,000 cycles, which is about the upper practical limit. It turns out, therefore, that

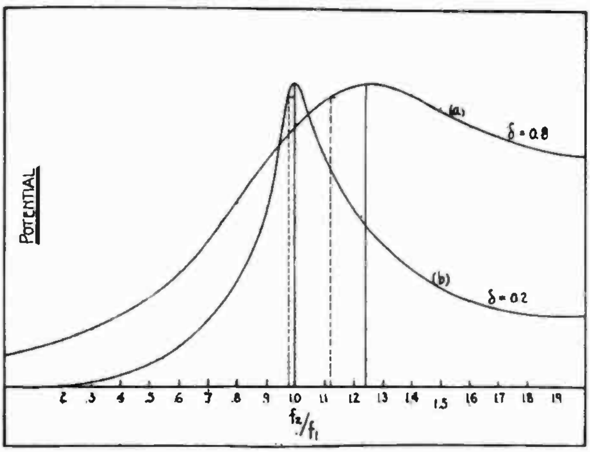


FIGURE 3

curve "b" corresponding to a decrement of 0.2 represents about as sharp tuning as can be used, and even then it is necessary to use the higher range of available heterodyne frequencies. It is to be noted that this tuning is by no means sharp as judged by the standards usually set for radio circuits.

With such tuning, frequencies 15 per cent greater and 30 per cent less than that to which the transformer is tuned are only reduced in amplitude by one half, and considerable energy within these frequencies would get directly thru the amplifier and produce the distortion just mentioned with harsh signal. In figures, it may be expected, if the amplifier were tuned to 3,000 meters, that mushy signals would be obtained for all waves between 3,900 meters and 2,550 meters.

If low heterodyne frequencies are to be employed then the tuning must be broader and the resonance curve "a" applies. Here, the allowable reduction of 5 per cent in response occurs for a change of about 10 per cent in frequency from the optimum value which means that the latter must be set for at least 50,000 cycles if a side frequency of 5,000 cycles is to get thru sufficiently to prevent distortion. With such broad tuning, however, even frequencies of half the value for which the transformers are designed get thru directly with very little loss and distortion with the mushy, harsh type of signal may be expected over a wide range of wave lengths.

For purposes of design of the transformers it is possible from the above considerations to decide on the most suitable heterodyne frequency, the sharpness of tuning and the approximate decrement and to determine roughly the constants of the transformer from the relations

$$\delta = \frac{r_2}{2 f L_2}$$

$$f = \frac{1}{2 \pi \sqrt{L_2 C_2}}$$

Still another point is involved here. In a pair of tuned coupled circuits such as must be used in the amplifier, the secondary and primary voltages are proportional inversely to the square root of the tuning capacities in the two circuits. That is

$$V_2 = \sigma V_1 \sqrt{\frac{C_1}{C_2}}$$

To get large secondary potentials, therefore, it is best to use small capacity and large inductance. Then, in order to keep the tuning or decrement to the desired value, the resistance must be increased, and these statements would hold without any qualification were the output of the vacuum tubes not definitely affected by the transformer load in their plate circuits.

When tuned, the secondary of a transformer introduces an effective resistance into the primary equal to

$$\frac{M^2 \omega^2}{r_2}$$

so that changing the resistance of the secondary to secure the decrement required to pass the side frequencies affects the load on the tube. What is desired is to get as high a potential V_1 across the transformer primary as possible. This requires the load impedance to be high as compared with the internal tube impedance. Increasing r_2 therefore militates against this and the best results can only be secured by careful adjustment of all of the factors, coupling, resistance, and inductance to the frequency involved.

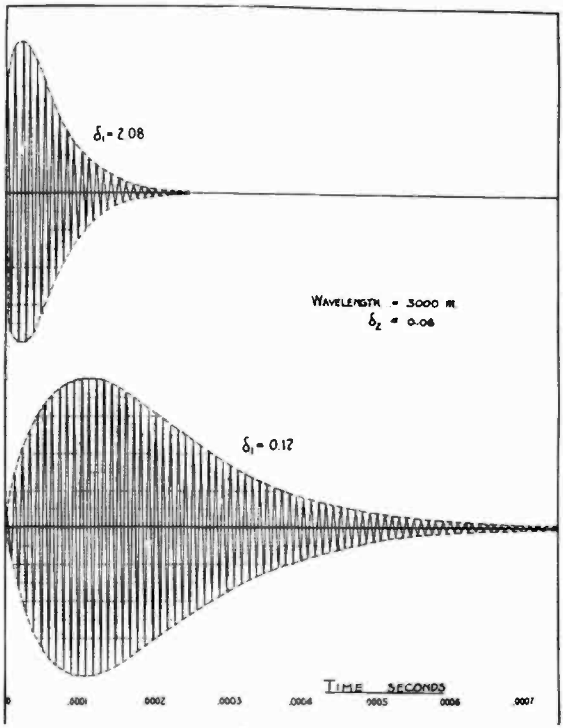


FIGURE 4

EFFECT OF TRANSMITTER DECREMENT AND ATMOSPHERICS

It appears that this type of amplifier functions most effectively on incoming waves of low decrement and that atmospheric disturbances which are always highly damped or else actually dead beat may be eliminated to a very considerable degree.

Curve "b" of Figure 4 shows a train of oscillations in a receiving circuit such as would be produced by a spark transmitter operating at 3,000 meters wave length and decrement 0.12. The decrement of the receiver for this curve was taken as 0.08. Curve "a" is similar but drawn for an excitation of high decrement, 2.08, approximating a static disturbance of the same frequency as that to which the receiver is tuned, that is, 3,000 meters. These curves can both be represented by equations of the form of equation (1) in which the Fourier's series gives the equation of the envelope curve of the oscillations.

For the curve "b," that is, the case of smaller damping, the different amplitudes of the harmonics and of the constant term in the representative series are as follows:

$$V = 7.86$$
$$V_1 = 12.23$$
$$V_2 = 5.92$$
$$V_3 = 3.72$$
$$V_4 = 2.09$$
$$V_5 = 1.56$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$V_{10} = 0.40$$

For curve "a" with high decrement the constants are:

$$V = 23.6$$
$$V_1 = 38.4$$
$$V_2 = 24.0$$
$$V_3 = 19.2$$
$$V_4 = 14.3$$
$$V_5 = 12.2$$
$$V_{10} = 6.3$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$V_{20} = 3.36$$

The amplitudes of the fundamental and various harmonics in the two cases are plotted in Figure 5 assuming the fundamental to be 1,000 cycles as in the usual spark transmission. It is seen that the amplitudes in the highly damped signal fall off much less rapidly than those of the more lightly damped signal. This means that in the former case a great deal of the total energy is contained in the harmonics, and if these are not passed thru the amplifier there will not only be distortion but loss in volume of signal as well. The use of a feebly damped spark transmission with an amplifier tuned just sharply enough to pass the principal harmonics or side frequencies produced therefore gives a system which largely eliminates static disturbances.

In this respect the present arrangement is more effective than the ordinary radio frequency amplifier. In the latter the presence of strong signal oscillations at the detector, after having passed the amplifier, amplifies the static in the same way that a locally produced frequency would, so that when the receiver is tuned to the incoming signal very loud sounds are caused by the static. These diminish rapidly, however, as the receiver is detuned, because the signal energy then falls off and the ratio of this, the equivalent local oscillation amplitude, to the static amplitude being thus reduced there results a much greater than proportionate decrease in the endodyne amplification effect on the static, as has already been shown by Major Armstrong in another paper.[1] But in the new amplifier only the fundamental and the first few harmonics of the static impulse are amplified by these interactions and thus much of the energy of such disturbances is lost.

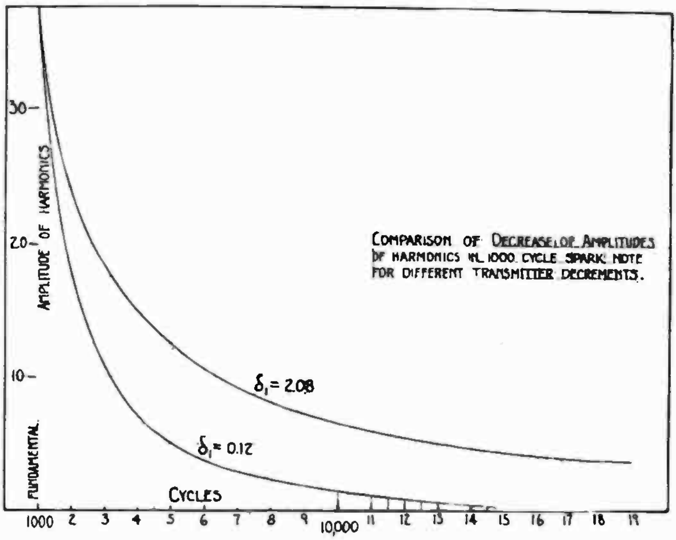


FIGURE 5

SUMMARY AND CONCLUSIONS

The above discussion has referred particularly to signals produced by spark transmitters, but the same general considerations are involved in telephone transmissions, except that in the latter the harmonic side frequencies to be considered will not generally exceed 2,000 cycles. The only point concerned in the case of sustained wave reception is that involving the passage of the incoming frequency directly thru the amplifier and this

[1] E. H. Armstrong, PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, April, 1917.

should be avoided with sustained waves for the same reasons that have been given to cover spark transmission.

Several practical considerations have been omitted from the discussion. Of these, one of the most important is the difficulty that is encountered in placing the circuits of a radio frequency amplifier with their transformers in a box in such a way that sharp tuning may be obtained and yet not have the whole or part of the system go over into oscillation. This involves careful adjustment of the various couplings and the resistances of the circuits and the proportions and arrangements are usually different for every wave length. It is suggested that an improvement might be made in this type of amplifier over the circuits that have been drawn by Major Armstrong, in which he uses air core tuned transformers in all of the stages of the amplifier, by the use of a tuned air core transformer behind the first detector tube feeding the first stage of the amplifier and with the stages following this coupled by means of carefully designed iron core transformers. The latter keep down stray field, and it has been found possible to build such transformers so as to get practically the maximum attainable amplification from the tube. By this arrangement the sharpness of tuning required in the amplifier is furnished by proper design of the first air core transformer, and the trouble experienced from coupling back, when several stages all tuned to the same frequency are employed, is reduced by the use of the iron core transformers which follow.

Two kinds of distortion are to be avoided. The first is caused by the passage of the incoming frequency directly thru the amplifier. The second is due to the more or less complete elimination of the harmonic side frequencies in passing thru the amplifier due to excessively sharp tuning. The type of amplifier in question is best suited to use on very short wave lengths, at least below 300 meters. At long wave lengths it is difficult to avoid distortion of the first of the two kinds mentioned, which, in the case of spark signals, results in a mushy, harsh note. Above 600 meters this type of distortion may be expected to occur over a band of wave lengths from 15 per cent to 30 per cent above and below that for which the amplifier is designed.

As regards an estimate of the allowable sharpness of tuning in different cases it would appear that this lies approximately between the limits set by decrements corresponding to 0.2, as about the sharpest tuning allowable, to about 0.8 for the broadest tuning. The latter would not be allowable except perhaps for the reception of very short waves. These figures apply only to the case where several tuned transformers are used in cascade in the amplifier. If the arrangement using one air core transformer and the balance iron core broadly tuned instruments as just described be used, the tuning of the first air core transformer might be made considerably sharper than this, of the order usually found in ordinary receiving tuners.

In general, the basic frequency to be used in the design of the amplifier may be higher for long wave lengths than for short up to a certain point, the practical limit being in the neighborhood of 400,000 or 500,000 cycles for the reception of 6,000 meter spark signals. For very long waves the beat frequency cannot be made so high.

The analysis indicates that the amplifier can be made to be freer from interference from highly damped spark stations and static disturbances than the usual types.

There is one other point that has not been mentioned tho I know it has already occurred to Major Armstrong himself. That is the question of the extent of the loss, if any, in effecting the change of incoming signal frequency to the value for which the amplifier is built. An experiment made[2] at Camp Alfred Vail in which the signal received on a simple non-regenerative tube was compared with that obtained by Major Armstrong's arrangement using a separate heterodyne, a rectifying tube for the super-audio note, and a detector tube, indicated that about equal signals were obtained by each method. Apparently, the heterodyne amplification in the second case just about makes up for the loss which accompanies the change in frequency.

Radio Laboratories,
Camp Alfred Vail, New Jersey,
December 4, 1919.

---

[2]By Mr. M. C. Batsel, Assistant Radio Engineer, Signal Corps, United States Army.

THE SUPER-HETERODYNE—ITS ORIGIN, DEVELOPMENT, AND SOME RECENT IMPROVEMENTS*

By

Edwin H. Armstrong

(Marcellus Hartley Research Laboratory, Columbia University New York)

The purpose of this paper is to describe the development of the super-heterodyne receiver from a wartime invention, primarily intended for the exceedingly important radio telegraphic direction-finding service in the Signal Corps of the American Expeditionary Force, into a type of household broadcasting receiver, which, with our present vision, appears likely to become standard.

The invention of the super-heterodyne dates back to the early part of 1918. The full technical details of this system were made public in the Fall of 1919.[1] Since that time it has been widely used in experimental work and is responsible for many of the recent accomplishments in long distance reception from broadcasting stations. While the superiority of its performance over all other forms of receivers was unquestioned, very many difficulties rendered it unsuitable for use by the general public and confined it to the hands of engineers and skilled amateurs. Years of concentrated effort from many different sources have produced improvements in vacuum tubes, in transformer construction, and in the circuits of the super-heterodyne itself, with the result that at the beginning of the present month there has been made available for the general public a super-heterodyne receiver which meets the requirements of household use.

It is a peculiar circumstance that this invention was a direct outgrowth of some experimental work undertaken to meet a very important problem confronting the American Expeditionary Force. This problem was the reception of extremely weak spark signals of frequencies varying from about 500,000 cycles to 3,000,000 cycles, with an absolute minimum of adjustments to enable rapid change of wave length. The

*Presented before The Institute of Radio Engineers, New York, March 5, 1924. Received by the Editor, April 26, 1924.

[1]Proceedings of The Institute of Radio Engineers, February, 1921. Presented December 3, 1919.

technical difficulties of this problem are now so well known that it was not necessary to consider them. Round in England and Latour in France, by some of the most brilliant technical radio work of the war, succeeded in producing radio frequency amplifiers covering the band from 500,000 to 1,000,000 cycles and tho covering a much more limited band, amplifiers operating on 2,000,000 cycles had been constructed. These results had been accomplished by the use of vacuum tubes and transformers of a minimum capacity. As this apparatus was used in the highly important intelligence services, all information was carefully guarded. When the United States entered the war, the facts that it was necessary to produce sensitive receivers for short wave lengths and that tube capacity would prove the bar to a straightforward solution of the problem were not known in this country. As a result, no attention was paid to the capacity in the type of vacuum tube which was adopted, and while the tube met the requirements of the lower frequencies admirably, it was impossible to use it effectively for the frequencies of importance in the direction-finding service.

During the early part of 1918, thru the courtesy and energy of General Ferrié and his staff, the American Expeditionary Force was supplied with apparatus of French manufacture. It was quite apparent, however, that this source of supply could not be a permanent one, and a solution of the problem became essential. During the early part of 1917, I had made a careful study of the heterodyne phenomena and their effect on the efficiency of amplification. With this work freshly in mind, the idea occurred to me to solve the problem by selecting some frequency which could be handled by the tubes available, building an effective amplifier for that frequency, and then transforming the incoming high frequency to this readily amplifiable value by some converting means which had no low limit; preferably the heterodyne and rectification. The principles and advantage of this method were explained in a paper presented before this INSTITUTE[3] and are now so well known that no further explanation is required here.

After much experimental work, an eight-tube set was constructed consisting of a rectifier tube, a separate heterodyne oscillator, three intermediate frequency amplifiers, a second

---

[2] PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, April, 1917. Presented October 4, 1916.

[3] This amplification is based on the ratio of the voltage applied to the second detector to the voltage at the loop terminals. The intermediate frequency amplification is unknown.

rectifier or detector, and two audio frequency stages. The intermediate frequency stages were coupled by tuned air-core transformers set for a frequency of about 100,000 cycles, with an adjustment for controlling the regeneration. The amplification of voltage measured at the input of the second detector with the amplifier just below the oscillating point, was about equivalent to a radio frequency amplification of 500.[4] This is illustrated in Figure 1 and the arrangement of its circuits in Figure 2. It gave satisfactory results except that the inclusion of a regenerative control on the intermediate frequency amplifier made skilled handling necessary, as the adjustment of the frequency of the oscillator changed the plate current of the detector tube and this, in turn, varied the resistance which that tube introduced into the amplifier system and upset the regenerative adjustment.

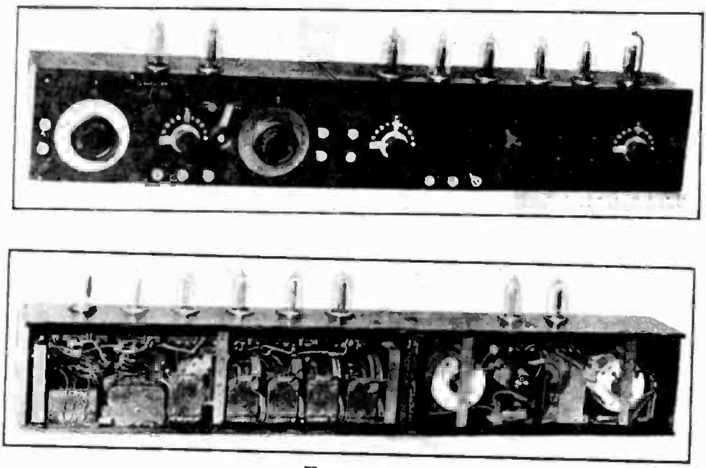


FIGURE 1

The Armistice ended development at this point, but in the fall of 1919, for the purpose of determining the results which could be obtained by pushing the super-heterodyne method of reception to the limit, a resistance-coupled intermediate frequency amplifier consisting of five high-μ (amplification factor) tubes was constructed. The voltage amplification of these five stages was probably between 5,000- and 10,000-fold. While greater amplification could have been obtained, the sensitiveness of a set composed of a two-tube frequency converter, a five-

[4] PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, February, 1921. Presented December 3, 1919.

tube intermediate frequency amplifier, a detector, and one-stage of audio, was such that on a three-foot (one-meter) loop, the sole criterion of reception was simply whether the signal was stronger than the atmospheric disturbances.

The sensitiveness of the super-heterodyne was demonstrated during the winter of 1919-1920, when the spark signals from amateur stations on the West coast and telephone signals from destroyers in Southern waters were received in the vicinity of New York on a three-foot (one-meter) loop. Probably the most striking demonstration of the capabilities of the method occurred in December, 1920, when Paul F. Godley, at Ardrosson, Scotland, received the signals of a large number of amateur stations located in the United States, many of them being spark stations. The super-heterodyne used by Godley consisted of a regenerative tube for the rectifier, a separate oscillator, four stages of re-

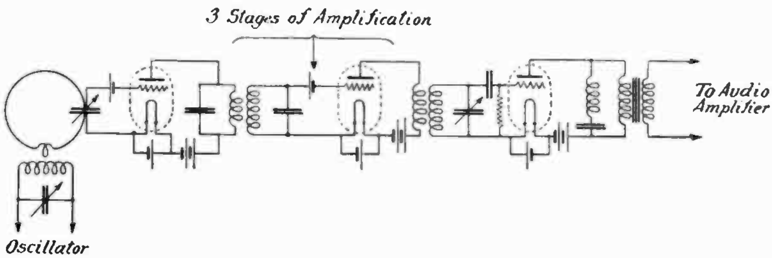


FIGURE 2

sistance-coupled intermediate frequency amplification, a second rectifier, and two stages of audio. While it is difficult to state definitely the actual voltage amplification obtained, it appears to have been between 3,000- and 5,000-fold.[5]

With the coming of the broadcasting art, and with the great increase in the number of stations and the consequent interference, the super-heterodyne began to take on a new importance —an importance which was based not on its superior sensitiveness nor on its selectivity, but on the great promise which the method offered in simplicity of operation. It was, and still is, the standard practice to furnish the public with receivers equipped with a variety of tuning adjustments for the purpose of amplifying the desired band of radio frequencies and excluding all others. As a matter of fact, many more adjustments than are on receivers should be used—more than could be placed in the hands of the average user. It would obviously be of the

[5] Based on standard previously described. This is without the second heterodyne which was used in receiving continuous waves.

greatest importance if in some way these tuning adjustments could be made in the laboratory by skilled engineers and sealed, leaving some relatively simple adjustment for the hands of the operator. The super-heterodyne offered the ideal solution. This solution lay in the construction of an intermediate frequency amplifier which would amplify a given frequency and a band 5,000 cycles above and below it and which would cut off sharply on either side of this desired band. The adjustments necessary to accomplish this could all be made by skilled men, and the only operations left for the user would be the two adjustments necessary to change the incoming frequency down to the band of the amplifier—adjustments which are not dependent on each other, which are of extreme simplicity, and which can be made equally well by the novice or the engineer. To determine just what could be accomplished along these lines, the writer, working in conjunction with Mr. Harry Houck, constructed during the spring of 1922, a set designed for the maximum usable sensitiveness and selectivity. The set-up consisted of one radio frequency stage (non-tuned transformer) a rectifier tube, and oscillator tube (used as a separate heterodyne), a three-stage iron-core transformer-coupled intermediate frequency amplifier designed to cover a a band of 20,000 to 30,000 cycles, a second detector tube, and two-stage of audio frequency amplification. UV-201 tubes were used. The set without the audio frequency amplifier is illustrated in Figure 3 and Figure 4. To prevent the intermediate frequency amplifier from oscillating, each stage was shielded separately. The use of a radio frequency stage ahead of the first detector possesses a number of advantages, but the chief one is in eliminating the reaction between the loop circuit and the oscillator circuit. Experience with the original type had shown that when an oscillator of ordinary power was used, it was necessary to couple it rather closely with the loop circuit in order to insure a sufficiently strong heterodyne current. This close coupling affected the tuning of both circuits, an adjustment of one changing the setting of the other. To avoid this trouble and to produce a system wherein a station could always be tuned in on exactly the same settings, a single stage of radio frequency amplification (non-tuned transformer) was used, and the oscillator was coupled into this transformer. This arrangement eliminated the reaction, reduced the radiation to a minimum and, in addition, removed the damping of the first rectifier from the loop circuit and improved its selectivity.

The results obtained with this set were about as expected.

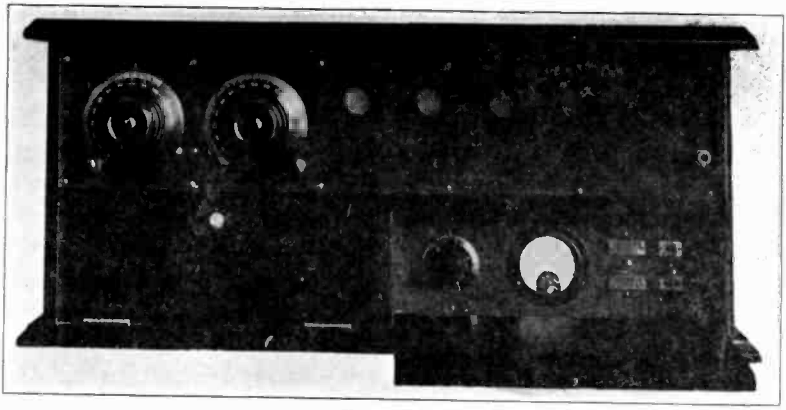


FIGURE 3

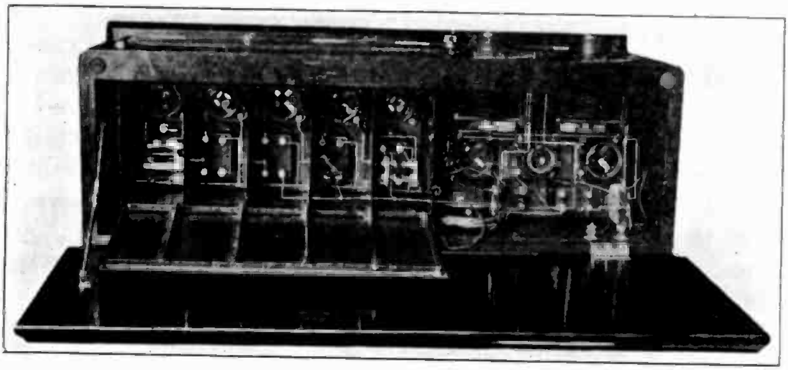


FIGURE 4

On a three-foot (one-meter) loop, the factor determining the reception of a station was solely whether the signal strength was above the level of the atmospherics. The selectivity was such that stations which had never been heard before on account of blanketing by local stations, were received without a trace of interference. While the performance of the set was much superior to any other receiver, it was apparant that the cost of construction and maintenance was prohibitive. The single item of a ten-ampere filament current will give some idea of the size of the storage battery and auxiliary apparatus required.

With the coming of the low filament consumption, or dry battery type of tube, the possibilities of producing a super-heterodyne for household use were tremendously improved. The set of Figure 3 was remodelled for the WD-11 tube, and its sensitiveness was brought to about the same value as obtained with the storage battery tubes. This was a long step forward, but still the cost was prohibitive.

It has been apparent ever since the question of the application of the super-heterodyne to broadcasting had been considered, that there were too many tubes performing a single function which were quite capable of performing a double one. The most outstanding case is that of the separate heterodyne oscillator. In view of our knowledge of the self-heterodyne, it appears quite obvious to perform the first rectification by means of a self-heterodyne oscillator and thereby save a tube. As a matter of fact, this was one of the very first things tried in France, but, except for very short wave lengths, it was never very successful when a high intermediate frequency was necessary. The reason was this: If a single tuned oscillating circuit was used, the mistuning to produce the proper beat caused a loss of signal strength which offset the gain of a tube. If two tuned circuits were used on the oscillator, one tuned to the signaling frequency and the other arranged to oscillate at the heterodyne frequency, then on account of the relatively small percentage difference in frequency a change in the tuning of one circuit changed the tuning of the other. The solution of this problem was made by Houck, who proposed an arrangement so simple and so effective that it completely solved the problem. Houck proposed to connect two tuned circuits to the oscillator, a simple circuit to the frequency of the incoming signal and a regenerative circuit adjusted to oscillate at such a frequency that the second harmonic of this frequency beating with the incoming frequency produced the desired intermediate frequency. The general arrangement is illustrated by Figure 5.

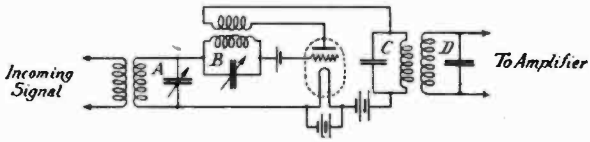


FIGURE 5

In the diagrammatic illustration, circuit A is tuned to the incoming signal, circuit B is tuned to one-half the incoming frequency plus or minus one-half the intermediate frequency, and the circuits C and D are both tuned to the intermediate frequency. The operation of the system is in line with ordinary self-heterodyne action. By reason of the asymmetrical action of the tube, there are created in the circuits a variety of harmonics. The second harmonic combines to produce beats with the incoming signals of the desired intermediate frequency, the tube rectifies them to produce the desired intermediate frequency and, thru *C* and *D*, the new frequency is supplied to the amplifier. On account of the fact that circuits *A* and *B* are tuned to frequencies differing by approximately 100 percent, a change in the tuning of one has no appreciable effect on the tuning of the other. This arrangement solved the oscillator problem and, in addition, practically eliminated radiation.

The next step in the reduction of the number of tubes was to make the radio frequency amplifier perform the function of amplifying intermediate frequency as well. This can be done with none of the difficulties inherent in audio frequency amplification, as the very small amplitudes of voltage handled by the first tube preclude the possibility of the grid becoming positive with respect to the filament. The general arrangement of circuits for carrying this out is illustrated by Figure 6. In this

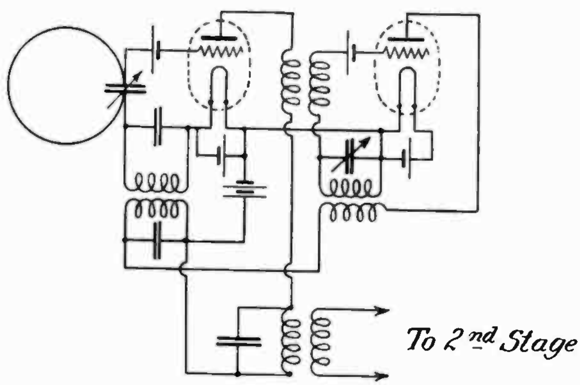


FIGURE 6

arrangement the signals received by the loop are amplified at radio frequency by the first tube and applied to the grid of a second harmonic oscillator by means of an untuned radio frequency transformer. The combined signaling and heterodyning currents are then rectified by the second tube, producing a current of the intermediate frequency which is applied to the grid of the first tube, amplified therein, and passed on to the second stage of the intermediate amplifier. A more practical method of carrying out this idea is illustrated in Figure 7. In this arrangement, a secondary of the first intermediate frequency transformer is connected to the grid of the first tube and in parallel with the loop circuit. Otherwise, the arrangements of Figures 6 and 7 are identical. The parallel type of circuit arrangement eliminates a variety of reactions which would give rise to oscillations of various frequencies and in addition, prevents the reception of long wave signals by the intermediate frequency amplifier. When this development had been completed, improvements in the design of the intermediate frequency transformers made it possible to obtain with two stages all the amplification which could be used.

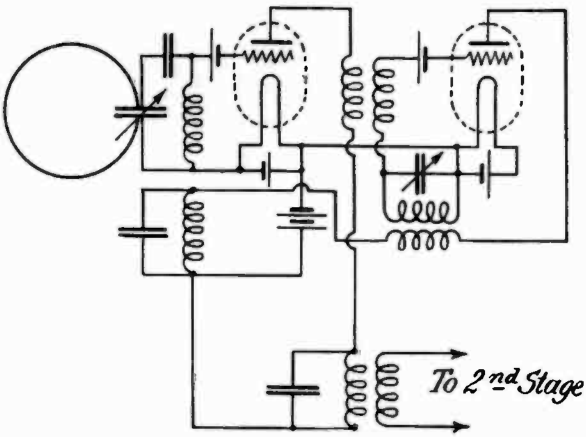


FIGURE 7

On account of the high amplification, signals from local stations overload the second rectifier and introduce distortion. Control of the amount of intermediate frequency amplification is essential. While there are numerous methods equally effective, the simplest one appears to be the control by means of the filament temperature of the second intermediate frequency amplifier.[6]

The features just described were all incorporated in the receiver, which is illustrated in Figures 8 and 9. The set measured 18 by 10 by 10 inches (45.6 by 25.4 by 25.4 cm.) and was completely self-contained—the batteries, loop antenna, and speaker mechanism being enclosed in the box. The results were highly satisfactory, and loud speaker signals (at night) in the vicinity of New York were obtained from stations in Chicago and Atlanta. It demonstrated that not only could a household receiver

[6]Although some form of potentiometer type of control of the voltage (a. c. applied to the grid of one of the amplifier tubes would obviously be better, the simplicity of the filament control has many advantages in manufacture.

of the super-heterodyne type be built, but that the first practical solution of the portable set was at hand.

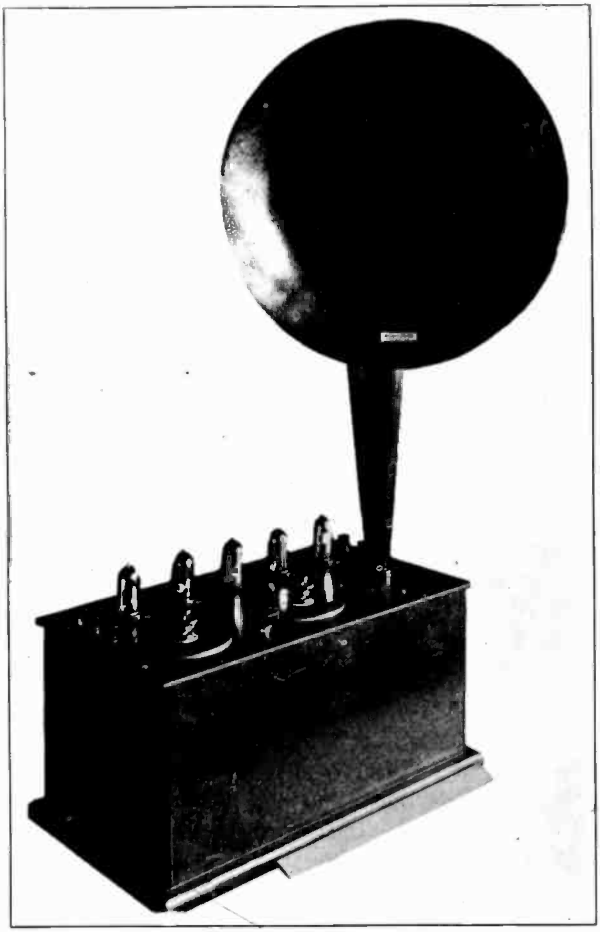


FIGURE 8

In this form, the capabilities of the set were brought to the attention of the Westinghouse Electric and Manufacturing Company and the Radio Corporation of America a little over a year ago. Its possibilities were instantly visualized by Mr. David Sarnoff, who immediately took steps to concentrate the resources of the research laboratories of the Radio Corporation of America, the Westinghouse Electric and Manufacturing Company, and the General Electric Company on this new development. From that point on it passed into a new phase—that of placing an in-

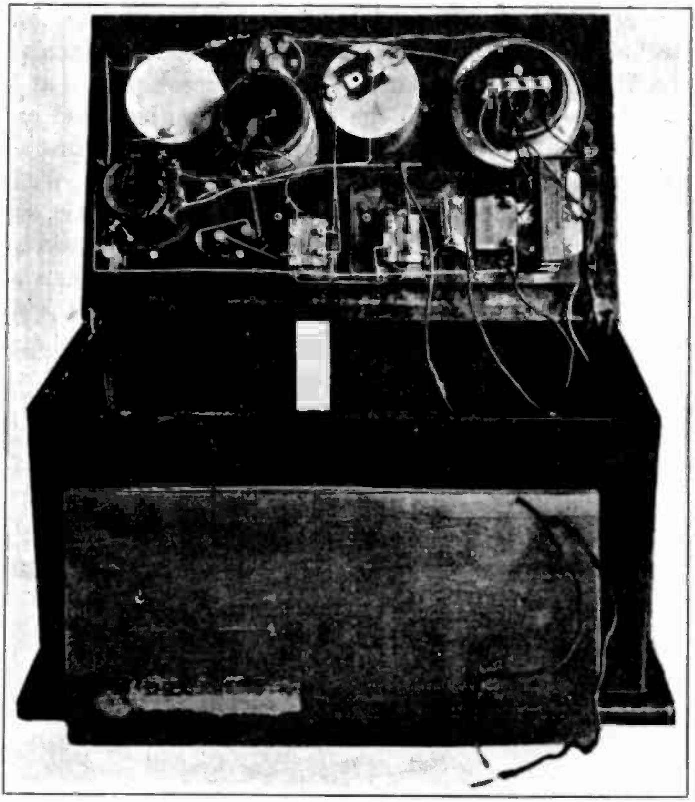


FIGURE 9

vention in a commercial form. In the limited time available, this was a most extraordinarily difficult proposition, and credit for its accomplishment is due to the untiring efforts on the part of the engineers of the above organizations. Many improvements and some radically new ideas of designs have been introduced, but it is the privilege of those responsible for them to present these. In the final development, an additional stage of audio frequency amplification was added in order to insure operation within steel buildings, particularly those within the city limits where signals are relatively very weak compared to suburban locations. This makes a six-tube set, but six tubes can be readily operated on dry batteries and the increase in sensitiveness is well worth the extra tube.

Some idea of the sensitiveness and the ease of operation of the set illustrated in Figures 10 and 11 may be gathered from an incident during the trans-Atlantic broadcasting tests of November and

December, 1923. On December 1st, two ladies, neither having any technical radio knowledge, received loud speaker signals from station 2LO, London, England. This was accomplished at Merrimac, Massachusetts, with the set and loop illustrated in Figures 10 and 11 and probably constitutes a record for the first radiophone reception from Europe with a portable receiver. With the same set and a three-foot (one-meter) loop, loud speaker signals from broadcast stations on the Pacific Coast were received in the vicinity of New York on an average of three or four times a week. The factor determining reception was simply whether the signal strength was above the level of the atmospheric disturbances.

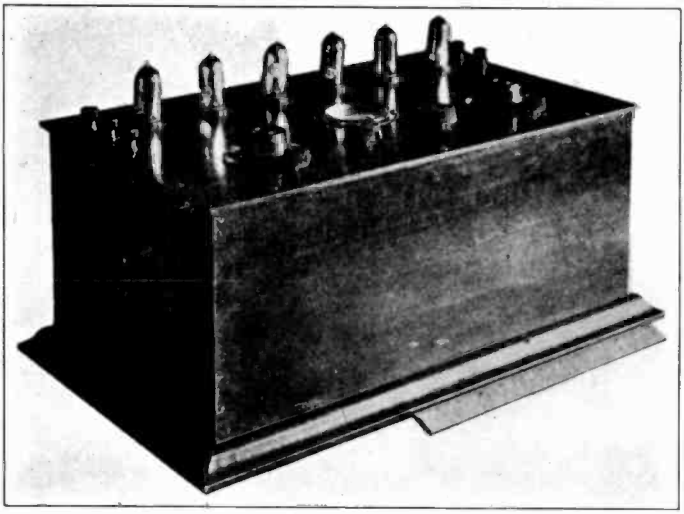


FIGURE 10

The type of super-heterodyne described is now available to the public in the two forms illustrated in Figures 12 and 13. Each of these sets incorporate the arrangements herein described. Their sensitiveness is such that, with a two-foot (61-cm.) loop and an unshielded location, the atmospheric disturbances are the criterion of reception. Here we reach a milestone in the development of broadcast receivers, for no increase in the distance of reception can now be obtained by increase in the sensitiveness of the receiver. Unless the power of transmitting stations is increased we are about at the limit of the distance which can be covered. Future improvement of this receiver will lie along the line of selectivity and simplifying the construction.

SUMMARY: This paper describes the development of the super-heterodyne receiver from a wartime invention into a commercial form of broadcast receiver apparatus now available to the general public. The success of the development is due to the low filament consumption vacuum tube and to the reduction in the number of tubes required by self-heterodyning, reflexing, and improvement in transformer design.

Instances are cited of trans-Atlantic and trans-continental reception of broadcast stations by completely portable sets constructed in accordance with the methods described.

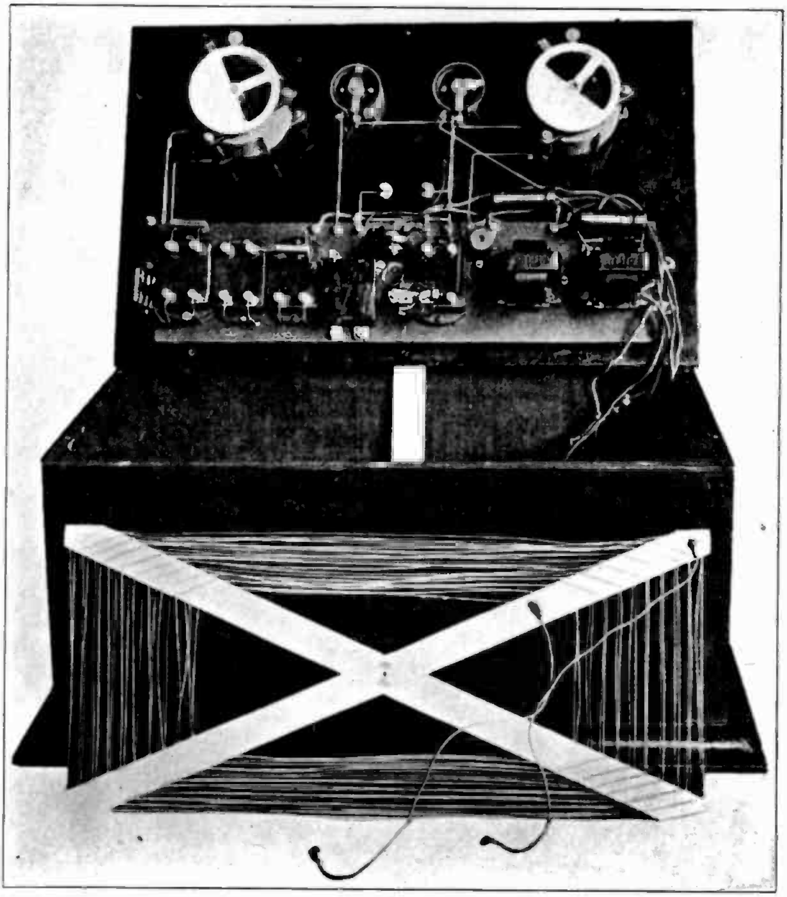


FIGURE 11

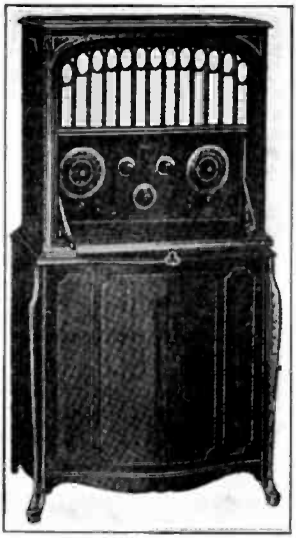


FIGURE 12

FIGURE 13

Proceedings of the Institute of Radio Engineers
Volume 24, Number 5 *May, 1936*

TECHNICAL PAPERS

A METHOD OF REDUCING DISTURBANCES IN RADIO SIGNALING BY A SYSTEM OF FREQUENCY MODULATION*

By

Edwin H. Armstrong

(Department of Electrical Engineering, Columbia University, New York City)

Summary—*A new method of reducing the effects of all kinds of disturbances is described. The transmitting and receiving arrangements of the system, which makes use of frequency modulation, are shown in detail. The theory of the process by which noise reduction is obtained is discussed and an account is given of the practical realization of it in transmissions during the past year from the National Broadcasting Company's experimental station on the Empire State Building in New York City to Westhampton, Long Island, and Haddonfield, New Jersey. Finally, methods of multiplexing and the results obtained in these tests are reported.*

PART I

IT IS the purpose of this paper to describe some recent developments in the art of transmitting and receiving intelligence by the modulation of the frequency of the transmitted wave. It is the further purpose of the paper to describe a new method of reducing interference in radio signaling and to show how these developments may be utilized to produce a very great reduction in the effects of the various disturbances to which radio signaling is subject.

Historical

The subject of frequency modulation is a very old one. While there are some vague suggestions of an earlier date, it appears to have had its origin shortly after the invention of the Poulsen arc, when the inability to key the arc in accordance with the practice of the spark transmitter forced a new method of modulation into existence. The expedient of signaling (telegraphically) by altering the frequency of the transmitter and utilizing the selectivity of the receiver to separate the signaling wave from the idle wave led to the proposal to apply the principle to telephony. It was proposed to effect this at the transmitter by varying the wave length in accordance with the modulations of the voice, and the proposals ranged from the use of an electrostatic micro-

* Decimal classification: R400×R430. Original manuscript received by the Institute, January 15, 1936. Presented before New York meeting, November 6, 1935.

phone associated with the oscillating circuit to the use of an inductance therein whose value could be controlled by some electromagnetic means. At the receiver it was proposed to cause the variations in frequency of the received wave to create amplitude variations by the use of mistuned receiving circuits so that as the incoming variable frequency current came closer into or receded farther from the resonant frequency of the receiver circuits, the amplitude of the currents therein would be correspondingly varied and so could be detected by the usual rectifying means. No practical success came from these proposals and amplitude modulation remained the accepted method of modulating the arc. The various arrangements which were tried will be found in the patent records of the times and subsequently in some of the leading textbooks.[1] The textbooks testify unanimously to the superiority of amplitude modulation.

Some time after the introduction of the vacuum tube oscillator attempts were again made to modulate the frequency and again the verdict of the art was rendered against the method. A new element however, had entered into the objective of the experiments. The quantitative relation between the width of the band of frequencies required in amplitude modulation and the frequency of the modulating current being now well understood, it was proposed to narrow this band by the use of frequency modulation in which the deviation of the frequency was to be held below some low limit; for example, a fraction of the highest frequency of the modulating current. By this means an economy in the use of the frequency spectrum was to be obtained. The fallacy of this was exposed by Carson[2] in 1922 in the first mathematical treatment of the problem, wherein it was shown that the width of the band required was at least double the value of the highest modulating frequency. The subject of frequency modulation seemed forever closed with Carson's final judgment, rendered after a thorough consideration of the matter, that "Consequently this method of modulation inherently distorts without any compensating advantages whatsoever."

Following Carson a number of years later the subject was again examined in a number of mathematical treatments by writers whose results concerning the width of the band which was required confirmed those arrived at by Carson, and whose conclusions, when any were expressed, were uniformly adverse to frequency modulation.

[1] Zenneck, "Lehrbuch der drahtlosen Telegraphy," (1912).
Eccles, "Wireless Telegraphy and Telephony," (1916).
Goldsmith, "Radio Telephony," (1918).

[2] "Notes on the theory of modulation," Proc. I.R.E., vol. 10, pp. 57–82; February, (1922).

In 1929 Roder[3] confirmed the results of Carson and commented adversely on the use of frequency modulation.

In 1930 van der Pol[4] treated the subject and reduced his results to an excellent form for use by the engineer. He drew no conclusions regarding the utility of the method.

In 1931, in a mathematical treatment of amplitude, phase, and frequency modulation, taking into account the practical aspect of the increase of efficiency at the transmitter which is possible when the frequency is modulated, Roder[5] concluded that the advantages gained over amplitude modulation at that point were lost in the receiver.

In 1932 Andrew[6] compared the effectiveness of receivers for frequency modulated signals with amplitude modulated ones and arrived at the conclusion that with the tuned circuit method of translating the variations in frequency into amplitude variations, the frequency modulated signal produced less than one tenth the power of one which was amplitude modulated.

While the consensus based on academic treatment of the problem is thus heavily against the use of frequency modulation it is to the field of practical application that one must go to realize the full extent of the difficulties peculiar to this type of signaling.

Problems Involved

The conditions which must be fulfilled to place a frequency modulation system upon a comparative basis with an amplitude modulated one are the following:

1. It is essential that the frequency deviation shall be about a fixed point. That is, during modulation there shall be a symmetrical change in frequency with respect to this point and over periods of time there shall be no drift from it.

2. The frequency deviation of the transmitted wave should be independent of the frequency of the modulating current and directly proportional to the amplitude of that current.

3. The receiving system must have such characteristics that it responds only to changes in frequency and that for the maximum change of frequency at the transmitter (full modulation) the selective characteristic of the system responsive to frequency changes shall be such that substantially complete modulation of the current therein will be produced.

[3] "Ueber Frequenzmodulation," *Telefunken-Zeitung* no. 53, p. 48, (1929).

[4] "Frequency modulation," Proc. I.R.E., vol. 18, pp. 1194–1205; July, (1930).

[5] "Amplitude, phase, and frequency modulation," Proc. I.R.E., vol. 19, pp. 2145–2176; December, (1931).

[6] "The reception of frequency modulated radio signals," Proc. I.R.E., vol. 20, pp. 835–840; May, (1932).

4. The amplitude of the rectified or detected current should be directly proportional to the change in frequency of the transmitted wave and independent of the rate of change thereof.

5. All the foregoing operations should be carried out by the use of aperiodic means.

The Transmitting System

An extensive experience with the various known methods of modulating the frequency convinced the writer as indeed it would anyone who has tried to work with this method of modulation at a high frequency that some new system must be evolved. During the course of this work there was evolved a method which, it is believed, is a complete solution of the transmitter problem. It consists in employing the modulating current to shift the phase of a current derived from a source of fixed phase and frequency by an amount which is directly proportional to the amplitude of the modulating current and inversely proportional to its frequency. The resulting phase shift is then put through a sufficient number of frequency multiplications to insure 100 per cent modulation for the highest frequency of the modulating current. By keeping the initial phase shift below thirty degrees substantial linearity can be obtained.

The means employed to produce the phase shift consisted of a source of fixed frequency, a balanced modulator excited by this source, and arrangements for selecting the side frequencies from the modulator output and combining them in the proper phase with an unmodulated current derived from the initial source. The phase relations which must exist where the combination of the modulated and unmodulated currents takes place are that at the moment the upper and lower side frequencies produced by the balanced modulator are in phase with each other, the phase of the current of the master oscillator frequency with which they are combined shall differ therefrom by ninety degrees.

The schematic and diagrammatic arrangements of the circuits may be visualized by reference to Figs. 1 and 2, and their operation understood from the following explanation. The master oscillator shown in these diagrams may be of the order of fifty to one hundred thousand or more cycles per second, depending upon the frequency of the modulating current. An electromotive force derived from this oscillator is applied in like phase to the grid of an amplifier and both grids of a balanced modulator. The plate circuits of the modulator tubes are made nonreactive for the frequency applied to their grids by balancing out the reactance of the transformer primaries as shown. The plate currents are therefore in phase with the electromotive force applied to the grid. The succeeding amplifier is coupled to the output transformer by a coil whose natural period is high compared to the frequency of the master oscillator and the electromotive force applied to the grid of this amplifier when the modulator tubes are unbalanced by a modulating voltage applied to the screen grids is therefore shifted in phase ninety

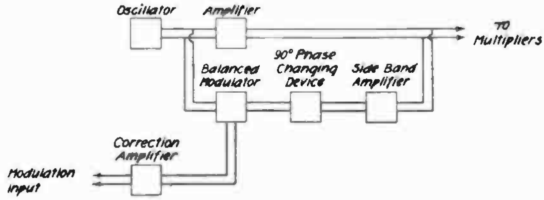


Fig. 1

degrees (or 270 degrees) with respect to the phase of the electromotive force applied to the grids of the balanced modulators. Hence it follows that the phase of the currents existing in the plate circuit of the amplifier of the output of the balanced modulator (at the peak of the modulation voltage) is either ninety degrees or 270 degrees apart from the phase of the current existing in the plate circuit of the amplifier of the

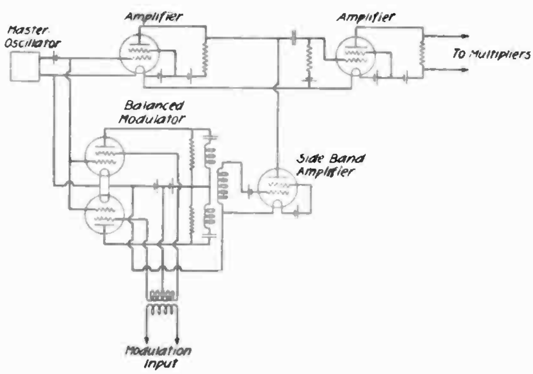


Fig. 2

unmodulated master oscillator current. Therefore the voltages which they develop across the common resistance load will be ninety degrees apart.

The resulting effect on the phase of the voltage developed across the resistance in the plate circuits of these two amplifiers when modulation is applied, compared to the phase of the voltage which would exist there in the absence of modulation will appear from Fig. 3. It will be observed from the vector diagrams that the phase of the voltage across the common resistance load is alternately advanced and retarded by the combination of the modulated and unmodulated components and that the maximum phase shift is given by an angle whose tangent is the sum of the peak values of the two side frequencies divided by the peak value of the unmodulated component. By keeping this angle

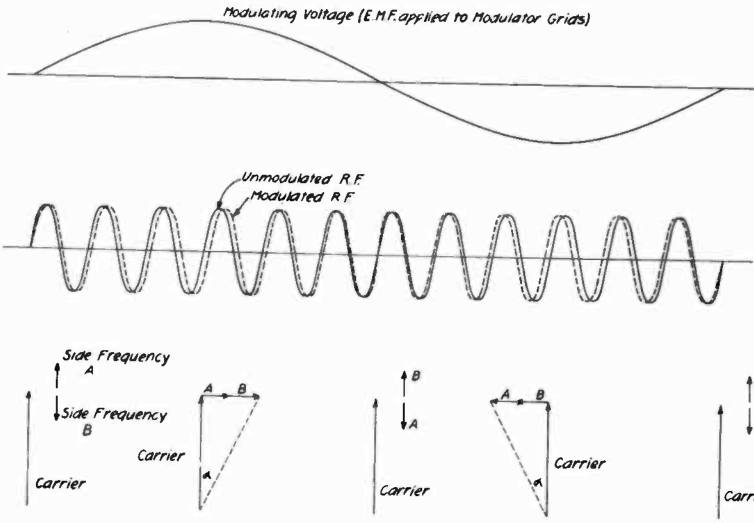


Fig. 3

sufficiently small (not greater than thirty degrees) it may be made substantially proportional to the amplitude of the two side frequencies and hence to the amplitude of the initial modulating current.[7] It will be observed that if the angle through which the phase is shifted be the same for all frequencies of modulation then the rate of increase or de-

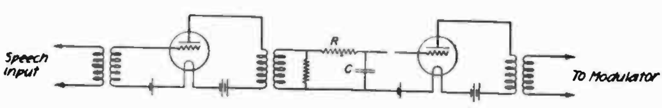


Fig. 4

crease of the angle will be proportional to the frequency of modulation and hence the deviation in frequency of the transmitted wave will be proportional to the frequency of the modulating current. In order to insure a frequency deviation which is independent of the modulation

[7] For the large angular displacements there will be an appreciable change in amplitude of the combined currents at double the frequency of the modulating current. This variation in amplitude is not of primary importance and is removed subsequently by a limiting process.

frequency it is necessary that, for a constant impressed modulating electromotive force, the angle through which the phase is shifted be made inversely proportional to the frequency of the modulating current. This is accomplished by making the amplification of the input amplifier inversely proportional to frequency by means of the correction network shown in Fig. 4. The network consists of a high resistance in series with a capacity whose impedence for the lowest frequency of modulation is relatively small with respect to the series resistance. The voltage developed across the capacity which excites the succeeding amplifier stage is therefore inversely proportional to frequency and hence it follows that the angle through which the current is advanced or retarded becomes directly proportional to the amplitude of the modulating current and inversely proportional to its frequency. The resulting phase shift must be multiplied a great many times before a frequency modulated current which can be usefully employed is produced. This will be clear from an examination of the requirements of a circuit over which it is desired to transmit a frequency range from thirty to 10,000 cycles. Since the lowest frequency is limited to a phase shift of thirty degrees it follows that for 10,000 cycles the phase shift will be but 0.09 degree. The minimum phase shift for 100 per cent modulation of the transmitted wave is roughly forty-five degrees. A frequency multiplication of 500 times is required, therefore, to produce a wave which is fully modulated[8] and capable of being effectively handled by the receiver in the presence of disturbing currents.

Under ordinary conditions this multiplication of frequency can be realized without loss of linearity by a series of doublers and triplers operated at saturation provided the correct linkage circuits between the tubes are employed. Where however the wide band frequency swing which will be described subsequently in this paper is employed unexpected difficulties arise. These also will be dealt with subsequently.

From the foregoing description it will be seen that this method of obtaining frequency modulation consists in producing initially phase modulation in which the phase shift is inversely proportional to the frequency of modulation and converting the phase modulated current into a frequency modulated one by successive multiplications of the phase shift. The frequency stability, of course, is the stability attainable by a crystal controlled oscillator and the symmetry of the deviation may be made substantially perfect by compensating such asymmetrical action in the system as may occur. With the method of phase

[8] One in which the side frequencies are sufficiently large with respect to the carrier to make it possible to produce at the receiver 100 per cent modulation in amplitude, without the use of expedients which affect unfavorably the signal-to-noise ratio.

shifting shown in Fig. 2 there is an asymmetry which is of importance when the frequency of modulation is high compared to the master oscillator frequency. It occurs in the plate transformer of the balanced modulator. The plate circuits of these tubes are substantially aperiodic and consequently the amplitudes of the upper and lower side frequencies are approximately equal and from this it follows that the electromotive forces induced in the secondary are directly proportional to the values of these frequencies. Where the master oscillator frequency is 50,000 cycles and a frequency of modulation of 10,000 cycles is applied, the upper side frequency may be fifty per cent greater than the lower. This inequality may be compensated by a resistance-capacity network introduced subsequent to the point at which the combination of carrier and side frequencies is effected but prior to any point at which loss of linearity of amplitude occurs. The level in the amplifiers ahead of the compensating network must be kept sufficiently low so that the operation of the system is linear. After the side frequencies are equalized amplitude linearity ceases to be of importance.

The performance of transmitters operating on this principle has been in complete accord with expectations. While the arrangements may seem complex and require a large amount of apparatus the complexity is merely that of design, not of operation. The complete arrangement, up to the last few multiplier stages may be carried out most effectively with receiving type tubes, these last multiplier stages consisting of power type pentodes for raising the level to that necessary to excite the usual power amplifiers.

The Receiving System

The most difficult operation in the receiving system is the translation of the changes in the frequency of the received signal into a current which is a reproduction of the original modulating current. This is particularly true in the case of the transmission of high fidelity broadcasting. It is, of course, essential that the translation be made linearly to prevent the generation of harmonics but it must also be accomplished in such a manner that the signaling current is not placed at a disadvantage with respect to the various types of disturbances to which radio reception is subject. In the particular type of translation developed for this purpose which employs the method of causing the changes in frequency to effect changes in amplitude which are then rectified by linear detectors, it is essential that for the maximum deviation of the transmitted frequency there shall be a substantial amplitude modulation of the received wave. At first sight it might appear that 100 per cent or complete modulation would be the ideal, but there are objections to approaching this limit too closely. It will, however, be clear that where the translation is such that only a few per cent amplitude modulation results from the maximum deviation of the frequency of the transmitted wave the receiver is hopelessly handicapped with respect to amplitude disturbances. This is true because even when the level of the voltage applied to the conversion system is kept constant by a current limiting device or automatic volume control there still remains those intervals wherein the incoming disturbances arrive in the proper phase to neutralize the signaling current in the detector, effecting thereby substantially complete modulation of the rectified current or the intervals wherein the disturbing currents themselves effect greater amplitude changes than the signal itself by cross modulation of its frequency.

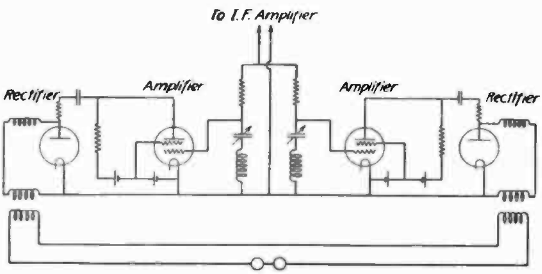


Fig. 5

An arrangement in which linear conversion can be effected without handicapping the system with respect to amplitude disturbances is illustrated diagrammatically in Fig. 5. Two branch circuits each containing resistance, capacity, and inductance in series as shown are connected to the intermediate-frequency amplifier of a superheterodyne at some suitable frequency. One capacity and inductance combination is made nonreactive for one extreme of the frequency band which the signal current traverses and the other capacity and inductance combination is made nonreactive for the other end of the band. The resistances are chosen sufficiently high to maintain the current constant over the frequency range of the band; in fact, sufficiently high to make each branch substantially aperiodic. The reactance characteristics taken across each capacity and inductance combination will be as illustrated in Fig. 6 by curves A and B. Since the resistances in series with the reactance combinations are sufficient to keep the current constant throughout the frequency band it follows that the voltages developed across each of the two combinations will be proportional to their reactances as is illustrated in curves A' and B'. The two voltages are applied respectively to the two equal aperiodic amplifiers, each of which is connected to a linear rectifier. The rectifiers are in series with equal output transformers whose secondaries are so poled that changes in the rectifier currents resulting from a change in the frequency of the received signal produce additive electromotive forces in their secondaries. Since amplifiers and rectifiers are linear the output currents will follow the amplitude variations created by the action of the capacity-inductance combinations. While the variation in reactance is not linear with respect to the change of frequency, particularly where the width of the band is a substantial percentage of the frequency at which the operation takes place, as a practical matter, by the proper choice of values together with shunts of high resistance or reactance

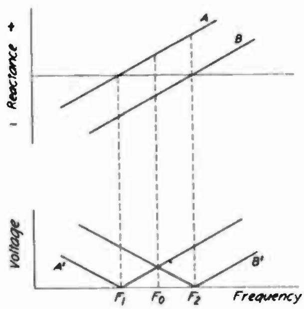


Fig. 6

these characteristics may be rendered sufficiently straight within the working range to meet the severest requirements of high fidelity broadcasting. The operation of the system is aperiodic and capable of effecting 100 per cent modulation if desired, this last depending on the separation of the two nonreactive points with respect to the frequency swing. Generally the setting of the nonreactive frequency points should be somewhat beyond the range through which the frequency is swung.

There is shown in Fig. 7 an alternative arrangement of deriving the signal from the changes in frequency of the received wave which has certain advantages of symmetry over the method just described. In this arrangement a single capacity-inductance combination with the nonreactive point in the center of the frequency band is used and the rectifiers are polarized by a current of constant amplitude derived from the received current. In this way, by properly phasing the polarizing current, which is in effect a synchronous heterodyne, differential rectifying action can be obtained. In Fig. 7 the amplified output of the receiver is applied across the single series circuit consisting of resistance R, capacity C, and inductance L. The reactance of C and L are equal for the mid-frequency point of the band and the reactance curve is as illustrated in *A* of Fig. 8. At frequencies above the nonreactive

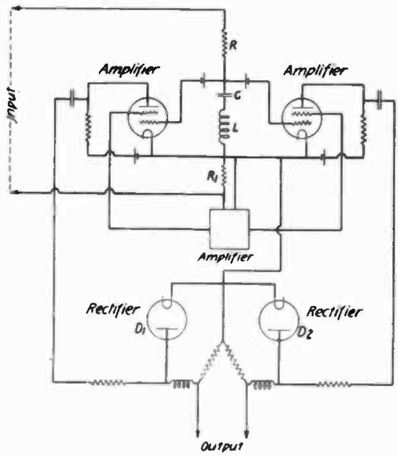


Fig. 7

point the combination acts as an inductance; at frequencies below the nonreactive point as a capacity and the phase of the voltage developed across the combination with respect to the current through it

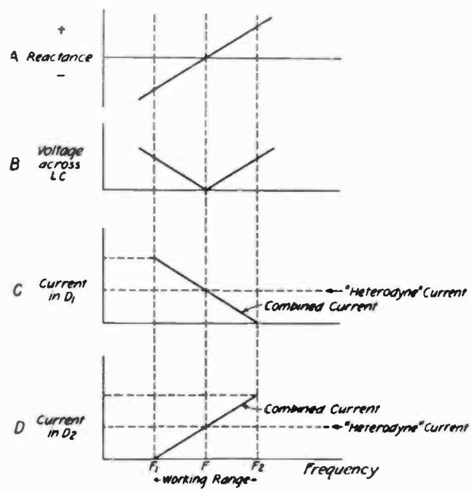


Fig. 8

differs, therefore, by 180 degrees above and below the nonreactive point. Since the current through the circuit is maintained constant over the working range by the resistance *R* and since the resistance of the capacity C and inductance L may be made very low the electromotive force developed across C and L is of the form shown in curve B. This curve likewise represents the variation in voltage with variation in frequency which is applied to the grids of the amplifiers and eventually to the two rectifiers D_1 and D_2.

The heterodyning or polarizing voltage is obtained by taking the drop across the resistance R_1, amplifying it, changing its phase through ninety degrees and applying the amplified voltage to the screen grids of the amplifiers in opposite phase. The characteristic of this amplifying and phase changing system must be flat over the working range. Under these conditions the signaling and heterodyning voltages are exactly in phase in one rectifier and 180 degrees out of phase in the other, and hence for a variable signaling frequency the rectifying characteristics are as shown in curves C and D the detector outputs being cumulatively combined for frequency changes. Adjustment of the relative amplitudes of the signaling and polarizing voltages in the rectifier controls the degree of amplitude modulation produced from 100 per cent down to any desired value.

PART II

With the foregoing description of the instrumentalities for transmitting and receiving frequency modulated waves it is now in order to consider the main object of the paper; the method of reducing disturbances and the practical results obtained by its use.

Method of Reducing Disturbances

The basis of the method consists in introducing into the transmitted wave a characteristic which cannot be reproduced in disturbances of natural origin and utilizing a receiving means which is substantially not responsive to the currents resulting from the ordinary types of disturbances and fully responsive only to the type of wave which has the special characteristic.

The method to be described utilizes a new principle in radio signaling the application of which furnishes an interesting conflict with one which has been a guide in the art for many years; i.e., the belief that the narrower the band of transmission the better the signal-to-noise ratio. That principle is not of general application. In the present method an opposite rule applies.

It appears that the origin of the belief that the energy of the disturbance created in a receiving system by random interference depended on the band width goes back almost to the beginning of radio. In the days of spark telegraphy it was observed that "loose coupling" of the conventional transmitter and receiver circuits produced a "sharper wave" and that interference from lightning discharges, the principle "static" of those days of insensitive and nonamplifying receivers was decreased. Further reduction in interference of this sort occurred when continuous-wave transmitters displaced the spark and when regeneration narrowed the band width of the receiving system. It was observed, however, that "excessive resonance" must not be employed either in telegraphic or more particularly in telephonic signaling or the keying and speech would become distorted. It was concluded in a qualitative way that there was a certain "selectivity" which gave the best results.

In 1925 the matter was placed on a quantitative basis by Carson[9] where in a mathematical treatment of the behavior of selective circuits when subjected to irregular and random interference (with particular reference to "static"), on the basis of certain assumptions, the proposition was established that "if the signaling system requires the transmission of the band of frequencies corresponding to the interval $\omega_2 - \omega_1$ and if the selective circuit is efficiently designed to this end, then the mean square interference current is proportional to the frequency band width $(\omega_2 - \omega_1)/2\pi$.

Hazeltine[10] pointed out that when a detector was added to such a system and a carrier of greater level than the interference currents was present, that for aural reception only those components of the interfering current lying within audible range of the carrier frequency were of importance and that Carson's theory should be supplemented by the use of a factor equal to the relative sensitivity of the ear at different frequencies.

With the discovery of shot effect and thermal agitation noises and the study of their effect on the limit of amplification quantitative relations akin to those enunciated by Carson with respect to static were found to exist.

Johnson,[11] reporting the discovery of the electromotive force due to thermal agitation and considering the problem of reducing the noise in amplifiers caused thereby, points out that for this type of disturbance the theory indicates, as in the Carson theory, that the frequency range of the system should be made no greater than is essential for the proper transmission of the applied input voltage, that where a voltage of constant frequency and amplitude is used one may go to extremes in

[9] J. R. Carson, "Selective circuits and static interference," *Bell Sys. Tech. Jour.*, vol. 4, p. 265, (1925).

[10] L. A. Hazeltine, Discussion on "The shielded neutrodyne receiver," Proc. I.R.E., vol. 14, pp. 408, 409; June, (1926).

[11] J. B. Johnson, "Thermal agitation of electricity in conductors, *Phys. Rev.*, vol. 32, no. 1, July, (1926).

making the system selective and thereby proportionately reducing the noise, but that when the applied voltage varies in frequency or amplitude the system must have a frequency range which takes care of these variations and the presence of a certain amount of noise must be accepted.

Ballantine[12] in a classical paper discussing the random interference created in radio receivers by shot and thermal effects obtained a complete expression for the noise output.[13]

Johnson and Llewellyn,[14] in a paper dealing generally with the limits to amplification, point out that in a properly designed amplifier the limit resides in thermal agitation in the input circuit to the amplifier, that the power of the disturbance in the output of the amplifier is proportional to its frequency range and that this, the only controllable factor in the noise equation, should be no greater than is needed for the transmission of the signal. A similar conclusion is reached in the case of a detector connected to the output of a radio-frequency amplifier and supplied with a signal carrier.

It is now of interest to consider what happens in a linear detector connected to the output of a wide band amplifier which amplifies uniformly the range from 300 to 500 kilocycles. Assume that the amplification be sufficiently great to raise the voltage due to thermal agitation and shot effect to a point sufficient to produce straight-line rectification and that no signal is being received. Under these conditions the frequencies from all parts of the spectrum between 300 and 500 kilocycles beat together to contribute in the output of the detector to the rough hissing tone with which the art is familiar. The spectrum of frequencies in the rectified output runs from some very low value which is due to adjacent components throughout the range beating with one another to the high value of 200 kilocycles caused by the interferences of the extremes of the band.

It is important to note that all parts of the 300- to 500-kilocycle spectrum contribute to the production in the detector output of those frequencies with which we are particularly interested—those lying within the audible range.

[12] Stuart Ballantine, "Fluctuation noise in radio receivers," Proc. I.R.E., vol. 18, pp. 1377–1387; August, (1930).

[13] Ballantine expressed his result as follows: "In a radio receiver employing a square-law detector and with a carrier voltage impressed upon the detector, the audio-frequency noise, as measured by an instrument indicating the average value of the square of the voltage (or current), is proportional to the area under the curve representing the square of the over-all transimpedance (or of the transmission) from the radio-frequency branch in which the disturbance originates to the measuring instrument as a function of frequency and proportional to the square of the carrier voltage."

[14] J. B. Johnson and F. B. Llewellyn, "Limits to amplification," *Trans. A.I.E.E.*, vol. 53, no. 11, November, (1934).

Assume now that an unmodulated signal carrier is received of, for example, 400 kilocycles and that its amplitude is greater than that of the disturbing currents. Under these circumstances an entirely new set of conditions arise. The presence of the 400-kilocycle current stops the rectification of the beats which occur between the various components of the spectrum within the 300- to 500-kilocycle band and forces all rectification to take place in conjunction with the 400-kilocycle carrier. Hence in the output of the rectifier there is produced a series of frequencies running from some low value up to 100 kilocycles. The lowest frequency is produced by those components of the spectrum which lie adjacent to the 400-kilocycle current, the highest by those frequencies[15,16] which lie at the extremity of the band; i.e., 300 and 500 kilocycles, respectively.

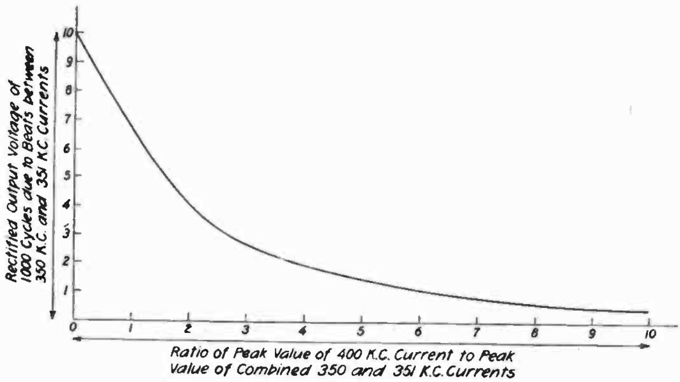


Fig. 9

The characteristics of the rectifiers and the magnitude of some of the effects involved in the above-described action may be visualized by reference to the succeeding figures. The actual demodulation of the beats occurring between adjacent frequency components by the presence of the 400-kilocycle current is shown by the characteristic of Fig. 9, which illustrates what happens to the output voltage of a rectifier produced by beating together two equal currents of 350 and 351 kilocycles, respectively, when a 400-kilocycle current is introduced in the same rectifier and its amplitude progressively increased with respect

[15] It has been pointed out by Ballantine[16] that it is improper to speak of the amplitude of a single component of definite frequency and that the proper unit is the noise per frequency interval. This is, of course, correct, but to facilitate the physical conception of what occurs in this system the liberty is taken of referring to the noise components as though they were of continuous sine wave form. The behavior of the system may be checked by actually introducing from a local generator such components.

[16] "Fluctuation noise in radio receivers," Proc. I.R.E., vol. 18, pp. 1377–1387; August, (1930).

to the amplitude of these two currents. The characteristic was obtained with the arrangement shown in Fig. 10, in which two oscillators of 350 and 351 kilocycles produced currents of equal strength in a linear rectifier, this rectifier consisting of a diode in series with 10,000 ohms resistance. The output of the rectifier is put through a low-pass filter, a voltage divider, and an amplifier. The 400-kilocycle current is intro-

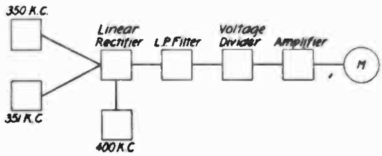


Fig. 10

duced into the rectifier without disturbing the voltage relations of the other two oscillators and the effect on the rectified output voltage observed as the 400-kilocycle current is increased. The purpose of the low-pass filter is to prevent the indicating instrument from responding to the 49- or 50-kilocycle currents produced by the interaction of the 350- and 351-kilocycle currents with the current of 400 kilocycles. The

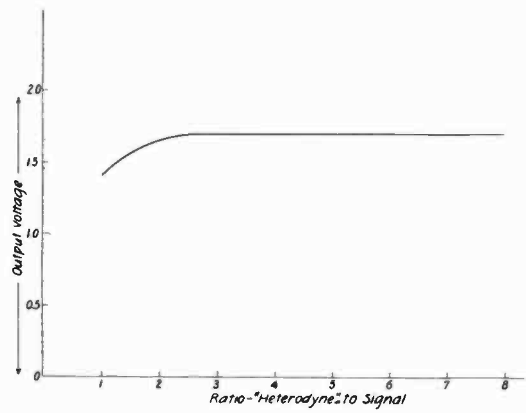


Fig. 11

linearity characteristic of the rectifier is shown in Fig. 11 where the voltage produced by the beats between a current of constant amplitude and one whose amplitude is raised from equality with, to many times the value of, the first current is plotted against the ratio of the two. The linearity of the rectifier is such that after the ratio of the current becomes two to one no further increase in rectifier output voltage results. In fact with the levels used in these measurements when the two currents are equal there is an efficiency of rectification of only about twenty per cent less than the maximum obtained.

It is important to note here that the only frequencies in the spectrum which contribute to the production of currents of audible frequency in the detector output circuit are those lying within audible range of the signal carrier. We may assume this range as roughly from 390 to 410 kilocycles. The frequencies lying beyond these limits beat against the 400-kilocycle carrier and of course are rectified by the detector but the rectified currents which are produced are of frequencies which lie beyond the audible range and produce therefore no effect which is apparent to the ear. It follows that if the signal carrier is somewhat greater in amplitude than the disturbing currents the signal-to-noise ratio for a receiver whose band of admittance covers twice the audible range will be the same as for one whose band width is many times that value. (There are, of course, certain second order effects, but they are of such minor importance that the ear cannot detect them.) The amplitude of the disturbances in the detector output, will vary in accordance as the components of the disturbing currents come into or out of phase with the signal carrier, the rectified or detector output current increasing above and decreasing below the level of the rectified carrier current by an amount proportional to the amplitude of the components of the 300–500-kilocycle band. The reasons for the independence of the signal-to-noise ratio of the band width under the circumstances which have been described should now be apparent. In any event, it can be readily demonstrated experimentally.

It is now in order to consider what happens when a current limiting device is introduced between the output of the amplifier and the detector input. (Assume signal level still above peak noise level.) Two effects will occur. One of the effects will be to suppress in the output circuit of the limiter all components of the disturbing currents which are in phase with, or opposite in phase to, the 400-kilocycle carrier. The other effect will be to permit the passage of all components of the disturbing currents which are in quadrature with the 400-kilocycle current.

Both the above effects are brought about by a curious process which takes place in the limiter. Each component within the band creates an image lying on the opposite side of the 400-kilocycle point whose frequency difference from the 400-kilocycle current is equal to the frequency difference between that current and the original component. The relative phase of the original current in question, the 400-kilocycle current and the image current is that of phase modulation—that is, at the instant when the original component and its image are in phase with each other, the 400-kilocycle current will be in quadrature with them both and at the instant that the 400-kilocycle current is in phase with one of these two frequencies, it will be out of phase with the other.

The relation (obtained experimentally) between the amplitudes of the original current and the image is illustrated by the curve of Fig. 12,

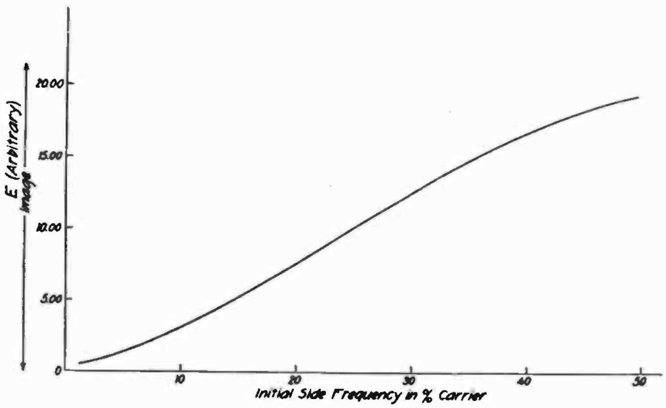


Fig. 12

which shows the relation between the amplitude of a 390-kilocycle current introduced into a limiter along with the 400-kilocycle current and the resulting 410-kilocycle image in terms of percentage amplitude of the 400-kilocycle current. It will be obvious from the curve that in the region which is of interest—that is, where the sidefrequencies are smaller than the mid-frequency—that the effect is substantially linear.

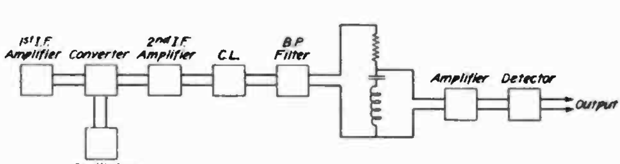


Fig. 13

With the above understanding of what takes place in the current limiter it is now in order to consider what happens when a selective system as illustrated in Fig. 13 is interposed between the limiter and the detector. (The band-pass filter is for the purpose of removing limiter harmonics.) A rough picture of what occurs may be had by considering a single component of the interference spectrum. Suppose this component to be at 390 kilocycles and that by the action already explained it has created its image at 410 kilocycles. These two frequencies are equal in amplitude and so phased with respect to each other and with respect to the 400-kilocycle carrier that no amplitude change results.

Assume now that the selective system has the characteristic *MN* which as shown in Fig. 14 is designed to give complete modulation for a ten-kilocycle deviation of frequency. Since at 390 kilocycles the reactance across the capacity-inductance combination is zero and at 410 kilocycles double what it is at 400 kilocycles it follows that the 390-kilocycle component becomes equal to zero but the ratio of the 410-kilocycle component to the 400-kilocycle carrier is doubled; that

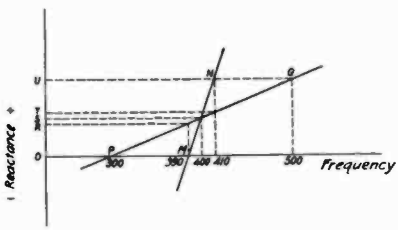


Fig. 14

is, it is twice as great as is the ratio in the circuits preceding the selective system. The change in amplitude, therefore, becomes proportional to *OU*. Therefore in combination with the 400-kilocycle carrier a variation in amplitude is produced which is substantially identical with that which would be obtained were the current limiter removed and the selective system replaced by an aperiodic coupling of such value that the same detector level were preserved.

Now consider what occurs when a selective system having the characteristic such as *PQ* and requiring a deviation of 100 kilocycles to produce full modulation is employed instead of one such as *MN*, where a ten-kilocycle deviation only is required. Assume the same conditions of interference as before. The 400-kilocycle voltage applied to the rectifier will be the same as before, but the *relative amplitudes of the 390- and 410-kilocycle voltages will only be slightly changed.* The 410-kilocycle voltage will be increased from a value which is proportional to *OS* to one which is proportional to *OT* and the 390-kilocycle voltage will be reduced from a value proportional to *OS* to one proportional to *OR*. The difference in value of the two frequencies will be proportional to the difference between *OS* and *OT* or *RT*, and the change in amplitude produced by their interaction with the 400-

kilocycle current will be likewise proportional to RT. The reduction in the amplitude of the disturbance as measured in the detector output by the use of a 200-kilocycle wide selective system as compared to the use of one only twenty kilocycles wide is therefore the ratio RT/OU. In this case it is ten per cent. The power ratio is the square of this or one per cent.

The above reasoning holds equally well if a balanced rectifying system is used where the characteristics of the selective system are as shown in Fig. 15. The output of the system insofar as voltages resulting from changes in frequency are concerned is the sum of outputs of the two sides of the balance.

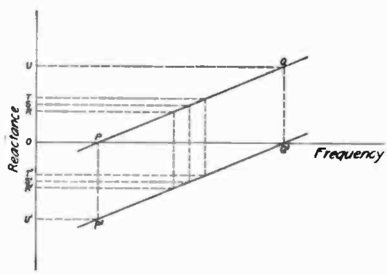


Fig. 15

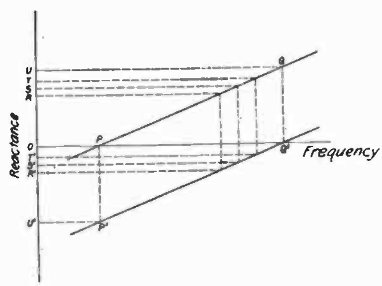


Fig. 16

It is of course clear that disturbing currents lying farther from the 400-kilocycle point than the ten-kilocycle limit will, by interaction with the 400-kilocycle current, produce larger values of rectified current than those lying within that band. *But the rectified currents produced in the detector output by those components of frequency which lie at a greater than audible frequency distance from the 400-kilocycle current will be beyond the audible range and hence will produce no disturbance which is audible.* (It is generally advisable to eliminate them from the audio amplifier by a low-pass filter to prevent some incidental rectification in the amplifier making their variations in amplitude audible.)

It remains only to consider what happens when the frequency of the 400-kilocycle current is varied in accordance with modulation at the transmitter. It is clear from Fig. 14 that when the selective system has the characteristic MN that a deviation of 10,000 cycles will produce complete modulation of the signal or a change in amplitude proportional to OU. Similarly, when the characteristic is according to the curve PQ it is clear that a 100,000-cycle deviation is required to produce complete modulation, which is likewise proportional to the same value OU. As the signal current is swung back and forth over the range of frequencies between 300 and 500 kilocycles the band of frequencies from which the audible interference is derived continually changes, the band progressively lying about ten kilocycles above and ten kilocycles below what we may call the instantaneous value of the frequency of the signal. The effect is illustrated by Fig. 16 and from this it will be seen that the amplitude of the disturbances in the output circuit of the rectifiers, which is proportional to the sum of RT and $R'T'$ will be constant. This will be true where the ratio of the amplitude of the signal to the disturbing currents is sufficiently large—where this condition does not exist then there are certain other effects which modify the results, but these effects will only be of importance at the limits of the practical working range.

Comparison of Noise Ratios of Amplitude and Frequency Modulation Systems

From the foregoing description it will be clear that as between two frequency modulation systems of different band widths the signal-to-noise power ratio in the rectified output will vary directly as the square of the band width (provided the noise voltage at the current limiter is less than the signaling voltage). Thus doubling the band width produces an improvement of 4 to 1 and increasing it tenfold an improvement of 100 to 1.

The comparison of relative noise ratios of amplitude and frequency modulation systems cannot be made on so simple a basis as there are a number of new factors which enter, particularly when the comparison is viewed from the very practical aspect of how much greater power must be used with an amplitude modulated transmitter than with a frequency modulated one. If the academic comparison be made between a frequency modulated system having a deviation of ten kilocycles and an amplitude modulated one of similar band width and the *same carrier* level (also same fidelity), it will be found that the signal-to-noise voltage ratio as measured by a root-mean-square meter will favor the frequency modulation system by about 1.7 to 1, and that the corresponding power ratio will be about 3 to 1. This improvement is due to the fact that in the frequency modulation receiver it is only those noise components which lie at the extremes of the band; viz., ten kilocycles away from the carrier which, by interaction with the carrier (when unmodulated) can produce the same amplitude of rectified current as will be produced by the corresponding noise component in the amplitude modulated receiver.

Those components which lie closer to the carrier than ten kilocycles will produce a smaller rectified voltage, the value of this depending on their relative distance from the carrier. Hence the distribution of energy in the rectified current will not be uniform with respect to frequency but will increase from zero at zero frequency up to a maximum at the limit of the width of the receiver, which is ten kilocycles in the present case. The root-mean-square value of the voltage under such a distribution is approximately 0.6 of the value produced with the uniform distribution of the amplitude receiver.

Similarly in comparing an amplitude modulation system arranged to receive ten-kilocycle modulations and having, of course, a band width of twenty kilocycles, with a 100-kilocycle deviation frequency

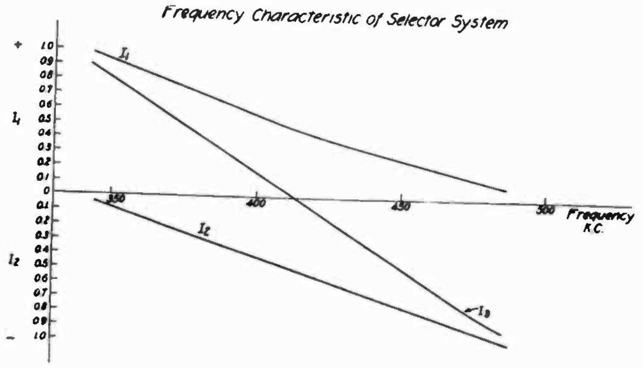


Fig. 17

modulation system (same carrier level and same fidelity) there will be an improvement in noise voltage ratio of

$$1.7 \times \frac{\text{deviation}}{\text{audio-frequency range}} \quad \text{or} \quad 1.7 \times \frac{100}{10} = 17.$$

The above comparisons have been made on the basis of equal carrier. The practical basis of comparison between the two is that of half carrier for the amplitude modulation and full carrier for the frequency modulation system. This results in about the equivalent amount of power being drawn from the mains by the two systems. On this basis the voltage improvement becomes thirty-four and the signal-to-noise power ratio 1156. Where the signal level is sufficiently large with respect to the noise it has been found possible to realize improvements of this order.

The relative output signal-to-noise ratios of an amplitude modulation system fifteen kilocycles wide (7.5-kilocycle modulation frequency) and a frequency modulation system 150 kilocycles wide (75-kilocycle deviation) operating on forty-one megacycles have been compared on the basis of equal fidelity and half carrier for amplitude modulation.

The characteristic of the selective system for converting frequency changes to amplitude changes, which was used, is shown in Fig. 17. The variation of the output signal-to-noise ratio with respect to the corresponding radio-frequency voltage ratio is illustrated in Fig. 18. The curves show that where the radio-frequency peak voltage of the noise measured at the current limiter is less than ten per cent of the

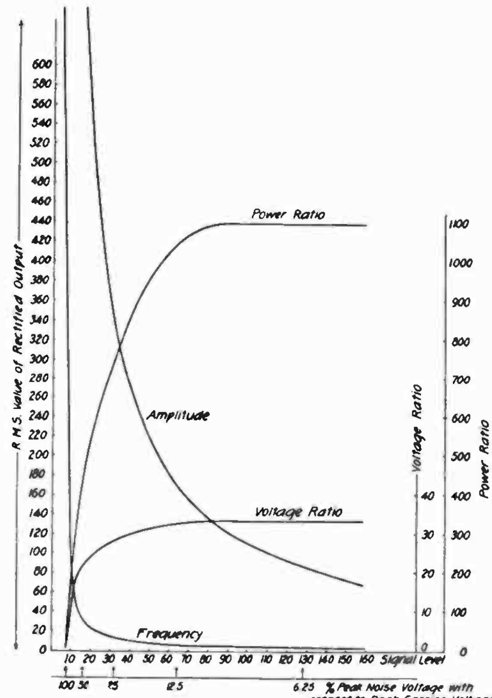


Fig. 18

signal peak voltage then the energy of the disturbance in the rectified output will be reduced by a factor which is approximately 1100 to 1. When the peak radio-frequency noise voltage is twenty-five per cent of the signal peak voltage then the energy of the disturbance in the rectified output has been reduced to about 700 to 1, and when it is fifty per cent the reduction of the disturbance drops below 500 to 1. Finally when the noise and signal peak voltages become substantially equal the improvement drops to some very low value. While it is unfortunate, of course, that the nature of the effect is such that the amount of noise reduction becomes less as the noise level rises with respect to the signal, nevertheless this failing is not nearly so important as it would appear. In the field of high fidelity broadcasting a signal-to-noise voltage ratio of at least 100 to 1 is required for satisfactory reception. It is just within those ranges of noise ratios which can be reduced to this low level that the system is most effective.

The arrangements employed in obtaining these characteristics and the precautions which must be observed may perhaps be of interest. As it was obviously impracticable to vary the power of a transmitter over the ranges required or to eliminate the fading factor except over short periods of time an expedient was adopted. This expedient consisted in tuning the receiver to the carrier of a distant station, determining levels and then substituting for the distant station a local signal generator, the distant station remaining shut down except as it was

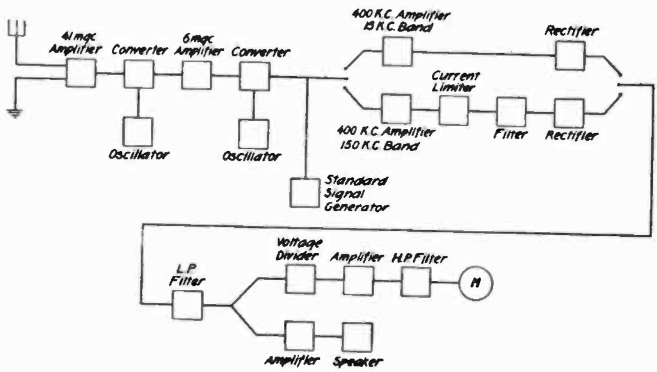


Fig. 19

called upon to check specific points on the curve. Observations were taken only when the noise was due solely to thermal agitation and shot effect.

Fig. 19 shows the arrangement of apparatus. The receiver was a two-intermediate-frequency superheterodyne with provision for using either a narrow band second intermediate amplifier with the amplitude modulation system or a wide band amplifier with the frequency modulation system. The band width of the amplitude modulation system was fifteen kilocycles or twice the modulation frequency range. The band width of the frequency modulation receiver was 150 kilocycles or twice the frequency deviation. Provision was made for shifting from one intermediate amplifier to the other without disturbing the remainder of the system. The forty-one-megacycle circuits and the first intermediate amplifier circuits were wide enough to pass the frequency swing of 150 kilocycles. Identical detection systems were used, the frequency modulation detector being preceded by a selective system for translating changes in frequency into changes in amplitude. The output circuits of the detectors were arranged to be connected alternately to a 7500-cycle low-pass filter with a voltage divider across its output. An amplifier with a flat characteristic over the audible range and a root-mean-square meter connected through a high-pass, 500-cycle filter provided the visual indication.

The standard signal was introduced into the input of the two branches of the second intermediate-frequency stage at 400 kilocycles. As long as the receiver is linear between the antenna and the point at which the standard signal is introduced it is immaterial whether the signal be of forty-one megacycles, six megacycles, or 400 kilocycles. This has been checked experimentally but 400 kilocycles was chosen on account of the greater accuracy of the signal generator on low frequencies.

The relative noise levels to be compared varied over such ranges that lack of linearity had to be guarded against and readings were made by bringing the output meter to the same point on the scale each time by adjustment of the voltage divider, and obtaining the relative voltages directly from the divider.

Two other precautions are essential. The absolute value of the noise voltage on the frequency modulation system becomes very low for high signal levels. If the voltages due to thermal agitation and shot effect are to be measured rather than those due to the power supply system the output meter must be protected by a high-pass filter of high attentuation for the frequencies produced by the power system. The cutoff point should be kept as low as possible since because of the difference in the distribution of energy in the rectified outputs of frequency and amplitude modulation receivers already referred to there is a certain error introduced by this filter which is small if the band width excluded by the filter is small but which can become appreciable if too much of the low-frequency part of the modulation frequency range be suppressed.

A second precaution is the use of a low-pass filter to cut off frequencies above the modulation range. Because of the wide band passed by the amplifiers of the frequency modulation part of the system there exists in the detector output rectified currents of frequencies up to seventy-five kilocycles. The amplitude of these higher frequencies is much greater than those lying within the audible range. The average detector output transformer will readily pass a substantial part of these superaudible frequencies which then register their effect upon the output meter although they in no way contribute to the audible disturbance.

The procedure which was followed in making the measurements we are considering consisted in tuning the receiver to the distant transmitter and adjusting the two detector levels to the same value for the respective carrier levels to be employed. This was done by cutting the carrier in half at the transmitter when the amplitude modulation detector level was being set and using full carrier for the adjustment of the frequency modulation detector. Each system was then modulated seventy-five per cent and output voltages checked against each other. If they were equal the modulation was removed and the relative noise voltages measured for the respective carrier levels. This gave the first point on the curve. The transmitter was then shut down and a local carrier introduced which gave the same level in the 400-kilocycle intermediate amplifier circuits as the half carrier distant signal. This level was directly ascertainable from the rectified detector current in the amplitude modulation system. From this point on the procedure was entirely within the control of the receiving station. The noise ratios could be compared at any signal level by adjusting the voltage introduced by the signal generator to any fraction of that of the distant signal, bringing the level in the amplitude modulation detector up to the same original value by adjustment of the amplification of the second intermediate amplifier (the frequency modulation detector stays at its point of reference because of the current limiter) and comparing the two output voltages. The level of the detector in the amplitude modulation receiver was of course set with the half carrier value of the signal generator and the output voltage measured at that level. The output voltage of the frequency modulation system was measured when twice that voltage was applied.

It is important to keep in mind just what quantities have been measured and what the curves show. The results are a comparison between the relative noise levels in the two systems (root-mean-square values) *when they are unmodulated.* In both an amplitude and in a frequency modulation receiver the noise during modulation may be greater than that obtained without modulation. In the frequency modulation receiver two principal sources may contribute to this increase, one of which is of importance only where the band for which the receiver is designed is narrow, the other of which is common to all band widths. If the total band width of the receiver is twenty kilocycles and if the deviation is, for example, ten kilocycles, then as the carrier frequency swings off to one side of the band, it approaches close to the limit of the filtering system of the receiver. Since the sides of the filter are normally much steeper than the selective system employed to convert the changes in frequency into amplitude variations and since the frequency of the signaling current will have approached to within the range of good audibility of the side of the filter a considerable increase in both audibility and amplitude of the disturbance may occur, caused by the sides of the filter acting as the translating device. This effect is obviously not of importance where a wider frequency swing is employed.

The other source of noise which may occur when the signal frequency swings over the full range is found in systems of all band widths. It was first observed on an unmodulated signal when it was noted that swinging the intermediate frequency from the mid-point to one side or the other by adjustment of the frequency of the first heterodyne produced an increase in the amplitude and a change in the character of the noise. The effect was noted on a balanced detector system and at first it was attributed to the destruction of the amplitude balance as one detector current became greater than the other. Subsequently when it was noted that the increase in the noise was produced by the detector with the smaller current and that the effect was most pronounced when the signal level was relatively low, the explanation became apparent. As long as the signal frequency was set at the mid-point of the band its level in the detector was sufficiently large to prevent the production of audible beats between the noise components lying respectively at the two ends of the band where the reactance of the selective systems is a maximum.

When however the signal frequency moves over to one side of the band the amplitude of the voltage applied to one of the detectors progressively decreases, approaching zero as the frequency coincides with the zero reactance point of the selective system. The demodulating effect of the signaling current therefore disappears and the noise components throughout the band, particularly those at the other side of it, are therefore free to beat with each other. The noise produced is the characteristic one obtained when the high-frequency currents caused by thermal agitation and shot effect are rectified in a detector without presence of a carrier. The effect is not of any great importance on the ordinary working levels for simplex operation, although it may become so in multiplex operation. It indicates, however, that where the signal-to-noise level is low, complete modulation of the received signal by the conversion system is not desirable and that an adjustment of the degree of modulation for various relative noise levels is advantageous.

In the course of a long series of comparisons between the two systems a physiological effect of considerable importance was noted. It was observed that while a root-mean-square meter might show the same reading for two sources of noise, one derived from an amplitude modulation, and the other from a frequency modulation receiver (both of the same fidelity) that the disturbance perceived by the ear was more annoying on the amplitude modulation system. The reason for this is the difference in the distribution of the noise voltage with respect to frequency in the rectified output currents of the two systems, the distribution being substantially uniform in the amplitude system but proportional to frequency in the frequency modulation system. Hence in the latter there is a marked absence of those frequencies which lie in the range to which the ear is the most sensitive. With most observers this difference results in their appraising a disturbance produced in the speaker by an amplitude modulation system as the equivalent of one produced therein by a frequency modulation system of about 1.5 times the root-mean-square voltage although of course the factor varies considerably with the frequency range under consideration and the characteristic of the individual's aural system.

On account of this difference in distribution of energy the correct method of procedure in making the comparison is that given in the article by Ballantine,[16] but lack of facilities for such determinations made necessary the use of a root-mean-square meter for the simultaneous measurement of the entire noise frequency range. The increase in noise voltage per frequency interval with the frequency may be readily demonstrated by means of the ordinary harmonic analyzer of the type now so generally used for the measurement of distortion. Because of the extremely narrow frequency interval of these instruments it is not possible to obtain sufficient integration to produce stable meter readings and apparatus having a wider frequency interval than the crystal filter type of analyzer must be used. The observation of the action of one of these analyzers will furnish convincing proof that peak voltmeter methods must not be used in comparing the rectified output currents in frequency and amplitude modulation receivers.

All the measurements which have been heretofore discussed were taken under conditions in which the disturbing currents had their origin in either thermal agitation or shot effect, as the irregularity of atmospheric disturbances or those due to automobile ignition systems were too irregular to permit reproducible results. The curves apply generally to other types of disturbances provided the disturbing voltage is not greater than that of the signal. When that occurs a different situation exists and will be considered in detail later.

There are numerous second order effects produced, but as they are of no great importance consideration of them will not be undertaken in the present paper.

The New York-Westhampton and Haddonfield Tests

The years of research required before field tests could even be considered were carried out in the Marcellus Hartley Research Laboratory at Columbia University. Of necessity both ends of the circuit had to be under observation simultaneously and a locally generated signal was used. The source of signal ultimately employed consisted of a standard signal generator based upon the principle of modulation already described and capable of giving 150,000 cycles swing on forty-four megacycles. The generator was also arranged to give amplitude modulated signals. Suitable switching arrangements for changing rapidly from frequency to amplitude modulation at either full or half carrier were set up and a characteristic similar to that of Fig. 18 ultimately obtained.

A complete receiving system was constructed and during the Winter of 1933–1934 a series of demonstrations were made to the executives and engineers of the Radio Corporation of America. That wholly justifiable suspicion with which all laboratory demonstrations of "static eliminators" should be properly regarded was relieved when C. W. Horn of the National Broadcasting Company placed at the writer's disposal a transmitter in that company's experimental station located on top of the Empire State Building in New York City. The transmitter used for the sight channel of the television system delivered about two kilowatts of power at forty-four megacycles to the antenna and it was the one selected for use. This offer of Mr. Horn's greatly facilitated the practical application of the system as it eliminated the necessity of transmitter construction in a difficult field and furnished the highly skilled assistance of R. E. Shelby and T. J. Buzalski, the active staff of the station at that time. Numerous difficulties, real and imaginary, required much careful measurement to ascertain their presence or absence and the relative importance of those actually existing. The most troublesome was due to the position of the transmitter, which is located on the eighty-fifth floor of the building and is connected by a concentric transmission line approximately 275 feet long with a vertical dipole antenna about 1250 feet above ground. Investigation of the characteristics of this link between transmitter and antenna showed it to be so poorly matched to the antenna that the resulting standing waves attained very large amplitude. The problem of termination afforded peculiar difficulties because of the severe structural requirements of the antenna above the roof and of the transmission line below it. It was however completely solved by P. S. Carter of the R.C.A. Communications Company in a very beautiful manner, the standing waves being practically eliminated and the antenna broadened beyond all requirements of the modulating system contemplated. With the transmitter circuits no difficulty was encountered at this time. The frequency of the system was ordinarily controlled by a master oscillator operating at 1733 kilocycles which was multiplied by a series of doublers and a tripler to forty-four megacycles. The multiplier and amplifier circuits were found to be sufficiently broad for the purposes of the initial tests.

The crystal control oscillator was replaced by the output of the modulation system shown in Fig. 20 in which an initial frequency of 57.33 kilocycles was multiplied by a series of doublers up to the input frequency of the transmitter of 1733 kilocycles. It was found possible to operate this apparatus as it is shown installed in the shielded room

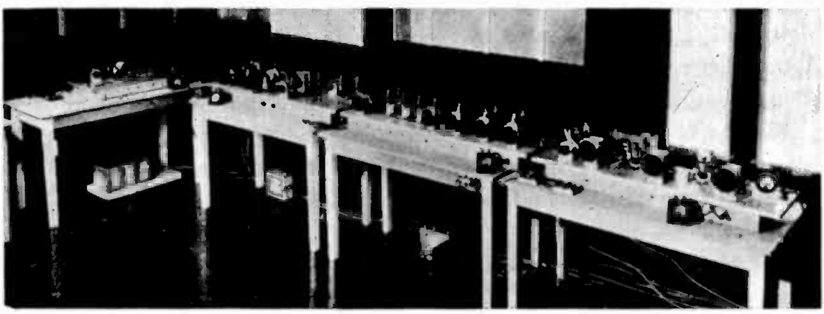


Fig. 20

of the television studio at the Empire State station as the shielding furnished ample protection against the effects of the high power stages of the transmitter located some seventy-five feet away.

The receiving site selected was at the home of George E. Burghard at Westhampton Beach, Long Island, one of the original pioneers of amateur radio, where a modern amateur station with all facilities, including those for rigging directive antennas, were at hand. Westhampton is about sixty-five miles from New York and 800 or 900 feet below line of sight.

The installation is illustrated in Figs. 21 and 22 which show both frequency and amplitude modulation receivers and some of the measuring equipment for comparing them. The frequency modulation receiver consisted of three stages of radio-frequency amplification (at forty-one megacycles) giving a gain in voltage of about 100. This frequency was heterodyned down to six megacycles where an amplification of about 2000 was available and this frequency was in turn heterodyned down to 400 kilocycles where an amplification of about 1000 could be realized. Two current limiting systems in cascade each with a separate amplifier were used. At the time the photograph was taken the first two radio-frequency stages had been discarded.

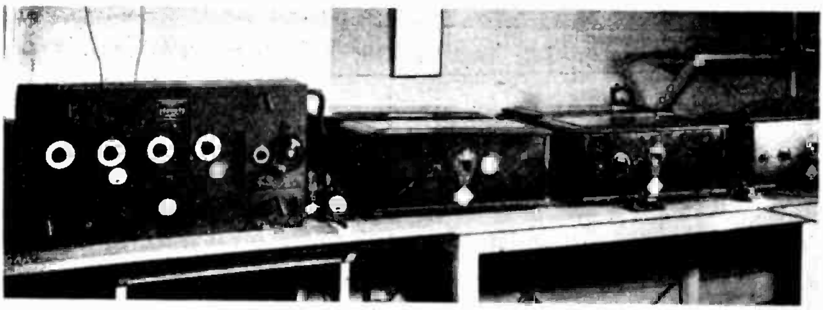


Fig. 21

The initial tests in the early part of June surpassed all expectations. Reception was perfect on any of the antennas employed, a ten-foot wire furnishing sufficient pickup to eliminate all background noises. Suc-

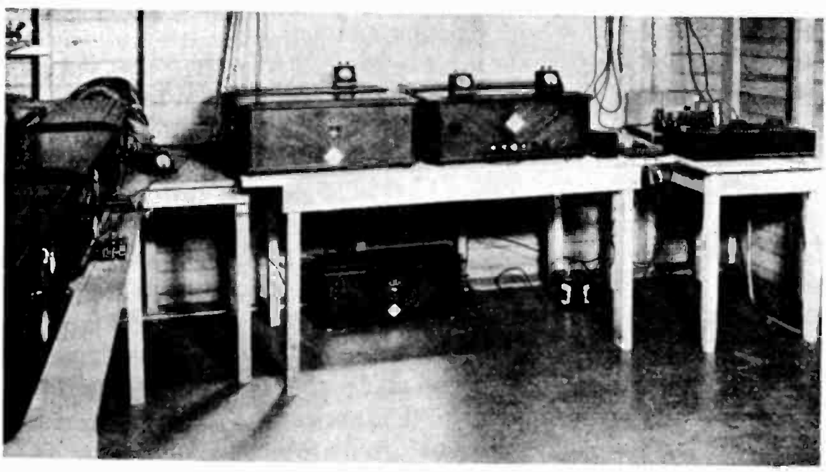


Fig. 22

cessive reductions of power at the transmitter culminated at a level subsequently determined as approximately twenty watts. This gave a signal comparable to that received from the regular New York broadcast stations (except WEAF, a fifty-kilowatt station approximately forty miles away).

The margin of superiority of the frequency modulation system over amplitude modulation at forty-one megacycles was so great that it was at once obvious that comparisons of the two were principally of academic interest.

The real question of great engineering and economic importance was the comparison of the ultra-short-wave frequency modulation system with the existing broadcast service and the determination of the question of whether the service area of the existing stations could not be more effectively covered than at present. The remainder of the month was devoted to such a comparison. With the Empire State transmitter operating with approximately two kilowatts in the antenna, at all times and under all conditions the service was superior to that provided by the existing fifty-kilowatt stations, this including station WEAF. During thunderstorms, unless lightning was striking within a few miles of Westhampton, no disturbance at all would appear on the system, while all programs on the regular broadcast system would be in a hopeless condition. Background noise due to thermal agitation and tube hiss were likewise much less than on the regular broadcast system.

The work at Westhampton demonstrated that in comparing this method of transmission with existing methods two classes of services and two bases of comparisons must be used. It was found that the only type of disturbance of the slightest importance was that caused by the ignition systems of automobiles, where the peak voltage developed by the interference was greater than the carrier level. In point-to-point communication this difficulty can be readily guarded against by proper location of the receiving system, and then thermal agitation and shot effect are the principal sources of disturbance; lightning, unless in the immediate vicinity, rarely producing voltages in excess of the carrier level which would normally be employed to suppress the thermal and shot effects. Under these conditions the full effect of noise suppression is realized and comparisons can be made with precision by means of the method already described in this paper. An illustration of the practical accomplishment of this occurred at Arney's Mount, the television relay point between New York and Camden of the Radio Corporation of America. This station is located about sixty miles from the Empire State Building and the top of the tower is only a few feet below line of sight. It is in an isolated spot and the noise level is almost entirely that due to the thermal and shot effects. It was noted by C. M. Burrill of the RCA Manufacturing Company who made the observations at Arney's Mount that with fifty watts in the antenna frequency modulated (produced by a pair of UX 852 tubes), a signal-to-noise ratio of the same value as the two-kilowatt amplitude modulation transmitter (eight-kilowatt peaks) was obtained.

The power amplifier and the intermediate power amplifier of the frequency modulation transmitter is shown in Fig. 23. The signal with fifty watts output would undoubtedly have had a better noise ratio than the two-kilowatt amplitude modulation system had full deviation of seventy-five kilocycles been employed, but on the occasion it was not possible to use a deviation of greater than twenty-five kilocycles. It was also observed at the same time that when the plate voltage on the power amplifier was raised to give a power of the order of 200 watts in the antenna a better signal-to-noise ratio was obtained than

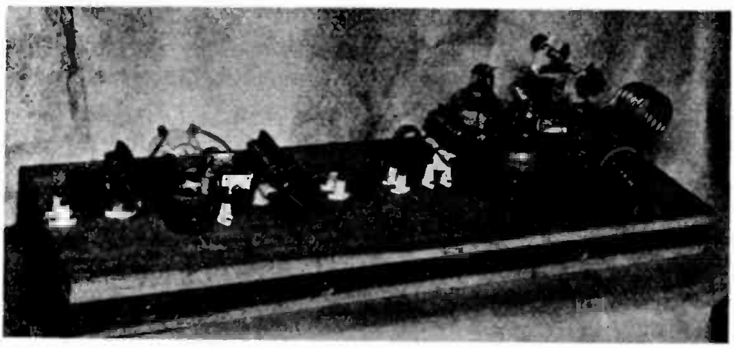


Fig. 23

that which could be produced by the two-kilowatt amplitude modulation. A casual comparison of the power amplifier stages of the frequency modulation transmitter shown in Fig. 23 with the water-cooled power amplifier and modulation stages of the Empire State transmitter is more eloquent than any curves which may be shown herein.

In the broadcast service no such choice of location is possible and a widely variable set of conditions must be met. Depending on the power at the transmitter, the elevation of the antenna, the contour of the intervening country, and the intensity of the interference there will be a certain distance at which peaks of ignition noise become greater than the carrier. The irregularity and difficulty of reproduction of these disturbances require a different method of comparison which will be hereinafter described.

As the site at Westhampton, which was on a section of the beach remote from man-made static, was obviously too favorable a site, a new one was selected in Haddonfield, New Jersey, and about the end of June the receiving apparatus was moved there and erected at the home of Harry Sadenwater. Haddonfield is located about eighty-five miles from New York in the vicinity of Camden, New Jersey, and is over 1000 feet below line of sight of the top of the Empire State Building in New York. Although the field strength at Haddonfield was considerably below that at Westhampton Beach, good reception was obtained almost immediately, the sole source of noise heard being ignition noise

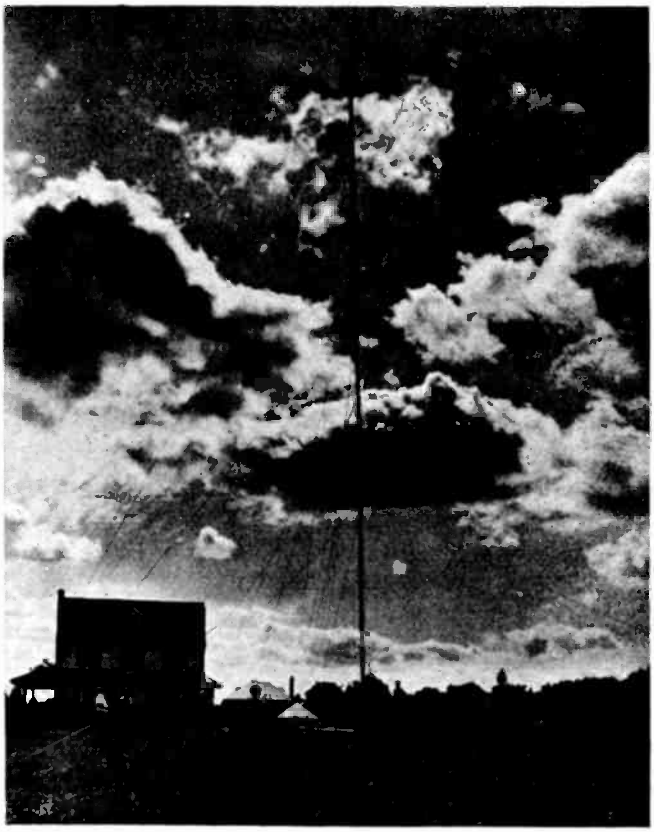


Fig. 24

from a few types of cars in the immediate vicinity of the antenna, or lightning striking within a few miles of the station. At this distance fading made its appearance for the first time, a rapid flutter varying in amplitude three- or four-to-one being frequently observable on the meters. The effect of it was not that of the selective fading so well known in present-day broadcasting. Very violent variations as indicated by the meters occurred without a trace of distortion being heard in the speaker. During a period of over a year in which observations have been made at Haddonfield, but two short periods of fading have been observed where the signal sank to a level sufficient to bring in objectionable noise, one of these occurring prior to an insulation failure at the transmitter.

It is a curious fact that the distant fading, pronounced though it may be at times, is not so violent as that which may be encountered at a receiving station located within the city limits of New York. The effect, which appears to be caused by moving objects in the vicinity of the receiving antenna, causes fluctuations of great violence. In was ap-

Fig. 25

parently first observed by L. F. Jones of the RCA Manufacturing Company within a distance of half a mile of the Empire State transmitter. It occurs continually at Columbia University located about four miles from the Empire State transmitter but no injurious effect on the quality of transmission has ever been noted.

While at first, because of the lower field strength at Haddonfield and the greater prevalence of ignition disturbances, the superiority over the regular broadcast service was not so marked as at Westhampton Beach, the subsequent improvements which were instituted at both transmitting and receiving ends of the circuit have more than offset the lower signal level. Some idea of their extent may be gained by comparison of the initial and final antenna structures. Fig. 24 shows the original antenna during course of erection, a sixty-five foot mast bearing in the direction of New York permitting the use of an eight-wave length sloping wire of very useful directive properties. Fig. 25

shows the final form on which the results are now much better than were originally obtained with the directional wire.

During the past summer, which was marked by thunderstorms of great severity in the vicinity of Philadelphia, it was the exception when it was agreeable or even possible to listen to the nightly programs of the regular broadcast service from the fifty-kilowatt New York stations. In some of the heaviest storms when lightning was striking within the immediate vicinity of the antenna, so close in fact that the lead-in was sparking to a near-by water pipe, perfectly understandable speech could be received on the frequency modulation system, although the disturbance was sufficient to cause annoyance on a musical program; but these periods seldom lasted more than fifteen minutes when the circuit would again become quiet. On numerous occasions the Empire State signal was better than that of the fifty-kilowatt Philadelphia station WCAU located at a distance of twenty miles from Haddonfield. Likewise during periods of severe selective side-band fading in the broadcast band which occurs even from station WJZ at Bound Brook, New Jersey, some sixty miles away, no signs of this difficulty would appear on the ultra-high-frequency wave.

Some of the changes which contributed to the improvement during the past year may be of interest. The introduction of the Thompson-Rose tube permitted the radio-frequency amplification required at forty-one megacycles to be accomplished with one stage and with considerable improvement of signal-to-noise ratio. It had a further interesting result. The tubes previously used for amplifying at this frequency were those developed by the Radio Corporation for the ultra-short-wave interisland communication system in the Hawaiian Islands. On account of the relatively low amplification factor of these tubes the shot effect in the plate circuit of the first tube exceeded the disturbances due to thermal agitation in the input circuit of that tube by a considerable amount. With the acorn type tube, however, the situation is reversed, the thermal noise contributing about seventy-five per cent of the rectified output voltage.

It should be noted here by those who may have occasion to make this measurement on a frequency modulation system that it cannot be made in the ordinary way by simply mis-tuning the input circuit to the first tube. To do so would remove the carrier from the current limiter and be followed by a roar of noise. The measurement must be made with a local signal of the proper strength introduced into one of the intermediate-frequency amplifiers. Under these conditions the antenna may be mis-tuned without interfering with the normal action of the limiter and the relative amounts of noise due to the two sources may readily be segregated.

Considerable trouble was caused during the early stages of the experiments by an order of the Federal Radio Commission requiring the changing of the frequency of the Empire State transmitter from forty-four to forty-one megacycles; this necessitating the realignment of the large number of interstage transformers in the modulating equipment shown in Fig. 20 and also the retermination of the antenna. It, however, led to the application of the idea inherent in superheterodyne design.

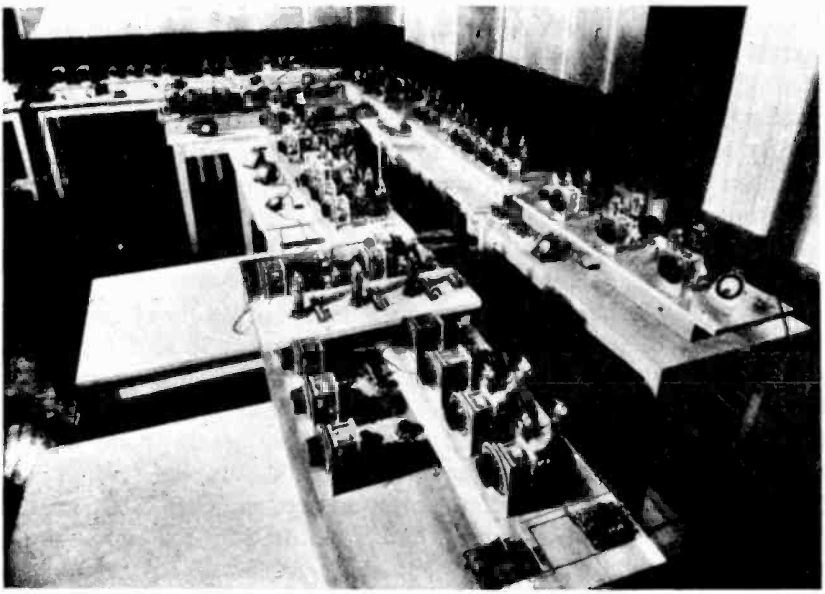


Fig. 26

While the circuits of the old modulator were temporarily modified and work carried on, a new modulation system was designed standardizing on an initial frequency of 100 kilocycles which was then multiplied by a series of doublers up to 12,800 kilocycles. By means of a local oscillator this frequency was heterodyned down to 1708 kilocycles, the new value of input frequency to the transmitter required to produce forty-one megacycles in the antenna. Any future changes in wave length can be made by merely changing the frequency of this second oscillator. The frequencies chosen were such that a deviation of 100 kilocycles could be obtained without difficulty, because of the extra number of frequency multiplications introduced. Fig. 26 shows the two modulation systems during the process of reconstruction with arrangements for making the necessary step-by-step comparisons between them.

Much attention was paid during the year to the frequency characteristic of the transmitter, which was made substantially flat from thirty to 20,000 cycles. This required careful attention to the characteristics of the doubler and amplifier circuits of the transmitter, and to John Evans of the RCA Manufacturing Company and to T. J. Buzalski I am indebted for its accomplishment. Continuous improvement of the transmitter and antenna efficiency was effected throughout the year, but of this phase of the development R. M. Morris of the National Broadcasting Company, under whose direction the work was carried on, is better qualified to speak. As the final step, the lines connecting the transmitter with the control board of the National Broadcasting Company at Radio City, from which the test programs were usually supplied, were equalized to about 13,000 cycles, and when this had been done the quality of reception at Haddonfield was far better than that obtainable from any of the regular broadcast stations.

Interference and Fading

Reference has heretofore been made to the difficulty of comparing the amounts of interference produced in amplitude and frequency modulation systems by the transient type of disturbance, particularly when, as in ignition noise, the peaks are greater in amplitude than the signal carrier. The best method of comparison seems to be that of observing how much greater signal level from the standard signal generator must be introduced into the receiving system when it is arranged to receive amplitude modulation than is required for the same signal-to-noise ratio on a frequency modulated system. The experimental procedure of making such comparison is to change the connection of the speaker rapidly from one receiver to the other, simultaneously changing the level of the local generator until the two disturbances as perceived by the ear are equal. At all times, of course, the amplification in the amplitude modulation receiver is correspondingly changed as the signal generator level is varied to apply the same voltage to the amplitude as to the frequency modulation detector so that the audio-frequency signal level which will be produced by the two systems is the same. The square of the ratio of the two voltages of the signal generator gives the factor by which the *carrier* power of the amplitude modulated transmitter must be increased to give equal performance. While the measurement is difficult to make, the following approximations may give some idea of the magnitudes involved.

If the peak voltage of the ignition noise is twice the carrier level of the frequency modulation system, about 150 to 200 times the power must be used in the carrier of the amplitude modulation system to reduce the disturbance level to the same value. When the peak voltage is five times as great, about 35 to 40 times the power in the amplitude modulation carrier is sufficient to produce equality.[17] These observations have been checked aurally and by the oscilloscope. The results of measurements where the disturbances are due solely to the thermal and shot effects have been compared to those obtained with the method previously described and are found to check with it. The chief value of this method of measurement, however, lies in the ability to predict with certainty the signal level required to suppress all ignition noise. An experimental determination made at Haddonfield shows that a signal introduced from the local generator which produces at the current limiter ten times the voltage of the Empire State signal is sufficient to suppress the disturbance caused by the worst offender among the various cars tested. These cars were located as closely as possible to the doublet antenna shown in Fig. 25, the distance being about forty feet. The increase in field strength necessary to produce this result can be readily obtained by an increase in the transmitter power to twenty or twenty-five kilowatts and the use of a horizontally directional antenna array. An increase in the field strength of three or four to one by means of an array is within the bounds of engineering design so that the practical solution of the problem of this type of interference is certainly at hand up to distances of one hundred miles.

So also is the solution of the problem at its source. It has been determined experimentally that the introduction of 10,000 ohms (a value of resistance which is not injurious to motor performance) into the spark plug and distributor leads of the car referred to eliminates the interference with the Empire State signal.

Since active steps are now being taken by the manufacturers of motor cars to solve the more difficult general problem, the particular one of interference with sets located in the home will thus automatically disappear. The problem of eliminating the disturbance caused by an automobile ignition system in a receiving set whose antenna is a minimum of fifty feet away from the car is obviously a much simpler one than that of eliminating the interference in a receiver located in the car or in another car a few feet away.

During the course of the experimental work in the laboratory a very striking phenomenon was observed in the interference characteristics between frequency modulation systems operating within the same wave band. The immunity of a frequency modulation system from interference created by another frequency modulated transmission is of the

[17] Linear detection was used in the amplitude modulation receiver but no limiting was employed.

same order of magnitude as the immunity with regard to tube noises. This property merits the most careful study in the setting up of a broadcast system at those wave lengths at which the question of interstation interference is a major factor. It is well known that when the carriers of two amplitude modulated transmitters are sufficiently close in frequency to produce an audible beat that the service range of each of them is limited to that distance at which the field strength of the distant station becomes approximately equal to one per cent of the field strength of the local station. As a consequence of this, the service area of each station is very greatly restricted; in fact the service area of the two combined is but a small percentage of the area which is rendered useless for that frequency due to the presence thereon of the two interfering stations. With the wide band frequency modulation system, however, interference between two transmissions does not appear until the field strength of the interfering station rises to a level in the vicinity of fifty per cent of the field strength of the local one. The reason for this lies in the fact, that while the interfering signal in beating with the current of the local station under such conditions may be producing a fifty per cent change in the voltage applied to the current limiter, the system is substantially immune to such variations in amplitude. The only way in which the interfering signal can make its presence manifest is by cross modulation of the frequency of the local signal. Since, under the conditions, this cross modulation produces less than a thirty-degree phase shift and since the characteristics of the wide band receiver are such that, at least within the range of good audibility, thousands of degrees of phase shift are necessary to produce full modulation, it is clear that a thirty-degree phase shift will not produce very much of a rectified output. For example, assuming two unmodulated carriers are being received, that their amplitudes have a ratio of two to one, and that their frequencies differ by 1000 cycles, then for a system having a wide band (of the order of 150,000 cycles) the modulation produced by the interaction of the two carriers would be of the order of one per cent of that produced by full modulation of the stronger carrier. This example, however, represents perhaps the worst possible condition as during modulation of either station, with the proper type of conversion system, the aural effect of the disturbance is greatly reduced. The whole problem of interference between unmodulated carriers may, however, be entirely avoided by separating them in frequency by an amount beyond the audible range. Hence it follows that with two wide band frequency modulated transmitters occupying the same frequency band that only the small area located midway between the two wherein the field strength of one station is less than twice the field strength of the other will be rendered useless for reception of either station. This area may well be less than ten per cent of the total area. Even in this area reception may be effected as a receiving station located within it has only to erect directional aerials having a directivity of two to one to receive either station. The two-to-one ratio of field strength which has been referred to as the ratio at which interference appears is not by any means the limit but rather one which can be realized under practically all conditions. Better ratios than this have been observed, but the matter is not of any great importance since by the use of the directional antennas referred to it becomes possible to cover the sum of the areas which may be effectively covered by each station operating alone, subject only to the limitations of the noise level. The problem of the interference due to overlapping has been completely wiped out. One precaution only should be observed—the unmodulated carriers should be offset in frequency by an amount beyond the audible limit.

In the above analysis it has been assumed, of course, that the distance between stations has been selected so that the "no-mans land" between stations is not sufficiently distant from either one to be within the zone where any large amount of fading occurs. If the distance between stations is such that the signal strength varies appreciably with time then the directivity of the receiving antennas must be greater than two to one.

Difficulties and Precautions

The principles which have been described herein were successfully applied only after a long period of laboratory investigation in which a series of parasitic effects that prevented the operation of the system were isolated and suppressed. The more important of these effects, which will be of interest to those who may undertake work in this field, will be referred to briefly.

It was observed in the early work in the laboratory that it was at times impossible to secure a balance in the detector system, and that the amplitudes of the currents in the rectifiers varied in very erratic fashion as the frequency of the first heterodyne was changed. Under these conditions it was not possible to produce any appreciable noise suppression. The effect varied from day to day and the cause defied detection for a long period of time. Ultimately the presence of two side frequencies in the detector circuits was discovered, one of these frequencies lying above and the other below the unmodulated intermediate frequency by an amount equal to the initial crystal frequency of the transmitter. It was then discovered that the trouble had its origin in the transmitting system and that a current having the fundamental frequency of the crystal, (in the present case 57.33 kilocycles), passed through the first doubler circuits in such phase relation to the doubled frequency as to modulate the doubled frequency at a rate corresponding to 57.33 kilocycles per second. This modulation of frequency then passed through all the transmitter doubler stages, increasing in extent with each frequency multiplication and appearing finally in the forty-four-megacycle output as a fifty-seven-kilocycle frequency modulation of considerable magnitude. In the first doubler tank circuit of the transmitter a very slight change in the adjustment of the tuning of the circuit produced a very great change in the magnitude of this effect. A few degrees shift in the tuning of the first doubler tank condenser, so small that an almost unnoticeable change in the plate current of the doubler occurred, would increase the degree of the modulation to such extent as to make the first upper and lower side frequencies in the forty-four-megacycle current greater than the carrier or mid-frequency current (when no audio modulation was applied). Under such conditions the proper functioning of the receiving system was impossible.

The delay in uncovering this trouble lay in the fact that it was obscured by the direct effect of harmonics from the transmitter doubler stages which had to be set up in an adjoining room and by the numerous beats which can occur in a double intermediate-frequency superheterodyne. To these effects were added an additional complication caused by the presence of harmonics in the circuits of the selective system resulting from the action of the limiter which the filtering arrangements did not entirely remove. The coincidence of one of these harmonics with the natural period of one of the inductances in the branch circuits likewise interfered with the effectiveness of the noise suppression. The causes of all these spurious effects were finally located and necessary steps taken to eliminate them.

With the removal of these troubles a new one of a different kind came to light, and for a time it appeared that there might be a very serious fundamental limitation in the phase shifting method of generating frequency modulation currents. There was found to be in the output of the transmitter at forty-four megacycles a frequency modulation which produced a noise in the receiver similar to the usual tube hiss. The origin of it was traced to the input of the first doubler or the output of the crystal oscillator where a small deviation of the initial frequency was produced by disturbances originating in these circuits. While the frequency shift in this stage must have been very small, yet on account of the great amount of frequency multiplication (of the order of 800 times) it became extremely annoying in the receiver; in fact for low levels of receiver noise that noise which originated in the transmitted wave was by far the worse. For a time it seemed as though the amount of frequency multiplication which could be used in the transmitter was limited by an inherent modulation of the frequency of the oscillator by disturbances arising in the tube itself. The proper proportioning of the constants of the circuits, however, reduced this type of disturbance to a point where it was no longer of importance and frequency multiplications as high as 10,000 have since been effectively used. On account of the very large amount of frequency multiplication, any troubles in these low-frequency circuits caused by noisy grid leaks, improper bypassing of power supply circuits, or reaction of one circuit upon another become very much more important than they would normally be. Difficulties of all these kinds were encountered, segregated, and eliminated.

Another source of trouble was discovered in the correction system. Because of the range in frequency required, particularly in multiplex work where thirty to 30,000 cycles was frequently used, the output voltage of the correction system at the higher frequencies became very much less than the input voltage, hence any leakage or feed-forward effect due to coupling through the power supply circuits developed a voltage across the output much higher than that required by the inverse frequency amplification factor as determined by the correction network. Hence, the frequency swing for the upper frequencies of modulation would frequently be several hundred per cent greater than it should be. Likewise, at the lower frequency end of the scale various reactions through the power supply were very troublesome. All these effects, however, were overcome and the correction system designed so that its accuracy was within a few per cent of the proper value.

From the foregoing it might be assumed that the transmitting and receiving apparatus of this system are inherently subject to so many new troubles and complications that their operation becomes impracticable for ordinary commercial applications. Such is not the case. The difficulties are simply those of design, not of operation. Once the proper precautions are taken in the original design these difficulties never occur, except as occasioned by mechanical or electrical failure of material. During the period of over a year in which the Empire State transmitter was operated, only two failures chargeable to the modulating system occurred. Both were caused by broken connections. Even the design problems are not serious as methods are now available for detecting the presence of any one of the troubles which have been here enumerated.

These troubles were serious only when unsegregated and en masse they masked the true effects and made one wonder whether even the laws of electrical phenomena had not been temporarily suspended.

Multiplex Operation

During the past year, two systems of multiplexing have been operated successfully between New York and Haddenfield and it has been

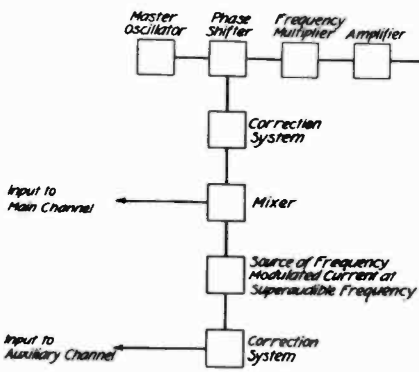


Fig. 27

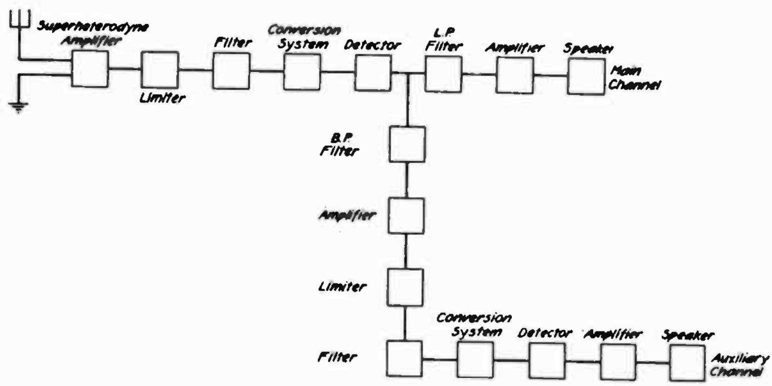


Fig. 28

found possible to transmit simultaneously the red and blue network programs of the National Broadcasting Company, or to transmit simultaneously on the two channels the same program. This last is much the simpler thing to accomplish as the cross-talk problem is not a serious one. The importance of multiplexing in point-to-point communication services has long been recognized. In broadcasting there are several applications which, while their practical application may be long deferred, are clearly within view.

Two general types of multiplexing were used. In one type a current of superaudible frequency is caused to modulate the frequency of the transmitted wave. The frequency at which the transmitted wave is caused to deviate is the frequency of this current and the extent of the deviation is varied in accordance with modulation of the amplitude of the superaudible frequency current. At the receiver detection is accom-

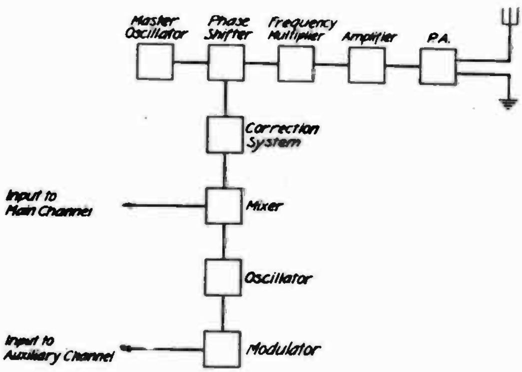


Fig. 29

plished by separating the superaudible current and its component modulations from the rectified audible frequency currents of the main channel and reproducing the original modulating current from them by a second rectification. The general outline of the system is illustrated in Figs. 27 and 28. The setting of the levels of the main and auxiliary channels must be made in this system of modulation with due regard

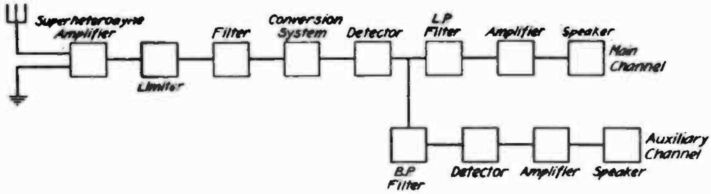


Fig. 30

to the fact that the deviation of the transmitted wave produced by the superaudible frequency current of the second channel is a variable one and changes between the limits of zero and double the unmodulated deviation.

In the second method of multiplexing a superaudible current produces a frequency modulation of the transmitted wave of constant deviation, the rate of the deviation being varied in accordance with the frequency of the superaudible current and modulation being produced by varying the frequency of this auxiliary current and thereby the rate at which the superimposed modulation of frequency of the transmitted wave changes. The operations which must be carried out at the receiver are the following: After suitable amplification, limiting, and filtering, an initial conversion and rectification produces in the output of the detector the audible frequencies of the main channel and a superaudible constant amplitude variable frequency current. This last is selected by means of a band-pass filter, passed through a second con-

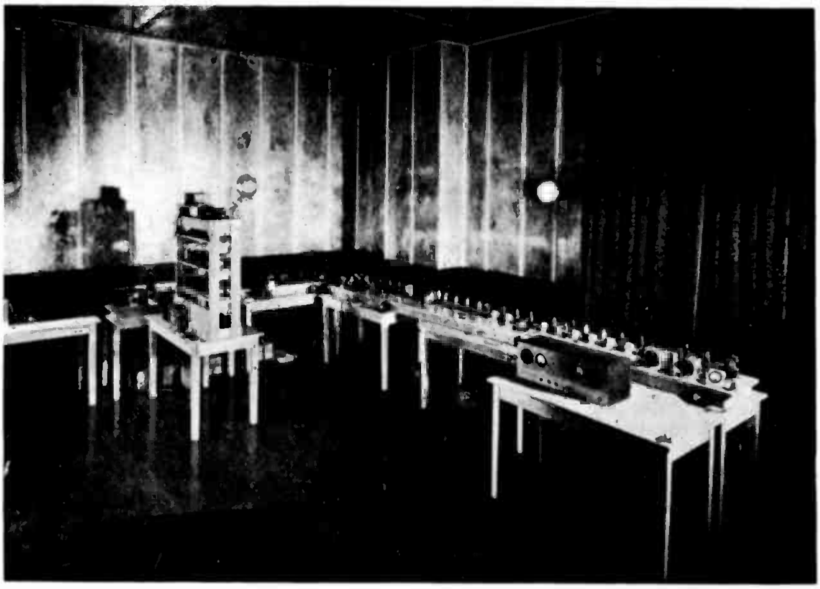


Fig. 31

version system to translate the changes in the frequency into variations of amplitude, and then rectified to recreate the initial modulating current of the auxiliary channel. The general arrangment of the system is illustrated in Figs. 29 and 30. This latter method of multiplexing has obvious advantages in the reduction of cross modulation between the channels and in the fact that the deviation of the transmitted wave produced by the second channel is constant in extent, an advantage being gained thereby which is somewhat akin to that obtained by frequency, as compared to amplitude, modulation in simplex operation. The subject of the behavior of these systems with respect to interference of various sorts is quite involved and wili be reserved for future treatment as it is beyond the scope of the present paper.

The final arrangement of the modulating equipment installed at the Empire State station is illustrated in Figs. 31 and 32. The main channel apparatus is shown on the five tables located on the right side of the room. The vertical rack in the left center contains three channels for transmitting facsimile by means of the amplitude modulation method of multiplexing already described. In Fig. 32, located on the four tables on the left of the room is shown the auxiliary channel of the frequency modulation type already described. The comparatively

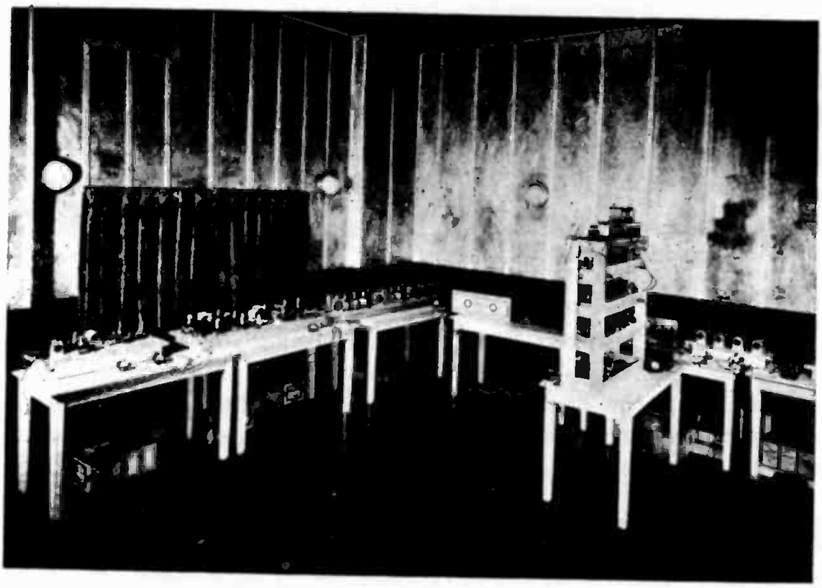


Fig. 32

low frequency of this channel was obtained by the regular method of phase shifting and frequency multiplication, the frequency multiplication being carried to a high order and the resultant frequency modulated current heterodyned down to twenty-five kilocycles (mid-frequency). A deviation up to ten kilocycles was obtainable at this frequency.

The receiving apparatus located at Haddonfield is illustrated in Figs. 33 and 34. Fig. 33 shows the modified Westhampton receiver and Fig. 34 the multiplex channels of the receiver. The vertical rack to the right holds a three-channel receiver of the amplitude modulation type. The two panels in the foreground constitute the frequency modulation type of auxiliary channel.

Some of the practical results may be of interest. It was suggested by C. J. Young of the RCA Manufacturing Company that it might be possible to transmit simultaneously a facsimile service at the same time that a high quality broadcast program was being transmitted. With the assistance of Mr. Young and Maurice Artzt this was accomplished

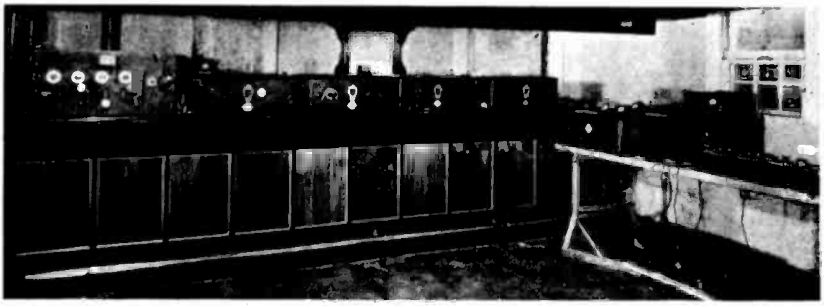


Fig. 33

over a year ago between New York and Haddonfield, New Jersey, the two services operating without interference or appreciable loss of efficiency at the distance involved. Two additional channels, a synchronizing channel for the facsimile and a telegraph channel, were also operated. The character of the transmission is illustrated in Fig.

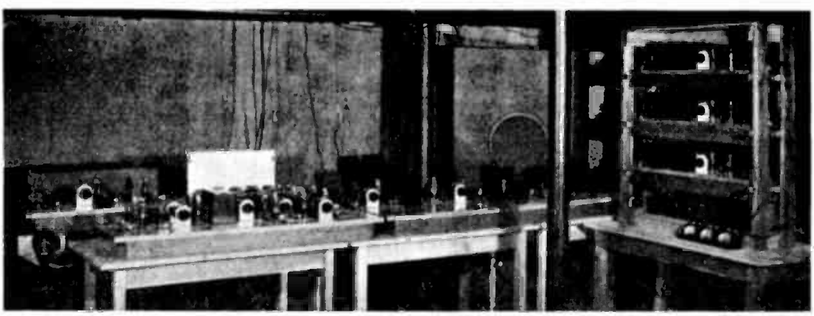


Fig. 34

35, which shows a section of the front page of the *New York Times*, This particular sheet was transmitted under considerable handicap at the transmitter as due to a failure of the antenna insulator on the forty-one-megacycle antenna it had become necessary to make use of the sixty-megacycle antenna for the forty-one-megacycle transmission. It is an interesting comment on the stability of the circuits that all four were kept in operation at the transmitter by one man, Mr. Buzalski,

New York Times.

NEW YORK, SATURDAY, NOVEMBER 24, 1934. TWO CENTS

VANDERBILT WRIT CALLS ON CAREW TO DEFEND RULING

Definite Decision Is Sought on Habeas Corpus Action to Pave Way for Appeal.

GUARDIAN FIGHT RESUMED

Reply to Mother's Plea Asserts She Is Not Qualified to Fill Such a Position.

SOVIET ARMY TO AID FRANCE IF INVADED, DEPUTIES ARE TOLD

Reporter of Budget Declares Russia Would Help Repel Any German Attack.

MILITARY FUNDS APPROVED

French War Minister Announces Term of Conscript Service Will Not Be Raised.

U. S. AND BRITA[...] AS NAVY TALKS [...] JAPAN TO DEN[...]

PACT ASSAILED BY SAITO

Ambassador Says Tokyo Will End It No Matter What Occurs at London.

HOLDS HONOR IS AT STAKE

Tells Philadelphia Audience Ratios Show Japan Is Looked Down On as 'Spoiled Child.'

OPPOSES BUILDING RACE

Declares Tokyo Is Willing to Enter an Agreement to Guard Philippine Independence.

HULL ENC[...] UNITY WIT[...]

Nov. 24, 1934

Four Channel Operation
(1) Music from chain
(2) Synchronizing channel
(3) Facsimile signal
(4) Telegraph channel
Empire State & Haddonfield N. J. W2XDG on 41 megacycles operating on 61 megacycle antenna.

Fig. 35

who was alone in the station on that day. The combined sound and facsimile transmission has been in successful operation for about a year, practically perfect copy being obtained throughout the period of the

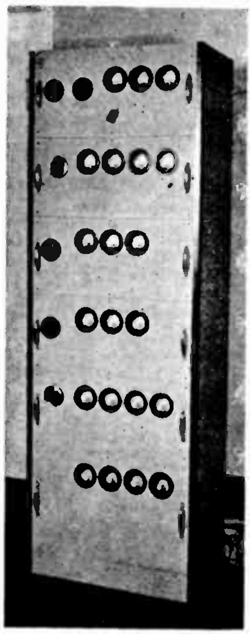


Fig. 36

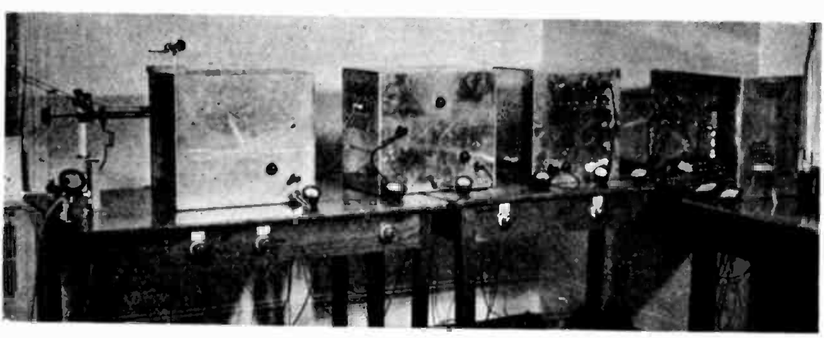


Fig. 37

severe atmospheric disturbances of the past Summer. The subject of this work and its possibilities can best be handled by Mr. Young, who is most familiar with it.

Acknowledgment

On account of the ramifications into which this development entered with the commencement of the field tests many men assisted in this work. To some reference has already been made.

I want to make further acknowledgment and express my indebtedness as follows:

To the staff of the National Broadcasting Company's station W2XDG for their help in the long series of field tests and the conducting of a large number of demonstrations, many of great complexity, without the occurrence of a single failure;

To Mr. Harry Sadenwater of the RCA Manufacturing Company for the facilities which made possible the Haddonfield tests and for his help with the signal-to-noise ratio measurements herein recorded;

To Mr. Wendell Carlson for the design of many of the transformers used in the modulating equipment;

To Mr. M. C. Batsel and Mr. O. B. Gunby of the RCA Manufacturing Company for the sound film records showing the comparison, at Haddonfield, of the Empire State transmission with that of the regular broadcast service furnished by the New York stations;

To Mr. C. R. Runyon for his development of the two-and-one-half-meter transmitters and for the solution of the many difficult problems involved in the application of these principles of modulation thereto;

To Mr. T. J. Styles and particularly to Mr. J. F. Shaughnessy, my assistants, whose help during the many years devoted to this research has been invaluable.

Conclusion

The conclusion is inescapable that it is technically possible to furnish a broadcast service over the primary areas of the stations of the present-day broadcast system which is very greatly superior to that now rendered by these stations. This superiority will increase as methods of dealing with ignition noise, either at its source or at the receiver, are improved.

Appendix

Since the work which has been reported in this paper on forty-one megacycles was completed attention has been paid to higher frequencies. On the occasion of the delivery of the paper a demonstration of transmission on 110 megacycles from Yonkers to the Engineering Societies Building in New York City was given by C. R. Runyon, who described over the circuit the transmitting apparatus which was used. A brief description of this transmitter is reproduced here.

The power delivered to the antenna was approximately 100 watts at 110 megacycles and the deviation (one half total swing) used during the demonstration was under 100 kilocycles. Fig. 36 illustrates the modulating equipment for this transmitter and the low power frequency multiplication stages. Fig. 37 shows the higher power frequency multiplier and power amplifier stages of the transmitter.

The rack shown in Fig. 36 consists of six panels. Panel number one at the top contains the correction system. Panel number two contains the master oscillator of 100 kilocycles and the modulator circuits. Panel number three contains a pair of doublers for multiplying the 100-kilocycle frequency to 400 kilocycles and the necessary filtering means for avoiding the modulation of the currents in the succeeding doubler stages by the 100-kilocycle oscillator current. Panel number four contains the doubling apparatus for raising the frequency to 3200 kilocycle and panel number five the multipliers for raising it to 12,800 kilocycles. Panel number five also contains a heterodyning and conversion system for beating the 12,800 kilocycles down to 2292 kilocycles. Panel number six contains a doubler for raising this to 4584 kilocycles and an amplifier for increasing the level sufficiently to drive the succeeding power stage. The output of this amplifier is fed through a transmission line to the metal box at the extreme right of Fig. 36 which contains a series of doublers and amplifiers for increasing the level and raising the frequency to 36,672 kilocycles. Adjacent to this box is a second box which contains a fifty-watt amplifier. This amplifier drives a tripler located in the third box and the tripler in turn drives the power amplifier located at the extreme left at 110 megacycles. The transmitter circuits were designed for total frequency swing of 500 kilocycles and may be effectively so operated. Because of the limitation of the receiver available at that time the demonstration was carried out with a swing of 200 kilocycles.