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 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.

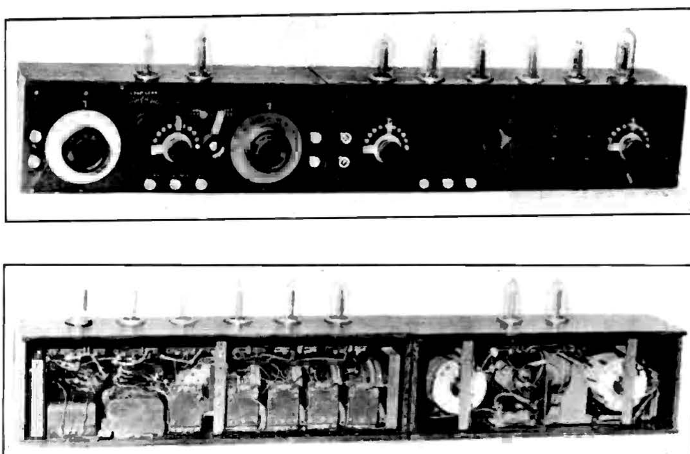

FIGURE 1

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-

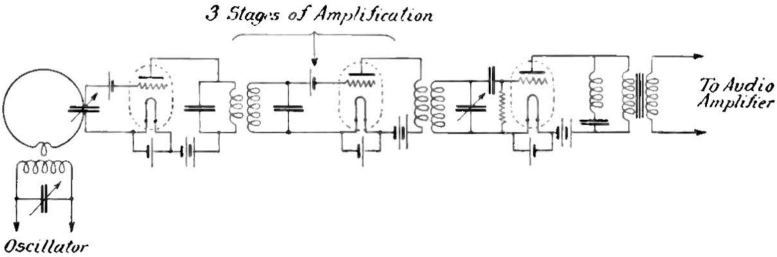

FIGURE 2

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.

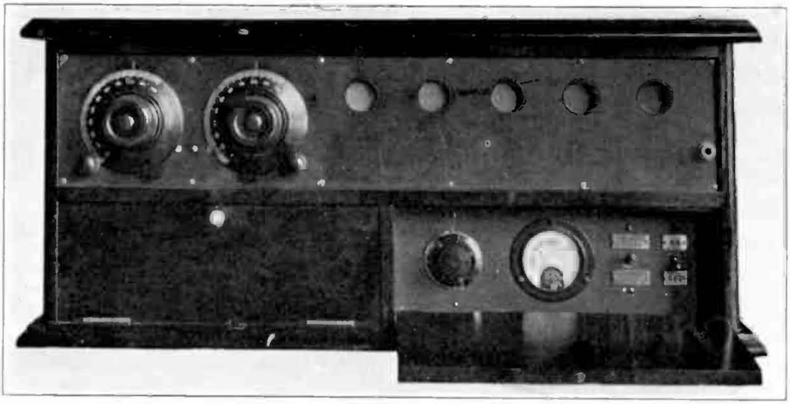

FIGURE 3

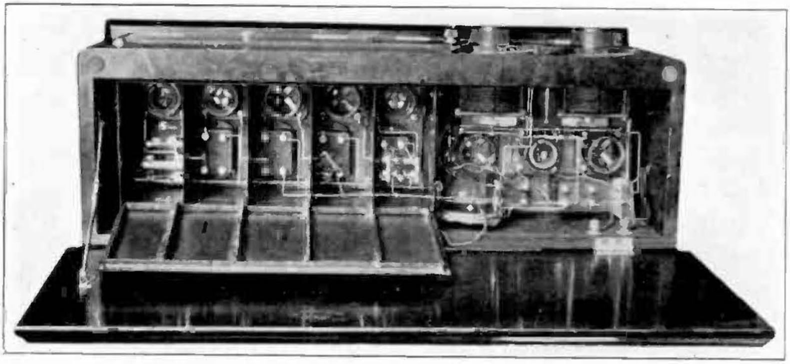

FIGURE 4

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.

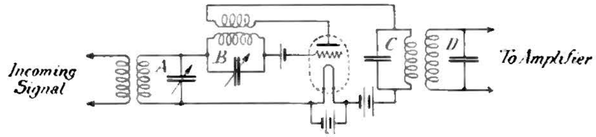

FIGURE 5

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

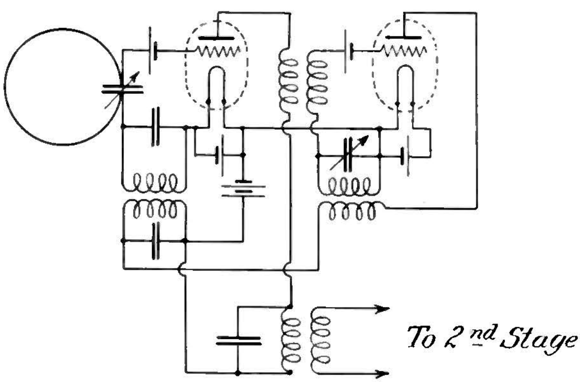

FIGURE 6

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.

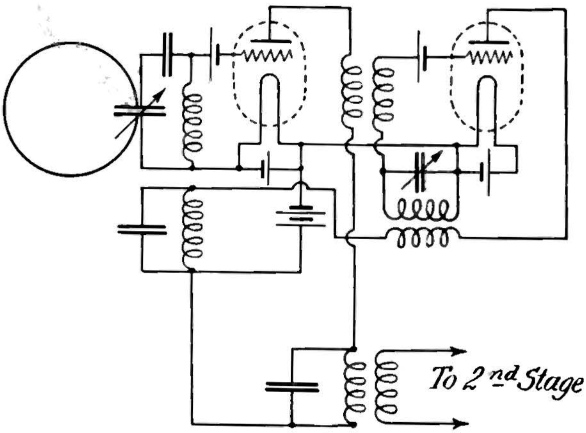

FIGURE 7

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.

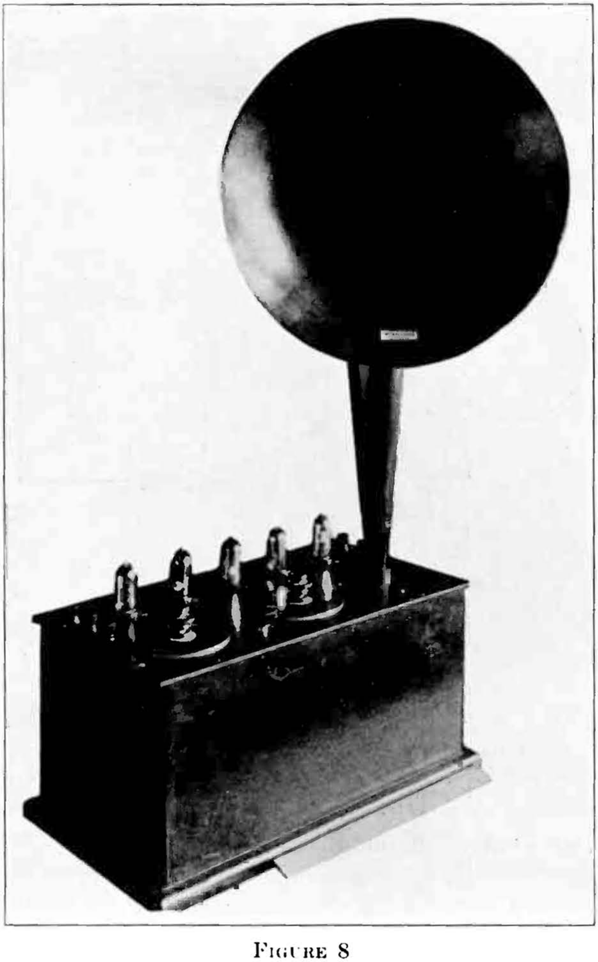

FIGURE 8

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-

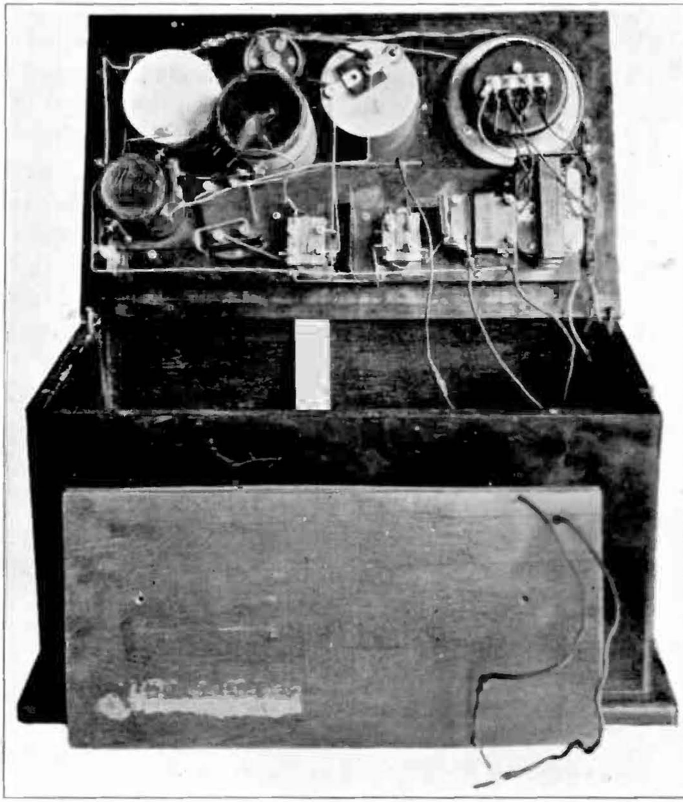

FIGURE 9

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.

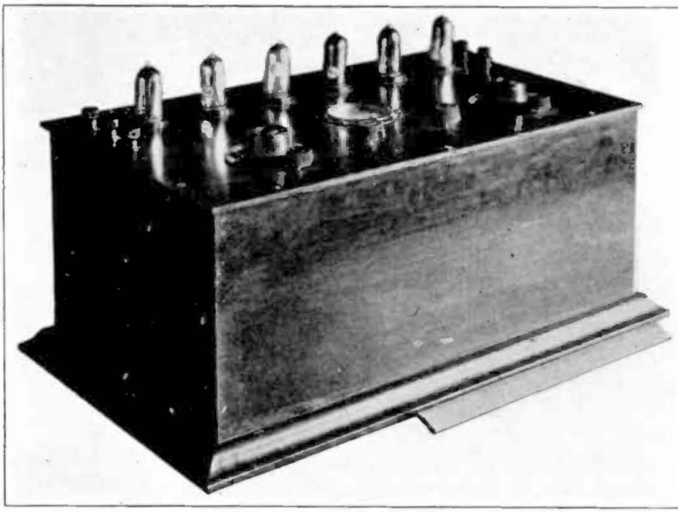

FIGURE 10

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.

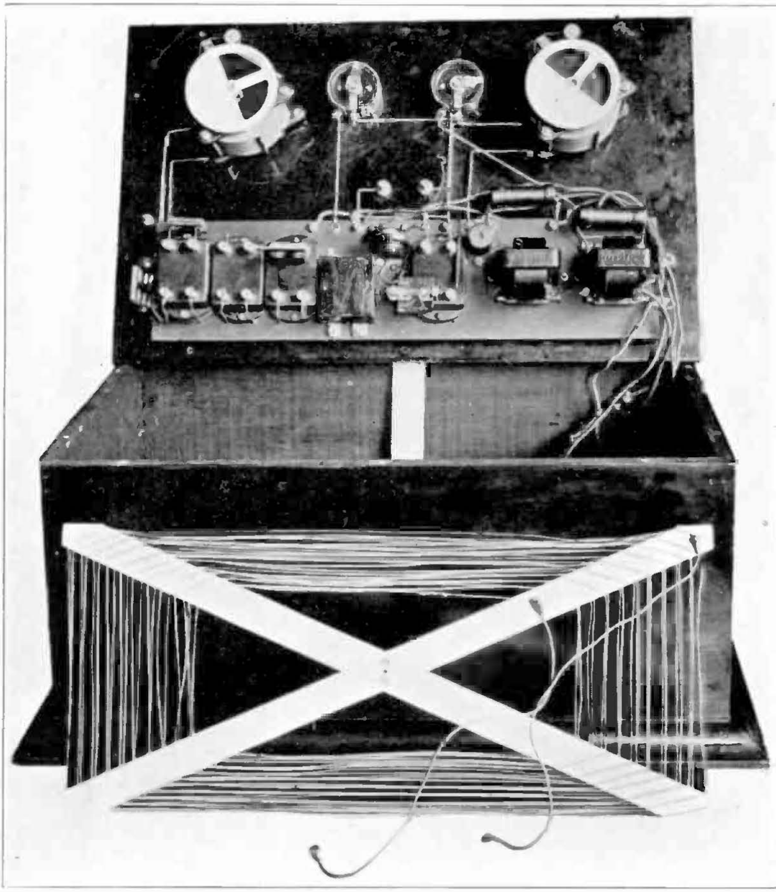

FIGURE 11

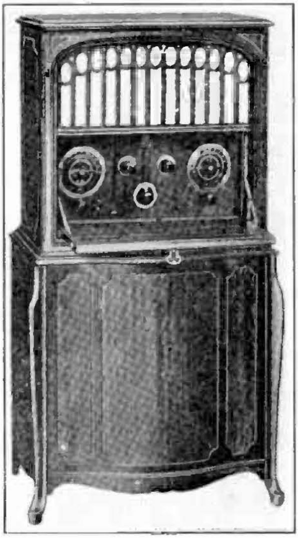

FIGURE 12

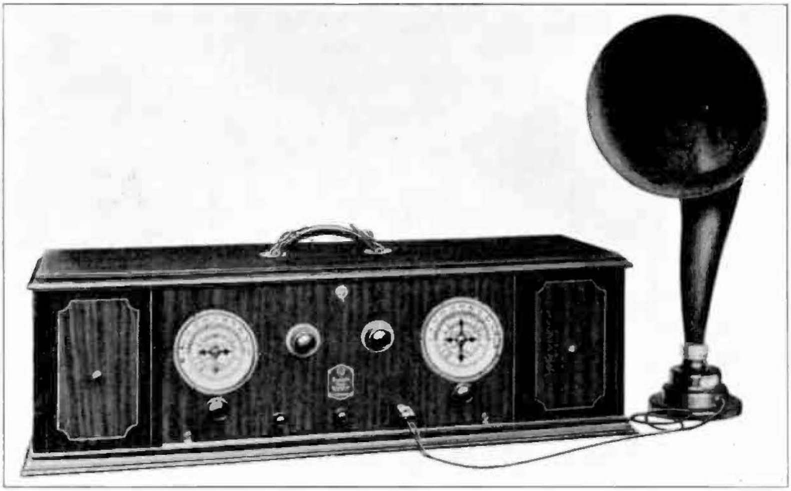

FIGURE 13