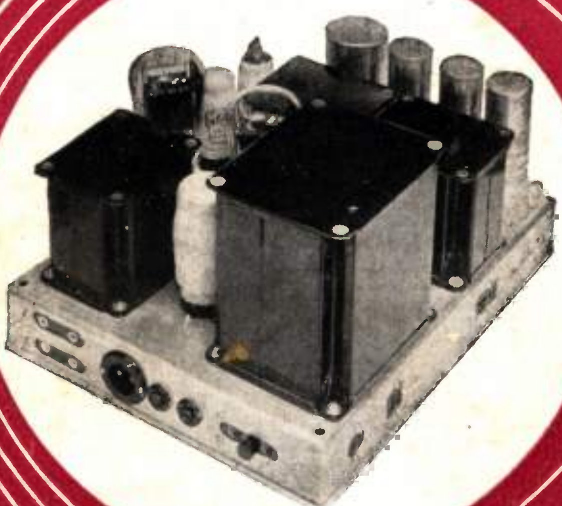
AN INTRODUCTION TO

FREQUENCY MODULATION



K. E. V. WILLIS A.R.C.S. B.Sc.

Incorporating: Modulation • The Carrier Wave • Amplitude Moderators • Superheterodyne Receiver • Frequency Modulation Receivers, etc.

AN INTRODUCTION TO FREQUENCY MODULATION

AN INTRODUCTION

TO

FREQUENCY MODULATION

by

K. E. V. WILLIS, A.R.C.S., B.Sc.

MODERN UNIVERSITY PRESS LTD.

1A, CADOGAN PLACE, KNIGHTSBRIDGE, S.W.1

FOREWORD

Until recent years, the amplitude modulation system has held an undisputed sway in the field of radio transmission. Consequently, this system is well understood, and, having withstood the test of time, it forms a very suitable "yardstick" against which the performance of other systems may be measured. For this reason readers will find that almost as much space has been devoted in this book to the amplitude modulation system as to the subject of frequency modulation.

No apology is felt to be necessary for the adoption of this arrangement, for it has been the intention of the author to give a comparative treatment in order that readers with but a general interest in the subject may gain a knowledge of the principles involved without recourse to other volumes.

Detailed descriptions of circuits have purposely been avoided, and serious students of the subject would be advised to supplement their reading of this book by a study of some of the more advanced works listed in the bibliography.

K.W.

CHAPTER ONE

Modulation

It is an established fact that intelligence cannot be conveyed between two separate points unless more than one frequency is employed. For example, the act of speaking calls into play a large number of frequencies in the audible range. A sustained musical note of single frequency, however, produces the sensation of sound by its action at the ear of the listener, and yet conveys no " message " in the process. To extend this analogy still further, it will be obvious that the message or intelligence would in this case be a melody or a chord, and both of these involve not one but a number of frequency components. A similar condition is that which arises when two points are linked by radio. The mere fact that a radio wave radiated by the aerial of a transmitting station can be detected at a remote point by suitable apparatus, does not imply that intelligence* can be exchanged between the two points, for to do so requires the use of more than the single frequency of the radio wave. The method by which the required component frequencies are incorporated in the radio wave to make possible the transmission of speech, music and other similar effects is termed " modula-

* The application of television requires that the term " intelligence " shall include not only speech and music (i.e., audio frequencies), but also the complex signals which give rise to the television image (video frequencies).

tion," and there are a number of modulation methods from which a choice may be made. Frequency modulation is but one system, so that it will be necessary to consider certain other methods in some detail before its action and merits can be appreciated to the full.

THE CARRIER WAVE

For the transmission of any form of intelligence by radio means, it is customary to employ a high-frequency (i.e., a radio-frequency) wave to fulfil the role of a carrier for the modulation which consists of audio or video frequencies. One of the reasons which underly this process is that radio-frequency waves may be propagated over very long distances without very serious loss of energy, whereas an audio-frequency wave propagated on its own would very quickly be attenuated to such an extent that it would become inaudible. It will be useful at this juncture to revise the chief characteristics of the electro-magnetic wave, for the carrier wave for all forms of modulation in the field of radio-communication is of this type. The carrier wave may therefore be represented very simply by the equation :

$$y = A \sin [2\pi f_c t]$$

This relates the instantaneous amplitude y of the wave at any time t to the maximum possible amplitude A. The wave performs f_c complete vibrations in unit time, usually taken to be the second, this being the frequency of the wave. When the above equation is plotted for increasing values of t, assuming f_c to remain constant, the well-known sine curve of figure 1(a) is obtained.

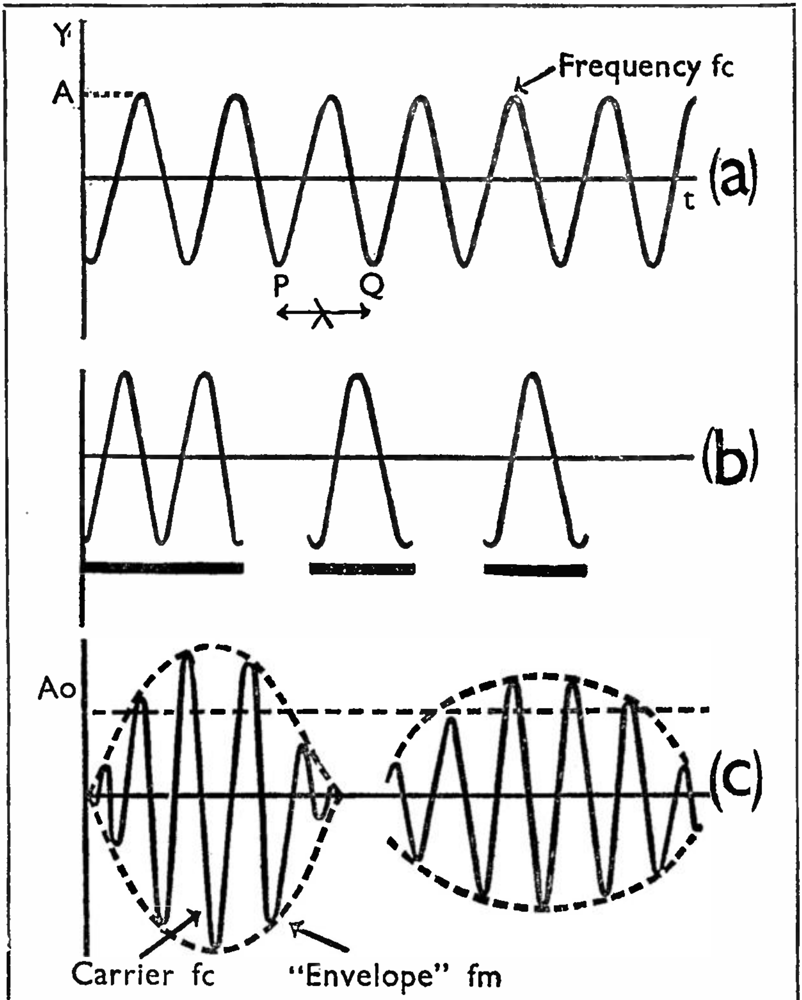


FIGURE 1

This is also the graphical representation of an electro-magnetic wave of constant maximum amplitude A, and of which the section PQ is one complete cycle. In unit time the wave performs f_c such cycles, and the distance through which the wave progresses in space during one cycle is the wavelength, λ, equal to PQ. Hence the velocity of propagation of the wave is given by:

$$\lambda f_c = v = 186{,}000 \text{ miles/second, in air.}$$

All electro-magnetic waves travel at the same velocity in any given medium, regardless of their frequency. The point reached along the actual graphical construction of the wave at any instant is termed the "phase" at that particular time. Phase really indicates the part of the complete cycle which has been traced out at any instant. Thus two separate waves initiated from a single source will be in identical parts of their series of cycles at every instant if they are "in phase." It is usual to use angular measure for the units of phase (e.g., radians or degrees), and these units are quite appropriate if the wave is considered as being derived from the projection on a moving plane of a vector which rotates with constant angular velocity given by $2\pi f_c$. This factor is called the "pulsatance" of the wave. The outcome of this reasoning is that a wave is considered to have suffered a phase change of 2π radians or 360° after having performed one complete cycle, after which it commences another cycle, identical with the first. Hence any part of the graphical sine curve, and therefore any state of the wave's variation, may be represented by the portion of the total 2π radians which has been traced out when the particular point is reached.

Figure 1(a) is representative of an unmodulated carrier wave of frequency f_c, and it is to this frequency that the receiver must be tuned to receive the signal. Reception of the carrier does not, of course, result in the exchange of intelligence, since no modulation has yet been impressed upon the carrier. The simplest form which the modulation may take is that indicated in figure 1(b), which shows a method of transmitting prearranged code signals which may be interpreted at the receiving station. The figure illustrates the way in which the carrier wave may be keyed in a definite sequence which, in the particular case shown, gives rise to the reception of the letter D of the International Morse Code. At first sight it might appear that the original stipulation demanding the use of more than one frequency for the conveyance of any form of intelligence has, in this case, been violated, but on further thought it will be realised that this is not so. The reason is, that a Fourier Analysis of a wave which is suddenly started and stopped in this way will disclose the existence of an infinite number of component frequencies which will appear in the keyed waveform. So simple a method of modulation is not applicable to the transmission of speech or music, and consequently it has been necessary to devise more complicated systems, the main features of which will be described.

Amplitude Modulation

According to the dictionary, the expression " to modulate " means to regulate or vary in tone. These definitions suggest that any form of modulation will consist essentially of varying some factor associated with the carrier wave in order to impress upon it the modulation frequencies to be transmitted. Generally speaking, there are three characteristics of the carrier wave which may be varied in order to modulate. These are amplitude, frequency and phase. Of these, only amplitude variations have been employed until recent years, and so important a part does this system play in the art of radio communication, that it is essential for a full understanding of other forms of modulation that its features should be described. When using amplitude modulation, the carrier wave is maintained at a constant frequency, but the amplitude of the wave is varied at a rate which is directly proportional to the modulation frequency it is desired to transmit. Figure 1(c) illustrates the effect of modulating a carrier wave f_c by an audio frequency f_m, the variations in amplitude being observed to take place about a mean value A_o. On the left-hand side of this figure a waveform arising from the modulation of the carrier by a source of large amplitude is shown, whilst to the right is an illustration of modulation to a lesser extent. An inspection of the diagrams will reveal that the " envelope " containing the high-frequency waves is the waveform of the modulation frequency. This is one of the simplest cases of amplitude modulation, for the modulation frequency has been assumed to remain constant throughout, a condition which will not prevail when, for example, speech or music is transmitted. It is the function of the apparatus at the receiving station to extract from the modulated carrier the radio-frequency and audio-frequency components, reproducing the latter in the loudspeaker or headphones.

Side-band Radiation

The amplitude modulation of a carrier wave is not quite as simple a process as it may at first sight appear. It is found that when a carrier of frequency f_c is modulated by any frequency f_m, then frequencies other than

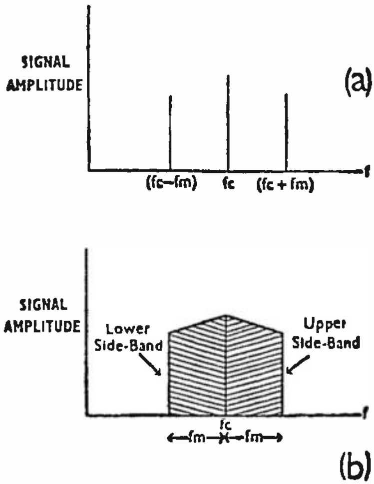


FIGURE 2

f_c and f_m are present in the resulting waveform. In the simple case previously considered when f_m remained constant throughout, the *radio-frequencies* produced on modulation are $(f_c + f_m)$ and $(f_c - f_m)$ in addition to f_c itself. These are shown in figure 2(a), and it is interesting to note that if f_m is an audio-frequency, then the two component frequencies formed during the modulation will be radio-frequencies, and as such will be radiated along with the carrier itself. In fact each may be tuned-in separately on the receiver. The situation becomes much more complicated when the modulation consists not of a single frequency but of a series of frequencies such as would be encountered during the transmission of speech or music. Each individual value of f_m present in the modulation wave will give rise to its pair of frequencies situated on either side of the carrier frequency, so that instead of a single pair of side frequencies, bands are now radiated both above and below the carrier frequency. These are termed the " side-bands," and their width is determined by the highest modulation frequency employed. Reference to figure 2(b), which shows the disposition of the side-bands about the carrier frequency, indicates that the signal will now occupy a definite portion of the radio-frequency spectrum instead of being a single-frequency radiation, as was the case when the wave was unmodulated. One fact which should be borne in mind is that no matter how complex the distribution of frequency components within the side-bands, their width never exceeds that determined by the highest modulation frequency. In deciding upon the width of the channel to be made available to any given modulated wave in the radio frequency spectrum, it is customary to assume that the highest modulation frequency required will be 5,000 cycles per second (5 kc/s). Thus the overall width of the side-bands will be 10 kc/s whenever this modulation frequency is operative. This assumption emphasises the fact that two stations which operate on carrier frequencies separated by less than 10 kc/s will probably encounter interference between the respective side-bands unless the modulation frequencies are purposely kept very low. Obviously, then, only a limited number of channels are available for broadcast services, a fact which has become increasingly important with the rapid growth of transmitting stations during recent years. The trend, therefore, has been to seek less-crowded frequency bands such as the short and ultra-short wavelength regions, in which to allocate frequencies to the growing number of services, and in addition to devise modulation methods which either reduce the width of the side-bands radiated or reduce their interference to a minimum. It is in this connection that the field of frequency modulation holds out a great promise for future application. There are indications that frequency modulation will not only minimise interference, but will also give a greater fidelity of reproduction and an improvement in the received strength over and above the receiver noise level. Even at this comparatively early stage of its development, frequency modulation is proving to be a marked improvement over the amplitude modulation system both from the point of view of the results obtained and the relatively simple apparatus with which these results may be achieved.

CHAPTER TWO

Frequency Modulation

In the opening chapter, reference was made to the three characteristics of a radio wave which, for the purposes of modulation, may be considered as variables, and one such system—that of amplitude modulation—has been dealt with in some detail. It is possible, however, to maintain a constant value of amplitude and to vary the carrier frequency in conformation with the required modulation frequency. This frequency modulation occurs about the carrier frequency, which, in the absence of modulation, is constant modulation frequency f_m. In this case it is given in figure 3(a), which illustrates a carrier of amplitude A undergoing frequency modulation by a constant modulation frequently f_m. In this case it is

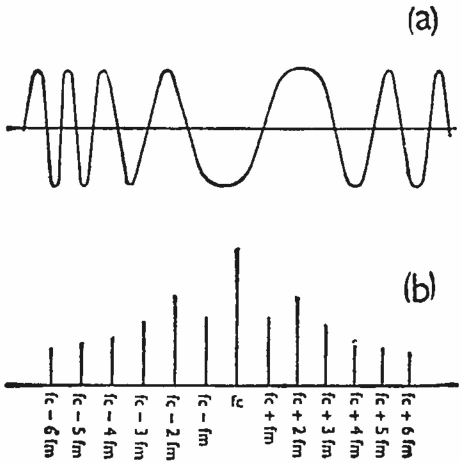


FIGURE 3

not easy to identify the " envelope " representing the modulation, but a change in the frequency of the wave (indicated by an increase or decrease in the number of complete cycles per unit time) about the mean or carrier frequency will be observed. There is not a necessarily rigid relation between the actual change in the frequency of the carrier and the modulation frequency. Instead, the requirement for frequency modulation is that for any modulation frequency f_m the carrier frequency shall change at the same rate, i.e., f_m times per second. The system is operated so that the amount of frequency change is governed by the amplitude of the modulation wave, and the term employed to describe the excursions of frequency both above and below the mean or unmodulated frequency is " the deviation." A numerical example will serve to illustrate this expression. If the modulation frequency is 2,000 cycles per second, the carrier frequency changes 2,000 times in every second. If the carrier frequency f_c is 10,000 kc/s, then a deviation of 100 kc/s indicates that the carrier will swing between the frequency limits of 9,900 kc/s and 10,100 kc/s. Assuming this to apply to the case when the modulation wave was of large amplitude, then for less depth of modulation the deviation produced would be considerably less, say of the order 100 cycles per second. Thus the carrier would undergo frequency changes of magnitude 100 kc/s when an intense modulation wave was impressed on the carrier, and of 100 cycles per second for a less intense wave. The deviation is therefore proportional to the intensity of the modulation wave, and is a factor which may be controlled by the design of the equipment. When deviation is large, say many times the highest modulation frequency, the system is designated " wide-band frequency modulation," and similarly " narrow-band " operation utilises only a small value of deviation whatever the amplitude of the modulation wave.

The idea of applying the frequency modulation system is by no means new. As early as 1920 its possibilities were under consideration, for it was felt that an appreciable reduction in the width of the side-bands would be apparent when a wave was frequency modulated. This hypothesis, attractive as it seemed, was shown by J. R. Carson in an analysis of the subject published in 1922 to be based on an incorrect assumption, and that no reduction in the " spread " of the transmitted signal could be expected. In fact, precisely the converse effect will be observed, because the side-frequencies become infinitely more complex when the carrier is frequency modulated than when the amplitude modulation is employed. For example, if the unmodulated carrier at f_c cycles per second is frequency modulated by a frequency f_m, pairs of side-frequencies, disposed regularly about f_c are radiated. The actual frequencies produced in this way have values $f_c \pm (f_m, 2f_m, 3f_m - - - nf_m)$ as indicated in figure 3(b). As before, the situation is further complicated when the modulation waveform consists of a number of frequencies, for then the side-frequencies become so numerous that extensive side-bands are radiated which have an appreciable amplitude even when they exist beyond the maximum deviation frequencies of the modulated carrier. This depressing picture did not encourage further work in the field of frequency modulation, and in fact no great progress was made until the year 1936, when E. H. Armstrong presented a scientific paper which laid the foundation stone of the present frequency modulation techniques.

Deviation Ratio

It is a fortunate circumstance that although a frequency modulated wave is built up essentially of an infinite number of side-frequencies disposed about the carrier frequency, it is unnecessary to transmit a signal of infinite band-width to obtain a faithful reproduction of the modulation frequencies at the receiving station. The critical factor is what is known as the deviation ratio, defined as the ratio of the peak deviation frequency to the modulation frequency, i.e.:

$$\textit{Deviation ratio} = \frac{\textit{Peak deviation frequency}}{\textit{Modulation frequency}}$$

If the system can be operated so that the deviation ratio is either equal to or greater than unity, then any side frequencies existing outside the peak deviation will be of such reduced amplitude that their effects may be neglected. Thus a minimum requirement will be a peak deviation frequency equal to the highest modulation frequency employed, but although such a condition gives a reproduction of the modulation which is reasonably free from distortion, it will prove to be far more satisfactory if a greater deviation ratio is employed, especially when a particularly high quality of reproduction is required.

When considering the deviation ratio it is very easy to make the mistake of assigning some intrinsic relation to the two separate factors, deviation and modulation frequency. Again it must be stressed that this is not so, the deviation being a function of the *intensity* of the modulation. Bearing in mind the large number of broadcast services which will eventually require allocations of frequency with sufficient separation to avoid their mutual interference, it will be apparent that the design of the frequency modulation transmitter must aim at a deviation ratio which will permit an adequate reproduction of the modulation frequencies at the receiver, and yet radiate a signal which occupies as small a portion of the frequency spectrum as possible. It will be reasonable to suppose that a broadcast station would aim at a greater fidelity of reproduction (and hence a greater deviation ratio) than would a radio-telephone communication link where complete intelligibility and not high quality would be the first consideration.

Phase Modulation

It is interesting to note the relation between phase modulation and frequency modulation, since to some extent the two are inseparable. So similar are the two systems, in fact, that an inexperienced listener would find it a very difficult matter to detect any major difference in the quality of reproduction of which these systems are capable. With phase modulation it is the phase of the carrier which is varied by the modulator. The relationship which holds between the phase and frequency modulation methods can be expressed quite simply by the equation:

$$d = f_m . \varnothing$$

where d is the deviation and f_m the modulation frequency, both measured in cycles per second, and $\varnothing$ is the phase change, expressed in radians. The equation may be interpreted by saying that phase modulation is a special case of frequency modulation in which the deviation is proportional not to the modulation amplitude but to the modulation frequency. The constant of proportionality is the phase change angle. But the phase modulation is not simply of academic interest, it is of major importance, for although it has not achieved a very great measure of popularity in itself, it has one very important application in that a suitably designed circuit will produce pure frequency modulation out of a phase modulated wave. This method, originally worked out by Armstrong, is sometimes termed indirect frequency modulation or corrected phase modulation, and it provides a particularly interesting method for the production of frequency modulation.

Having now dealt with the main aspects of frequency modulation in their relationship with other systems, it is convenient to investigate the requirements in the design of apparatus both for the reception and production of such waves.

CHAPTER THREE

The Superheterodyne Receiver

For a number of reasons the superheterodyne receiver is to be preferred in modern radio communication systems. Although there may be some advantage to be gained from the use of a "straight" receiver circuit for certain fixed-frequency services, there can be no doubt that for outstanding all-round performance the superheterodyne is an obvious first choice. Following the general scheme of the previous sections it is intended that the superheterodyne used for the reception of frequency modulation signals shall be dealt with in comparison with the circuit employed in connection with an amplitude modulated service. Therefore it is necessary to make a quick revision of the superheterodyne receiver itself before considering how it must be adapted for the reception of frequency modulation.

Figure 4(a) is what is known as a block diagram and it illustrates the arrangement of circuits to be found in a conventional superheterodyne receiver. The signal, which is either unmodulated or amplitude modulated, is received at the aerial and usually is conveyed to the receiver by means of a transmission line. This permits the aerial to be situated high up and away from objects which are likely to screen it. The transmission line feeds the signal into the radio frequency amplifier stage of the receiver which comprises one or more valve amplifiers which increase the amplitude of the signal by voltage amplification. From the output of this stage the amplified signal is coupled to the frequency-changer circuit, where it is mixed with a second signal injected from a local oscillator unit. If the frequency of the incoming carrier is denoted f_c and that of the oscillator unit f_o, then the principle frequencies arising from the action of the frequency-changer circuit will be $(f_c + f_o)$, $(f_c - f_o)$, or if the oscillator is operating on a frequency higher than

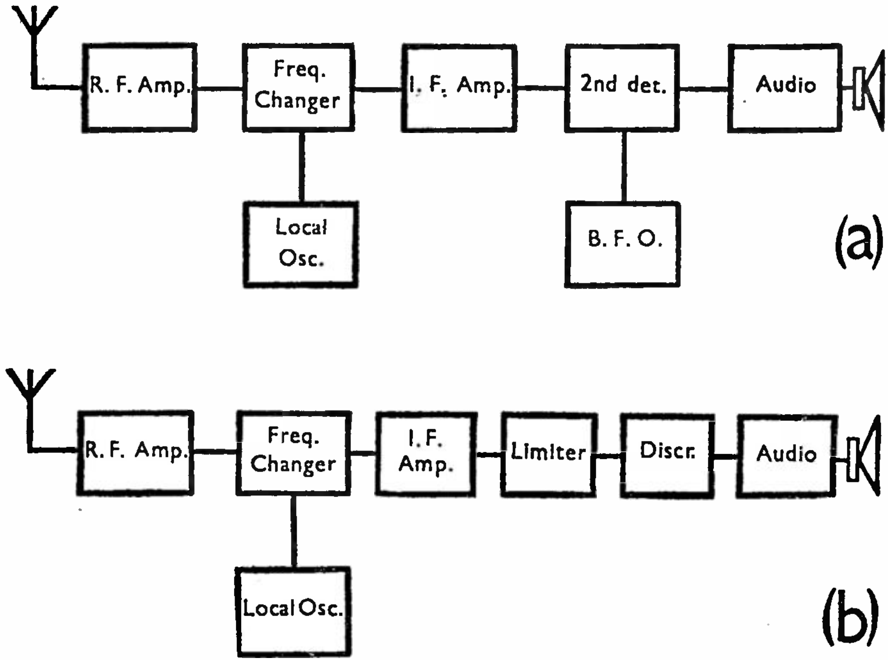

R. F. Amp.
Freq. Changer
I. F. Amp.
2nd det.
Audio
Local Osc.
B. F. O.
(a)
R. F. Amp.
Freq. Changer
I. F. Amp.
Limiter
Discr.
Audio
Local Osc.
(b)

FIGURE 4

that of the carrier ($f_o - f_c$). The frequency changer valve has a tuned anode load which is in resonance with one of these " difference " frequencies, usually ($f_o - f_c$), this being termed the " intermediate frequency." The voltage which appears, at any instant, across this load is taken to a series of amplifiers which operate at the intermediate frequency. The tuning controls on the local oscillator and radio-frequency stages are ganged together, and the circuits electrically " tracked," with the result that whatever the value f_c of the incoming carrier, the difference ($f_o - f_c$) remains the same, this permitting the intermediate frequency amplifiers to operate at a frequency which is always constant. This is a great advantage, since the amplifiers may be adjusted for peak performance at this frequency. After having undergone this amplification, the signal, now much increased in amplitude but changed to intermediate frequency, is taken to the demodulator or second detector stage. It must not be imagined that the process of changing frequency has eliminated the modulation from the carrier, for any amplitude changes inherent in the received wave will exist in the wave at intermediate frequency. It is the function of the second detector stage to separate the radio and audio frequencies. In this stage, therefore, the circuit is so designed that the mean value of current through the valve rises and falls in sympathy with the amplitude variations of the signal, and thus the current which flows through the valve load follows the " envelope " representing the modulation frequencies. This means that for a resistive load, at any rate, the voltage across it is a faithful replica of the modulation waveform, and this may be extracted for further amplification by the audio amplifiers which follow this stage. The final audio stage is normally a power stage feeding a loudspeaker unit. During the reception of morse signals which are at constant amplitude there can be no change in the mean current through the second detector valve, except an instantaneous one when the carrier is keyed on and off. Hence no periodic voltage variations appear across the load resistance. A second local oscillator must be used in this case, known as a " beat-frequency " oscillator, and this emits a signal which produces audible beats with the signal frequency itself, and this audible tone may be amplified in the normal way. The following are a few points of some importance in the consideration of the superheterodyne receiver:

1. The amount of useful amplification which may be applied to any signal entering the receiver is governed largely by the noise level, most of which originates in the early circuits of the receiver and which increases with the signal as it passes through the various amplifier stages of the receiver. The ratio of " signal to noise " is therefore of great significance in the operation of the receiver. By far the greater part of this noise is produced by thermal agitation and " shot " effects in the radio-frequency amplifier stages, so that these circuits must be designed to give as much signal amplification as possible without a similar increase in the generation of noise. It is interesting to note here, for future reference, that most extraneous electrical noise entering the set by way of the aerial is made up of impulses which are strictly " amplitude " variations as opposed to frequency change effects.

2. The superheterodyne contains a large number of tuned circuits incorporated in its signal-frequency and intermediate frequency stages. When all these circuits are correctly aligned, the overall selectivity of the receiver (i.e., its ability to separate two incoming signals on adjacent frequencies) can be exceptionally high if the receiver is a good one. Operation in a condition such as this may mean that the whole of the side-bands necessary for the reproduction of good quality is not admitted, and so the higher modulation frequencies are not heard in the receiver. It is necessary, therefore, to compromise between sharpness of selectivity and quality of reproduction. This is essentially a problem of design of the equipment. The final choice depends, of course, on the requirements of the particular service in conjunction with which the receiver is to be operated.

3. In operating a superheterodyne receiver under conditions which demand the maximum possible selectivity, it is possible to use single-side band communication. With this system only one of the two side-bands is utilised, so that the effective bandwidth of the signal is halved. This may be brought about either by employing receiver tuned circuits with exceptionally sharp response curves or by modifying the transmitter design in order to suppress one side band in the signal radiated. It may be shown by resorting to mathematics that there still remain sufficient frequency components after the elimination of one side-band to give a faithful reproduction of the modulation frequencies. Furthermore, since power must be consumed in radiating the side-bands, the removal of one of them at the transmitter end must result in a power

saving which will be appreciable in high-power transmitters.

The Second Detector Stage

The function of the second detector or demodulator stage is to respond to changes in amplitude of the incoming signal and to pass them on to the audio amplifiers in the form of a voltage appearing across the valve load resistance and varying at the modulation rate. Thus the stage as a whole receives and transmits an amplitude change, so it will be of no value in a receiver designed for the reception of frequency modulated signals, since the amplitude of these signals will not vary. The requirement now is a device which accepts frequency variations and converts them to amplitude variations before passing them on to the audio stages for amplification. When this fact was realised it was thought that as electrically-produced noise is largely an amplitude phenomenon, the inclusion in the circuit of a device which responded only to frequency changes would effectively remove all receiver noise. To what extent this condition is fulfilled will be more apparent when the frequency modulation receiver has been discussed in greater detail.

CHAPTER FOUR

THE FREQUENCY MODULATION RECEIVER

In the main there is very little difference between the receiver employed for frequency modulated signals and that required for amplitude modulation. The frequency modulation arrangement is illustrated in figure 4(b), and on examination it will be observed that the second detector stage is now replaced by a unit known as the discriminator, this being the frequency to amplitude conversion device. Preceding this stage is a unit which has no counterpart in the normal receiver. It is the amplitude limiter, usually referred to simply as " the limiter." With the exception of these two stages the receivers are almost identical in every other detail. This suggests that it is possible to construct a receiver which could be employed for the reception of either system simply by selecting the appropriate circuits by means of suitable switchgear.

THE DISCRIMINATOR

The fundamentals of the discriminator circuit are best described by reference to actual circuits, firstly taking one of the earliest models ever to be used in this connection and then discussing some more modern circuits. The basic requirement is a circuit with an output voltage which is directly proportional to the frequency change of the incoming signal. In other words the frequency-amplitude conversion must be linear, and moreover a rather difficult analysis indicates that unless the conversion efficiency is high, many of the advantages inherent in the use of frequency modulation are lost. One of the first discriminators to be used is illustrated in figure 5. It

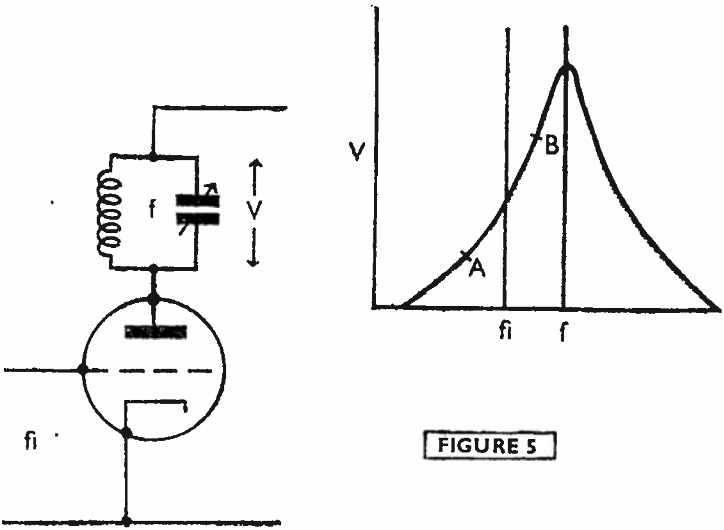


FIGURE 5

is really a simple parallel-tuned circuit connected in the anode of a valve which receives the intermediate frequency in its grid circuit. The tuned circuit is resonant either just above or below the intermediate frequency, and in the case shown it is tuned to the higher side. If the intermediate frequency is f_i, then the incoming signal which is assumed to be frequency modulated swings about this mean frequency. The voltage appearing across the tuned circuit is governed by the shape of its response curve. Referring again to the figure, if the deviation is confined to the region AB a sensibly linear response is obtained. This circuit has a number of disadvantages and was soon replaced by more advanced methods. For instance, the conversion efficiency depends upon the steepness of the response curve, and there are serious practical limitations in increasing this gradient. Secondly, the output voltage is necessarily of a low value because the circuit must be operated in a detuned condition, which prevents the maximum amplification being realised. The deviation which this circuit will handle is restricted by the fact that on the low-frequency side the output rapidly approaches zero, whilst on the high-frequency side a maximum output is soon attained, after which it falls again as the deviation increases. Only the range AB is a safe working deviation.

Another circuit is of the type frequently employed for the automatic control of oscillator frequencies, such as the local oscillator of a receiver. The circuit—a frequency discriminator—has an output voltage which is proportional to the incoming frequency variation referred to some chosen standard, i.e., it is proportional to the deviation. This output voltage is applied to one of the oscillator valve electrodes, usually the grid, in the form of a D.C. bias which restores the frequency to the reference value. If the discriminator is correctly adjusted, it will give zero output voltage when the incoming signal is at its required value, so no bias is produced and no restoration action takes place. This type of circuit is obviously well suited to the demodulator stage of the frequency modulation receiver for it apparently fulfils all the essential requirements of this stage. It is met very frequently in two main forms, one an amplitude or double-tuned discriminator, and the other a phase-difference circuit. Both are of sufficient importance to warrant a detailed account of their action.

The Double-tuned Discriminator

In this circuit a valve is fed from the output of the limiter stage and contains in its anode two tuned circuits connected in series. Sometimes these circuits are connected in the anode of the limiter itself, but whatever the arrangement they are parallel tuned, with the result that the voltage across either circuit will reach a maximum when the frequency fed to the grid of the valve is the resonant frequency of the particular circuit. The tuning can be arranged so that one circuit is in resonance at a frequency above the incoming intermediate frequency, and the other at a frequency an equal amount below it. In this way, approximately equal voltages are produced across the individual circuits when the incoming signal is, for the sake of example, 1,000 cycles per second above the intermediate frequency and 1,000 cycles per second below it. The outputs from the individual loads are taken separately to a pair of diode detector valves which have resistive loads so arranged that the voltages which appear across them are in opposition. The advantage of this circuitry is that when the incoming signal is at intermediate frequency, the output is zero, but for a signal of given deviation from this frequency, an output voltage will be produced which is positive in the case of one diode load and negative for the other, the effect being to double the output voltage which could be obtained from any one tuned circuit connected in the previous valve. An exact voltage doubling would only be obtained if both the diode units had identical characteristics, but by an adjustment of the values of the load resistances, this is not too difficult an achievement. Figure 6 shows the basic circuit of a discriminator of this type, and the diodes are seen to be connected " back to front " to give the voltage opposition across the resistive loads R_1 and R_2. The purpose of the condensers C_1 and C_2 is to provide a low impedance path for the radio-frequency component and yet offer high impedance to audio frequencies, i.e., to a direct current varying at an audio rate. The figure includes a typical response curve, and it is the gradient of this curve which determines the efficiency of the conversion. The linearity of the discriminator also depends upon this curve, so the design of the tuned circuits is very important. Once again it will be necessary to confine the deviation to the range AB for a linear relation between output voltage and deviation. This circuit is very satisfactory in many ways, but it has been discarded to a large extent in favour of the phase difference discriminator which, though similar in its construction and general characteristics, is rather simpler in its construction. There are, in fact, a very large number of variations of the double-tuned discriminator circuit.

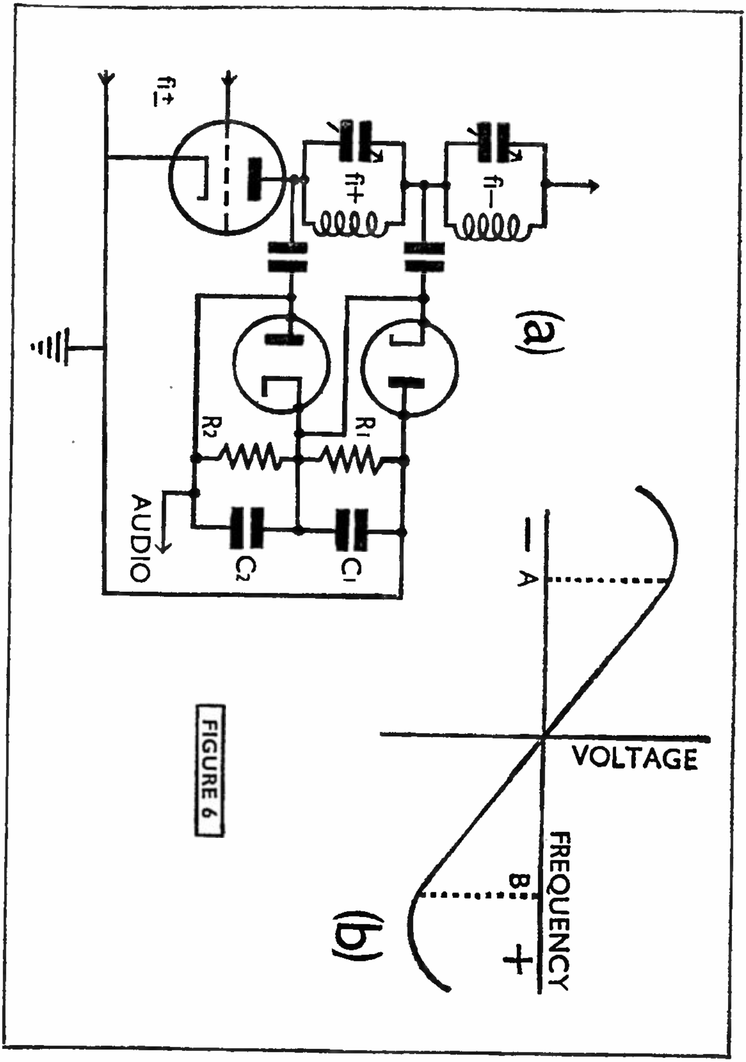


FIGURE 6

The Phase-difference Discriminator

This circuit is frequently referred to as the Foster-Seeley discriminator, being named after its designers who, in 1937, introduced it as a means of applying automatic frequency control to the local oscillator of a receiver. The design and action of the circuit is centred upon a specially designed transformer which is included in the anode circuit of the limiter valve. This transformer is operated at intermediate-frequency, and has both primary and secondary tuned for resonance at this frequency. In addition, the secondary is centre-tapped and feeds two diode detectors as before. Further coupling between primary and secondary is achieved by means of the small condenser C which is indicated in figure 7. The diodes

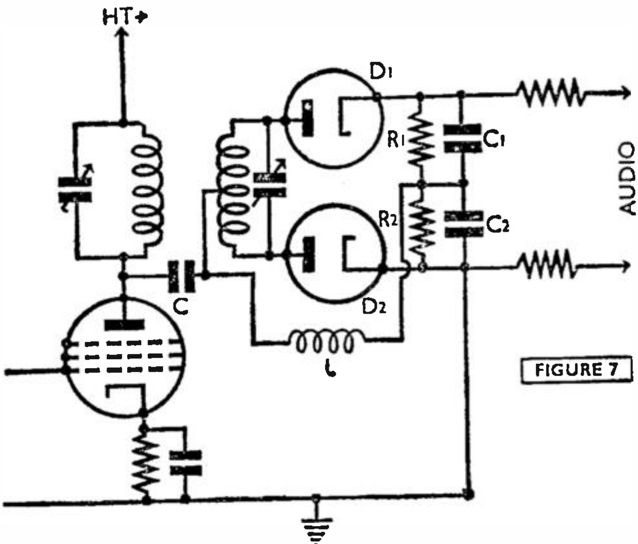


FIGURE 7

D_1 and D_2 are employed with their load resistances in series, the junction of these loads being connected back to the transformer secondary centre-tap via a small inductance L, actually a radio-frequency choke. With this arrangement, each diode receives one-half of the voltage appearing across the transformer secondary, while the primary voltage appears across the radio-frequency choke due to coupling via C and C_2. The output D.C. voltages from the two loads are connected in opposition, and if the circuit is correctly aligned to the mean frequency (which is of course the intermediate frequency), then the voltages across R_1 and R_2 will be equal but opposite and no voltage is registered across the extremities of the combined loads. A point of interest is that when receiving the mean frequency, the discriminator is completely insensitive to amplitude disturbances because it is a balanced circuit and the output is effectively cancelled by the opposing diodes. The analysis of the circuit action is somewhat complex mathematically, but the main features depend upon the fact that the voltage across the transformer primary is out of phase by 90° (either lagging or leading) on the secondary voltage, so that on a graphical representation the two voltages would be out of step by one-quarter of a cycle. Phase changes in the primary are not directly related to the circuit action, however. In the secondary winding a complicated sequence of events occurs. There are, for example, two methods by which the secondary is energised, viz., by the direct capacity coupling C and by the mutual induction of the two windings. When subjected to the mean frequency, the voltage and current in the secondary winding are in phase, but immediately a deviation occurs this simple phase relationship is disturbed. Briefly, the operation of the circuit is as follows. Suppose the mean frequency to increase under the action of the modulation. The signal received by D_1 also increases and exceeds the value of signal applied to D_2. Thus the voltage across the load R_1 will exceed that across R_2, and let it be assumed that this is a positive voltage with respect to the junction of the loads. If the mean frequency now swings to the other side and becomes less than the mean value, then the voltage across R_2 exceeds that across R_1, and a negative voltage results. The general effect will be a response curve of output voltage against frequency which has the same shape as that obtained for the double-tuned discriminator, figure 6(b). This curve is, in fact, a generalised response curve for all discriminators. There are, of course, a number of rather difficult design problems to be overcome in the production of a well-balanced discriminator. The choice of the inductance to capacity (L/C) ratio of the transformer windings is very important, since the design of these components is restricted by the fact that the circuits must resonate at the mean or intermediate frequency. Therefore, what theoretically might prove to be a suitable number of secondary turns for an efficient conversion, could well be an inefficient arrangement from the point of view of the optimum L/C ratio. The following points will prove very useful as an aid to the alignment of a phase-difference discriminator circuit:

(1) The primary tuning changes the symmetry of the response curve. If it is tuned to too high a frequency then the output peak at the higher

frequency is of greater amplitude than that at the lower frequency.

(2) The secondary tuning controls the mean frequency, at which the output voltage will be zero.

(3) The mutual coupling between the transformer windings governs the frequency separation of the positive and negative peaks of output voltage. As the coupling increases, so will the frequency separation.

Since these operations are more or less independent of one another, the adjustment of the discriminator characteristic is not a very difficult task.

The Amplitude Limiter

For the frequency modulation receiver to operate under the best possible conditions, steps must be taken to eliminate any amplitude variations which might reach the audio amplifier stages.

Since the audio stages are concerned with the amplification of amplitude changes resulting from the discriminator action, any unwanted amplitude modulation of the carrier by noise, interference, etc., should be reduced before the signal enters the discriminator stage. For this reason a circuit known as the amplitude limiter is included in the receiver, usually occupying a position in the circuit at the end of the chain of intermediate-frequency amplifiers, and having the discriminator transformer primary winding as an anode load. There are numerous methods of limiting amplitude. One circuit utilises an oscillator which has a more or less constant output voltage, but whose frequency is controlled by the incoming frequency-modulated signal. The oscillator follows the excursions of carrier frequency, but maintains a constant amplitude. This method is quite effective, but for most purposes is too complicated, and as a result it has become common practice to employ the simplest of all limiter circuits, the saturated valve-amplifier.

The Saturated Valve-amplifier as Amplitude Limiter

A typical saturated amplifier limiter is illustrated by the circuit diagram of figure 8(a). In essence it is a leaky-grid detector in which the grid condenser C_1 and the grid leak R_1 have a combined time constant which is very small, usually of the order five to ten microseconds. Suitable values for these components would therefore be 100 micro-micro-farads for C_1 and .05 to .1 megohm for R_1. The transformer T_1 in the grid-cathode circuit is an intermediate-frequency transformer with its primary winding in the anode of the last intermediate-frequency amplifier valve. T_2 is the discriminator transformer. The value of the common anode and screen voltage on the limiter valve is purposely kept very much lower than the usual operating potentials of the valve, and normally is not much more than 50 volts. If the input signal suffers an increase in amplitude, the bias voltage appearing across the grid leak R_1 becomes more negative and the anode current of the limiter falls. By a correct choice of circuit constants, the bias voltage is made to follow any amplitude changes very closely, and it is to permit the handling of sudden bursts of noise or interference that the time constant of C_1 R_1 is small

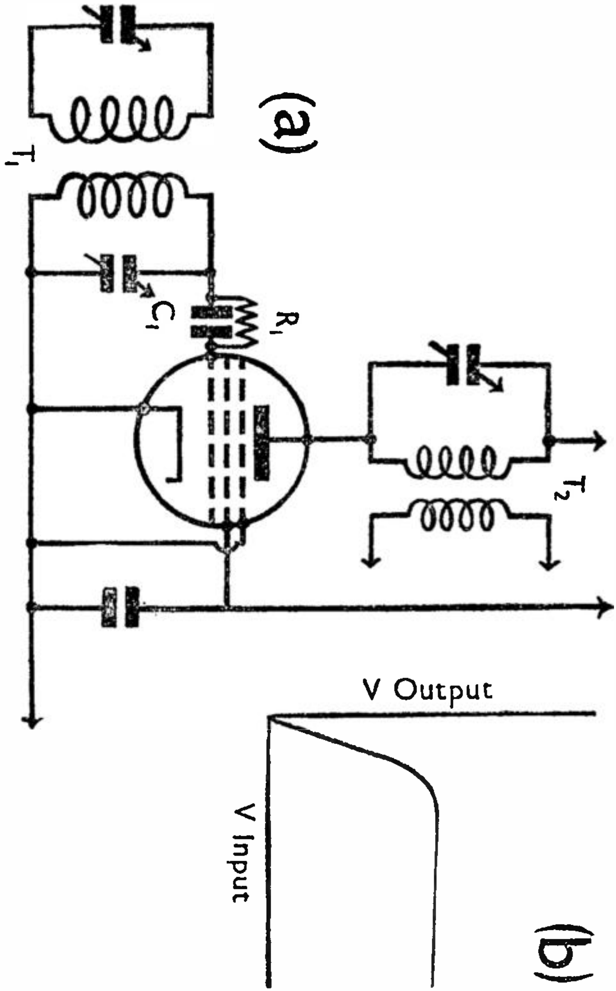


FIGURE 8

and therefore the recovery of this network is very rapid. For the limiter to operate in this manner the valve and the part of the characteristic employed must be chosen very carefully. A typical limiter curve is illustrated in figure 8(b), and it will be apparent from this diagram that, provided the input voltage to the stage is sufficiently high, a completely flat response is obtained. The high input voltage condition is satisfied by positioning the limiter stage after the intermediate-frequency amplifiers, at which point the signal has acquired a considerable amplitude.

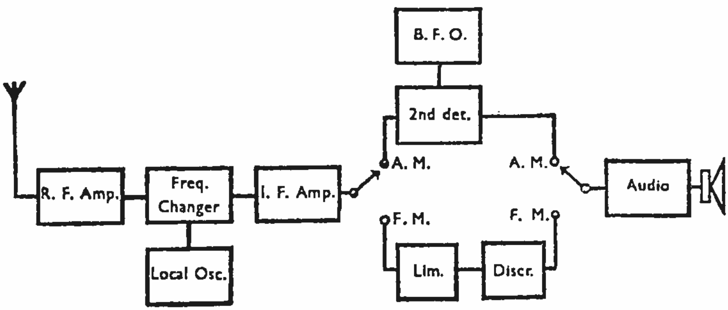


FIGURE 9

THE DUAL-SYSTEM RECEIVER

Reference has been made to the similarity which exists between receivers for amplitude and frequency modulation systems, the two being identical with the exception of the units which have been discussed in detail. Apart from these, the only other particular in which the frequency-modulation receiver may differ from the more usual type of receiver is that its tuned circuits must have response curves which admit the deviation appropriate to the received signal. This applies mainly to the intermediate-frequency stages, since the R.F. circuits will usually pass the comparatively small deviation. Figure 9 will suggest a possible method of switching which may be incorporated in a receiver which is required to operate on both systems. There is nothing complicated about the arrangement, and in fact a number of receivers have already been produced to this general design.

CHAPTER FIVE

The Radio Transmitter

All radio transmitters follow the same general scheme of design, the essential differences between transmitters being the actual circuits of the individual stages. Any piece of apparatus which is required to transmit a modulated radio wave will incorporate the following units:

1. A master-oscillator unit, consisting of an oscillator of very high stability. It is often frequency-stabilised by the application of piezo-electric crystal control, though for certain

frequency modulation circuits this will not be permissible. The function of the oscillator is to initiate the carrier wave, the frequency at which the oscillator operates being equal to (or, at any rate, related to) the frequency at which the carrier is radiated.

2. A number of amplifier stages which raise the output of the oscillator to a useful level. The last amplifier in the chain is of the power type, for it is this stage which energises the transmitting aerial. When, and this is frequently the case, the master-oscillator valve is operated at a frequency which is a sub-harmonic of the carrier frequency, the amplifier chain includes an appropriate number of frequency-multiplier stages.

3. The modulator, and in this must be included the speech and microphone amplifiers, or other units depending upon the type of intelligence to be radiated. The modulator design is naturally determined by the modulation system for which the transmitter has been constructed.

4. Power supply units, required to feed all stages with both high-tension and valve filament supplies.

5. Some form of aerial system, necessary for the radiation of the wave.

Once again the most convenient treatment will be to consider, if somewhat briefly, the chief features of the amplitude modulated transmitter, for its principles are well-established and its efficiency proved by long service. Until recently, this system was employed more or less to the exclusion of all others, and it therefore provides a standard against which we may compare the frequency modulation equipment.

The A.M. Transmitter

The amplitude modulation transmitter incorporates all the units previously specified, but it is a feature of this type of modulation that the modulator, in almost every case, acts in conjunction with the final amplifier stage of the transmitter, i.e., the stage driving the aerial. There are two or three variations in the way in which the modulator may be connected to this stage, however. If, for example, the modulation frequencies are fed into the grid of the amplifier valve, the result is a grid-modulated amplifier, the modulator output causing variations in grid-bias voltage on this stage. These variations are in sympathy with the modulation wave. Another method, the most popular, is to couple the modulator into the anode of the amplifier valve so that an anode effect is produced on modulation. It must not be thought that these two methods are themselves modulation systems. They are simply alternative ways of producing an amplitude modulated wave. An important point which now arises is that whatever amplitude method is selected, the modulator itself must be capable of providing considerable audio power. When anode-modulation is used, the modulator should be able to provide audio power which is at least 50% of the power input to the final amplifier. The grid system is less exacting in its demand for power, requiring only about 10% of the amplifier input. This advantage is offset to a great extent by the fact that a grid-modulated amplifier is a difficult design problem, and its adjustment a highly critical process if a good quality of reproduction is essential. The problem of audio power becomes quite serious when one considers transmitters of over 200 watts input to the final amplifier. Taking the extreme case of a broadcast station, with input rated in *kilowatts*, the provision of several *kilowatts* of audio power is not only expensive, but a difficult piece of radio-engineering.

It would be possible, perhaps, to effect a saving in audio power by applying the modulation to a lower power stage in the transmitter, but this involves such a careful design of the following radio-frequency stages that it is not altogether an acceptable solution.

It will be a very strong point in favour of the adoption of the frequency modulation system if it should transpire during this discussion that high audio powers were unnecessary when using that system.

The F.M. Transmitter

In the successful design of the frequency-modulated transmitter, three fundamental requirements must be fulfilled. In the first place, the mean frequency must remain constant throughout the period during which the modulation is applied, for otherwise there would be a frequency shift of the received signal. Next, the deviation must be proportional to the amplitude of the modulation wave, and be unaffected by the actual modulation frequency. Finally, the carrier amplitude should be unchanged by modulation. These factors are controlled by finer points in the design of the equipment, and any attempt to discuss them would be beyond the scope of this general treatment. It is sufficient to record that modern circuit techniques are capable of satisfying these requirements.

Frequency modulation is applied to a low-power stage of the transmitter, and this suggests that high audio power will not be essential. It is the *frequency* of the carrier which is to be varied by the modulator, and therefore the audio frequencies may be fed into a circuit at the very head of the chain of stages, i.e., to the master-oscillator unit, for it is this unit which governs the frequency of the emitted wave.

The two main methods of accomplishing frequency modulation are the direct and indirect systems.

Direct F.M. (The Reactance Valve)

For direct frequency modulation, the master-oscillator should not be crystal controlled, for then the stability would be so great that frequency variations could not be produced to any extent. For this reason it is customary to employ any good self-excited oscillator circuit, the oscillation frequency being controlled by a parallel-tuned circuit of inductance and capacity (i.e., a " tank " circuit). It is convenient to regard the tank circuit as a pair of reactive elements in parallel connection, so that a change in magnitude of the reactances causes a change in the frequency of the oscillator, and a subsequent frequency shift of the carrier wave. Furthermore, a periodic reactance variation, if this might be accomplished, would produce sympathetic changes in carrier frequency. Since this is precisely what is required of the frequency modulation transmitter, it remains simply to design a circuit which periodically changes its reactance when

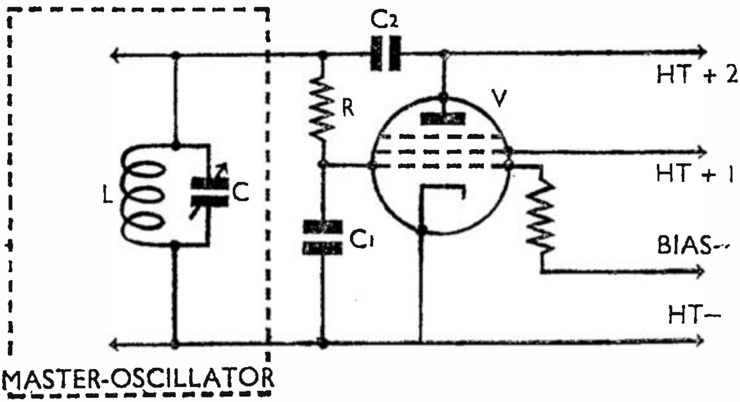


FIGURE 10

energised by the output from the modulator. Connecting this circuit across the oscillator tank circuit would then produce the required frequency excursions of the carrier wave.

Fortunately, this problem may be solved by a relatively simple piece of circuitry, but before describing this, it is of historical interest to mention what is possibly the most elementary of the variable reactance devices. An instrument similar to a telephone earpiece is connected up so that the output from the modulator energises the magnetising coils. The diaphragm, which may be set up as one electrode of a parallel-plate condenser, experiences a backward and forward movement which follows the current changes through the coils, and so the capacity of the condenser will itself vary according to the same sequence of variations. Connecting this condenser across the oscillator tank circuit would produce frequency modulation of the oscillator output waveform. This is a somewhat mechanical device, and needless to say it soon gave way to the products of a more advanced circuit technique which led to the development of a purely electrical version of the variable reactance element. The reactance valve, as it was named, utilises a valve of high internal resistance, usually a pentode, connected in such a manner that its output, anode to cathode, is reactive. The magnitude of this reactance is controlled by the grid voltage of the valve. Another operating condition for the circuit is that the current through the valve must be out of phase by 90° on the current which flows through the oscillator tank circuit, across which the reactance valve is to be connected. This may be accomplished quite simply by employing the circuit of figure 10, which is intended only to illustrate the essentials of the design and is not a practical circuit. The resistance R is necessarily large compared with the reactance of the condenser C_1 at the frequency of resonance of the tank circuit LC of the master-oscillator. Under these conditions, the voltage across C_1, which incidentally also appears across the grid-cathode circuit of the reactor valve, lags on the voltage across the tank circuit LC by 90°. At resonance, the current through the tank circuit is in phase with the voltage across it, and further, the grid voltage and anode current of the valve are in phase. This means that the current through the valve will lag by 90° on that through the tuned circuit, so that the condition of operation is satisfied, and the output is reactive. If, by some means, the anode current is varied at the modulation frequencies, the output displays a sympathetic reactance variation. This is usually performed by feeding the modulation in series with the grid-bias supply to the valve, the changes in bias modifying the flow of anode current. The modulator is therefore required only to generate a voltage which varies at the modulation rate and which is fed to the grid of the reactor valve to produce in it changes of anode current. This voltage need not be of very large magnitude, and obviously no very great expenditure of audio-power is required to produce the desired reactance changes which bring about the frequency modulation. The deviation which results from this simple method should be quite linear with applied grid voltage provided that operation is confined to the linear region of the reactance valve anode-current/grid voltage characteristic. Since the deviation is obviously dependent upon the amplitude of the waveform applied to the grid, reasonably pure frequency modulation should be obtained.

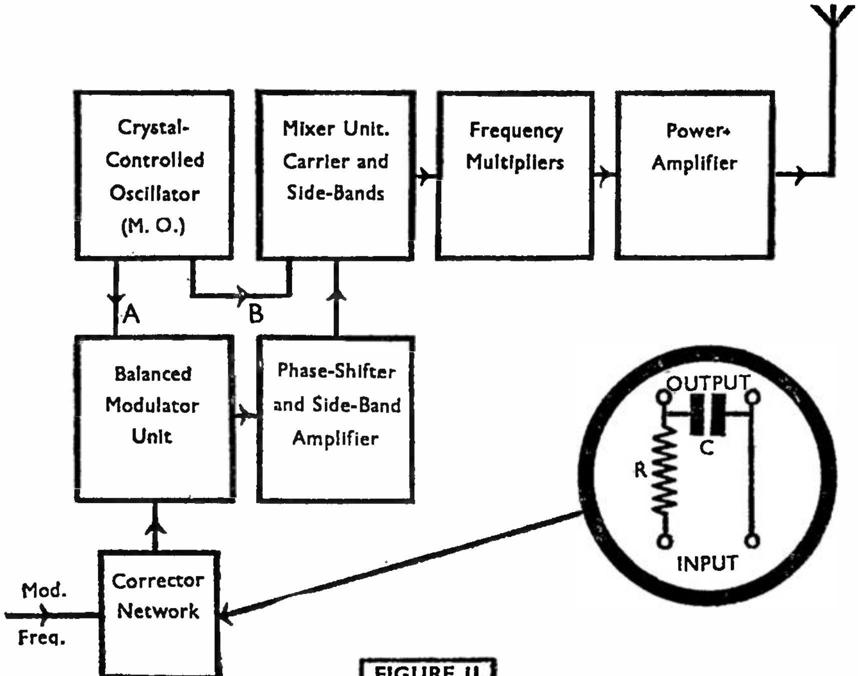
As one might expect, certain disadvantages are encountered which limit the application of this simple circuit. For example, in any valve circuit there are resistive elements present due to the very nature of the valve itself. In the reactance valve circuit such elements give rise to a certain amount of unwanted amplitude modulation. This effect can be overcome by the use of a more ambitious circuit employing two valves in a balanced variable reactance circuit. Another point, possibly of less importance, is that the reactance valve is unsuitable for use in conjunction with master oscillators operating at ultra-high frequency unless the oscillator has been designed for exceptional frequency stability. The usual solution which is adopted is to employ a master-oscillator on a lower frequency and to follow it by sufficient frequency multiplier stages to give the required carrier frequency. It must not be overlooked that the deviation itself is multiplied the same number of times, and the excursions of the oscillator must be restricted to a frequency band, which, after the appropriate multiplication, is of the desired band-width. Deviations of some 75 kilocycles per second have been a popular choice during recent years.

As an interesting example of circuit application, the reactance valve circuit is sometimes found in conjunction with the phase-difference discriminator. In this way, any frequency drift due to the oscillator is corrected by the discriminator circuit, so that the mean frequency is reasonably constant during modulation.

Generally speaking, the reactance-valve modulator is not suited to the precision requirements of a broadcasting station, and therefore such installations usually prefer to use the more complex system of indirect F.M.

The Armstrong Modulation System

Mention has been made of the way in which phase modulation may be employed to produce pure frequency modulation. This is the basis of the indirect system of modulation developed by E. H. Armstrong, and it is frequently referred to as the Armstrong Phase Modulation system. Like any other method, Armstrong's has certain inherent disadvantages, but it is well suited to the broadcast station because of the very great stability of mean frequency which is a feature of the circuit. It is based on a master-oscillator unit which is crystal-controlled. The point was raised previously that the reactance valve does not lend itself to the modulation of an oscillator which is crystal controlled, because the stability of such an oscillator would be far too great. Although it must be expected that future research will produce other modulation methods, there is no doubt that at this time the Armstrong circuit allows a far greater frequency stability than any other method. Once again the block diagram has been chosen to illustrate the circuit arrangements (figure 11). The sequence of events during the operation of the modulator is as follows. The output from the crystal-controlled master-oscillator unit is divided to feed two separate channels. With reference to the figure, let us first consider path A. The radio-frequency signal enters a balanced amplitude modulator, which is a clever piece of circuit design. Its amplitude modulates the wave, forming the upper and lower side-bands in the usual way, but arranges for the carrier frequency to be cancelled out, so that the wave leaving this stage is composed of a pair of side-bands only. These are taken into a unit where a small condenser shifts their



Crystal-Controlled Oscillator (M. O.)
Mixer Unit. Carrier and Side-Bands
Frequency Multipliers
Power-Amplifier
A
B
Balanced Modulator Unit
Phase-Shifter and Side-Band Amplifier
OUTPUT
C
R
INPUT
Corrector Network
Mod. Freq.

FIGURE II

phase 90° with respect to the carrier itself. The remainder of the unit amplifies the side-bands. The carrier proper in an unmodulated state is meanwhile taken into a mixer unit via path B. The side-bands are also fed into this mixer, and it can be shown that the result of mixing a carrier with side-bands which are 90° out of phase is a phase-modulated carrier wave. Phase modulation, it will be remembered, is a special case of frequency modulation, the deviation being directly proportional to the modulation *frequency*, instead of being related to the modulation *amplitude* as it must be for pure frequency modulation. The introduction of a very simple network makes possible the conversion of the phase modulation to pure frequency modulation. This network is labelled the " Corrector network " in the figure, and the inset shows a very simple combination of a resistance R and a condenser C which would be sufficient to perform the conversion. Suppose that R is a high resistance, and C is of such value that its reactance is low even at the lowest modulation frequency employed. Then the voltage appearing across C will depend upon frequency, so we have produced an amplitude effect from a frequency variation. To be more specific, the output from the corrector network is inversely proportional to the modulation frequency. The overall result is to produce frequency modulation, (corrected phase modulation in this case). Armstrong's method, however, produces a deviation of very small order, say 25 cycles per second only. This is one of the greatest objections to the circuit, for in order to provide a carrier deviation of 75 kc/s (75,000 cycles per second), the carrier must be multiplied up in frequency some 3,000 times! The multiplier stages are indicated in the figure, the output finally driving a power amplifier which energises the transmitting aerial. Taking the same figure of a multiplication of 3,000, in order to produce a carrier, on, say, 30 megacycles per second, the master-oscillator would have to function at 3 kc/s! To multiply by this amount naturally complicates the circuit and involves an increase in cost. It is not very satisfactory to multiply by more than four times in any one circuit, so that for a multiplication of even 256 times, four stages are required.

The authorities controlling the allocation of frequency channels to broadcast services are particularly concerned that the carrier frequency stability of the emission should be of a very high order. For this reason the Armstrong modulator is to be found in most installations of any importance.

CHAPTER SIX

Radio-wave Propagation

A radio signal radiated from a transmitting aerial can arrive at a remote point via a number of paths. There is first the direct path, the ground-wave, the signal passing directly between the two stations. If we assume that the path of a radio wave in space is more or less a straight line, then the distance over which the ground wave is able to be received is limited by the earth's curvature. This is not to suggest that distances greater than the " horizon " cannot be covered by a ground wave, for other factors such as refraction enter into the argument, but this is one reason why the coverage is restricted. In addition to the direct path, there is the indirect ray which arrives at its destination after reflection from the ionosphere. This is a ray which, travelling outwards from the earth, would be lost, unless an ionised layer in the upper regions of the atmosphere reflected it back towards the earth. This, of course, happens frequently, the whole concept of long-distance communication on the short waves being dependent upon this phenomenon. From the point of view of a frequency modulated wave, the importance of the dual path is to be found in the different distances which the direct and indirect rays would travel if received at the same point. Roughly speaking, the problem is one of triangulation. The two points forming the base of the triangle represent the sending and receiving stations respectively. The apex of the triangle is the point of reflection from the ionised layer. The base line is the direct ray and the other two sides —together greater than the third—are collectively the indirect ray. Clearly there must be a path difference between the rays. Since, also, the wave is frequency modulated, it is changing frequency many hundred times per second, so that very complicated phase differences exist between the two rays. This is selective fading, and it is responsible for distortion in the reproduced signal, for obviously the receiver has to cope with the effects of both rays, and the reproduction suffers accordingly. The distortion is usually worse with a frequency modulated signal than with one which is amplitude modulated. This fact rules out frequency modulation transmissions on the short-wavebands. The width of the channel required for the transmission of the side-bands rules out the medium wavelength region of the spectrum, and so we are left only with the ultra-short wavelength region for the allocation of frequency modulation channels. The region most favoured up to 1939 was that between 40 and 50 megacycles per second, and allowing each service some 200 kc/s (i.e., deviation ±100 kc/s maximum), this would provide 50 separate channels. These wavelengths are suitable for working over " optical " ranges, i.e., to the horizon, and if the transmitter is set on a high point the coverage can be considerable. Unfortunately, considerable electrical noise from automobile ignition and electrically operated machinery is experienced when operating at these wavelengths.

As regards the aerial systems, no special types are required, but any good ultra-high frequency aerial will be satisfactory. Whether or not a beam aerial is employed depends upon the requirements of the station. Prior to 1939, most of the aerials used seemed to be of the horizontal type, but there seems to be no reason why the vertical aerial should not be used, except, perhaps, that most impulsive noise is vertically polarised. Using a horizontal aerial would minimise its effects.

"Capture" Effect

The "capture" effect is one which is peculiar to frequency modulated signals. If two signals occupy the same channel, interference is inevitable to some extent. When both signals are amplitude modulated, the effect is often to render both unintelligible. When the signals are both frequency modulated, however, the stronger signal completely "captures" the weaker, and suppresses it. This normally requires that the signal strength of the two carriers should differ by a ratio of about 2 : 1 which, incidentally, is not as great a difference as might be supposed. If the signal strengths are identical, and the channel common to both signals, both will be rendered unintelligible, and the only solution is to separate the channels or to beam the radiations in different directions in order to bring about the required difference in signal strengths received.

Interference and Noise

Considering first the noise inherent in the receiver itself, it has been stated already that the bulk of it originates in the first stage of the receiver. It is a fact that when a carrier is received the noise level should fall, because of all the complex frequency components of which the noise is composed, only those with frequencies differing from the carrier by an audible frequency will have any effect when the carrier is received. The importance of the signal-to-noise ratio has been stressed, but it is interesting to observe that for the frequency modulation system, the reproduced noise is only about half that present in the case of a similar amplitude modulated system. This noise reduction is one of the main advantages of frequency modulation. The above figure of a 2 : 1 reduction in noise is based on a deviation ratio of unity. The ideal ration (2 :1) is never achieved in practice, but Armstrong gives the practical value as 1.7 : 1. For higher deviation ratios, the improvement in noise reduction is obtained by the ratio

1.7 × Deviation Ratio : 1

This again is an idealistic value, but it suggests that with a reasonably large deviation ratio, the noise reduction can be improved very considerably.

The presence in the circuit of the amplitude limiter which suppresses all amplitude variations is a further aid to noise elimination, because most extraneous noise is of an " amplitude " type of disturbance, whilst receiver noise is produced as a result of the beating together of individual " noise frequencies," the resultant beats being amplitude effects. The noise, after being " ironed out " by the limiter, may give trouble by causing a phase modulation of the signal, however. The theory of the circuit action under the combined effect of frequency modulated signal plus noise indicates that the noise will be least suppressed during the period when the higher audio frequencies are being reproduced. Briefly, this is because the greater the noise level, the greater the phase modulation effect after limiting. For a given depth of phase modulation, the deviation produced will be greater for higher modulation frequencies. This follows from the simple relation between phase and frequency modulated waves. Therefore, during high modulation frequencies, the signal-to-noise ratio decreases, which is unfortunate since the optimum signal-to-noise occurs for modulation frequencies which are very small, e.g., 10 to 100 cycles per second, and these are seldom, if ever, employed. One way out of this problem is to use the *pre-emphasis* method at the transmitter, whereby the higher modulation frequencies are given extra amplitude. Then, at the receiver, corresponding *de-emphasis,* a reduction of the amplification given to higher frequencies by the audio amplifier, restores the balance. Careful design will give a resultant flat response curve to the audio amplifier and yet avoid the loss in signal-to-noise during " highs " in the modulation.

For the case of a wave which is interfering with another on a nearby frequency, the effect is usually to produce audible beats by the heterodyne action. These beats are amplitude variations, and therefore the circuit deals with them as for an impulse of noise. As it may well be appreciated, the whole problem of what happens in the circuit under conditions such as have been described, is not the simplest of matters. The only satisfactory treatment of the case is by resort to mathematics, and in particular a vector analysis of the problem can be very enlightening.

F.M. VERSUS A.M.

To sum up, the main advantages of frequency modulation over the amplitude modulation system will be mentioned. There are possibly three advantages worthy of attention, but it must be realised that F.M. is really in its infancy, and the whole subject is, to a certain extent, controversial. Quite a number of qualified engineers are not yet convinced that the frequency modulator is everything that it has been made out to be. Eventually, when the system has been fully exploited, sufficient scientific evidence will be available to settle the matter one way or the other. At the moment the choice is perhaps purely a matter of opinion. However, these are the three factors on which the claims for supremacy of frequency modulation are based:

(a) The improvement in signal-to-noise ratio under the conditions which have been stated. To anyone who has listened to a frequency modulation broadcast for any length of time, there can be little doubt as to the justification for this claim.

(b) The reduction of interference between stations operating on adjacent frequencies. This advantage must be considered in the light of all the available evidence, not only from frequency modulated signals but also that which may be brought to light by an analysis of the theory of the interference of two amplitude modulated waves. We have seen that the effect of an amplitude variation is to produce a phase modulation of the carrier, with the result that the interference will be greater at frequencies furthest away from the mean. That is to say, that two frequency modulation signals will interfere with one another to a lesser degree if they are, for the sake of argument, 1 kc/s apart than if their carrier frequencies differed by, say, 5 kc/s. The separation of 5 kc/s is taken to be about the worst possible from the point of view of interference. Therefore two systems can be operated very close together without serious interference. In addition to this, the " capture " effect will reduce the weaker of two adjacent signals. The best arrangement therefore appears to be:

(i) Set up a station to serve a particular area only, by direct ray transmission, and " beam " the signal on that area to improve the reception. The maximum range, assuming a signal in the neighbourhood of 40 mc/s, would be some 100 miles. Any signal not beamed on the area, or one beamed on it but originating from a point more than 100 miles distant, would be of reduced strength and be subject to the " capture " effect of the signal intended for the area.

(ii) Arrange for any two services which produce equivalent signals at any receiving point to be well separated in frequency, or if this is impossible, at least make their separation less than the 5 kc/s value*, which seems to give rise to the greatest interference.

(c) The saving in transmitter power which may be effected by the use of frequency modulation. An overwhelming point in favour of the frequency modulation transmitter is its high efficiency compared with the amplitude modulation

* Note.—This value of 5 kc/s was suggested as a result of measurements carried out in the U.S.A. by engineers of the N.B.C. It should not be considered as being an absolute constant.

equipment. In the case of an amplitude modulation transmitter operating at full 100% modulation, the amplitude of the carrier on modulation peaks has to be raised to twice the unmodulated level. This requires an increase in peak carrier power of four times that of the unmodulated carrier. With a frequency modulated wave, however, the carrier level is, or should be, unchanged by modulation, so that this power drain is not experienced. Considering R.M.S. values, the frequency modulation transmitter will consume only half the power from the supply than would its amplitude modulated counterpart. This can be a very worthwhile saving in the case of a broadcast transmitter.

It is hoped that the foregoing chapters have succeeded in giving a broad survey of present-day knowledge of the frequency modulation art. In an endeavour to keep the treatment simple, possibly the text has suffered somewhat, but all of the statements made in this book may be amplified by reference to the increasing number of technical papers and text-books on the subject.

BIBLIOGRAPHY

J. R. CARSON: "Notes on the Theory of Modulation," *Proc.I.R.E.* (February, 1922).

E. H. ARMSTRONG: "A Method of Reducing Disturbances by a System of Frequency Modulation," *Proc.I.R.E.* (May, 1936).

D. E. FOSTER and S. W. SEELEY: "Automatic Tuning, Simplified Circuits and Design Practice," *Proc.I.R.E.* (March, 1937)

K. R. STURLEY: "Frequency Modulation," *Hulton Press.*

M. G. CROSBY: "The Service Range of Frequency Modulation," *R.C.A. Review* (January, 1940).

GLOSSARY OF TERMS

Amplitude Limiter.—A device, usually a valve circuit, which tends to restrict variations in amplitude of an alternating current flowing into the circuit.

Carrier Wave.—A high-frequency radio wave which is employed to carry the intelligence it is required to transmit.

Corrector Network.—A circuit of resistance and capacity which converts phase variations into frequency variations.

Deviation.—The change measured from some standard reference level, such as the change in frequency about some mean value.

Discriminator.—A device which accepts frequency variations and converts them into amplitude changes.

Frequency, of a Wave.—The number of vibrations of the wave every second. When the variations are sufficiently slow to be registered by the human ear, the frequency is said to be an "audio-frequency."

Modulation.—A method of superimposing intelligence on a wave which would otherwise convey only energy between two points.

Phase, of an Alternating Current.—A factor which indicates the part of the cycle reached by the wave at the particular time considered, with respect to some other quantity.

Reactance Valve.—A valve which is incorporated in a circuit which produces a reactive element which varies in sympathy with amplitude variations fed into the grid of the valve.

Side-bands.—The bands of frequencies, often complex, which are formed when a carrier wave of single frequency is modulated.

INDEX

AN INTRODUCTION TO

FREQUENCY MODULATION

Details of information covered

Aerial
Amplitude Modulation
Amplitude Limiter
Amplitude of Wave
Audio Frequency
Balanced Modulator
"Capture" Effect
Carrier Wave
Channel Width
Corrector Network
De-emphasis
Detector
Deviation Ratio
Discriminator
Discriminator Foster-Seeley
Dual System Receiver
Electromagnetic Wave
Envelope Modulation
Fidelity of Reproduction
F.M. Receiver
Frequency Changer
Fourier Analysis
Generation of Noise
Ionised Layer
Interference
Intermediate Frequency
Local Oscillators
Modulation
Modulation Side-bands
Narrow-band F.M.
Phase Modulation
Pre-emphasis
Propagation of Radio Waves
Pulsatance
Radio-frequency
Reactance Valve
Resistive Load
Selectivity of Receiver
Side-bands
Side-bands, Single
Signal-to-noise Ratio
Spread of Signal
Superheterodyne Receiver
Tank Circuit
Tracking of Receiver Circuits
Transmission of Intelligence
Transmitter
Transmitter A.M.
Transmitter F.M.
Velocity of Radio Wave
Wavelength
Wide-band, F.M.
etc., etc.