CHAPTER XII.—RADIO-TELEPHONY

1. The production of continuous and interrupted continuous waves for radio-telegraphic purposes having been discussed in Chapter IX it is now proposed to consider the principles of radio-telephony. Radio-telephony is the transmission of speech, music or any other form of sound by means of electro-magnetic waves. As a means of communication radio-telegraphy and radio-telephony may be compared as follows :—

(i) Telephony possesses the advantages of verbal over written communication, i.e. saving of time, and rapid adjustment of minor points of detail. Arrangements are generally possible whereby the communication is carried out directly by the responsible officers, without the intervention of an operator, since knowledge of the morse code is not necessary. This is the most outstanding advantage from the service point of view.

(ii) Its disadvantages, compared with W/T are, first, that no written record of direct communication is available, and if such a record is required, the rapidity of communication is less than by telegraphy. Second, it is difficult to transmit by telephony messages in code, while the use of plain language involves risk of interception. Third, listening-through is not possible except by the use of complicated arrangements, which are impracticable for service use. Fourth, interference is more troublesome than with W/T.

SOUND

2. A brief reference to the phenomena connected with the process of hearing has already been made. (Chapter X.) Sound is invariably caused by the vibration of a material body, and the ensuing undulation of the medium in which the source of sound is situated constitutes a sound wave. In the absence of some such medium, no sound wave is produced, the usual method of demonstrating this being to enclose an electric buzzer in the bell-jar of an air-pump. When the bell-jar contains air at normal atmospheric pressure, the passage of sound waves is only slightly hindered by the presence of the jar, and the buzzer is distinctly heard. As the jar is evacuated, however, the sound becomes progressively fainter, and eventually becomes inaudible. This experiment shows that sound waves are of a nature different from light waves, for the bulb of an ordinary incandescent lamp (i.e. the " vacuum ", as opposed to the " gas-filled" type) is evacuated to the highest degree possible, but the radiation of light is unhindered by the absence of a material medium. So far as our normal experience is concerned, the medium in which sound waves are most often propagated is air, but any form of matter, solid, liquid or gaseous, will transmit sound, the motion of the particles in the medium being to and fro in the direction of propagation, in contrast to the corresponding motion in, say, a surface wave in water ; it is easily observed that in the latter instance the particle movement is in a plane perpendicular to the direction of propagation of the wave. Surface waves in water are an example of what are called transverse waves, while sound waves exemplify a type of undulation known as longitudinal wave motion.

Characteristics of sound

3. Sounds are conveniently divided into two classes, namely those which produce a pleasing effect upon the ear, and those which are unpleasant. The former are called musical sounds and the latter noises, but it is difficult to draw a definite line of demarcation. A sound wave possesses three characteristic properties by which it can be distinguished. These are :—

(i) Its wave form, also referred to as its quality or timbre.
(ii) Its intensity.
(iii) Its pitch.

The wave form in turn may be broadly divided into two classes, namely, repetitive and non-repetitive, which corresponds to the previous division into musical sounds and noises, and it may be said that a musical sound is that caused by a body whose vibrations possess a repetitive wave form, while noises are produced by bodies in a state of irregular vibration. In telephony, the reproduction of noises and musical sounds are of equal importance. In the service we are mainly concerned in the transmission of speech, which may be said to consist of an irregular succession of both musical and non-musical sounds, the former being the vowel sounds and the latter the consonants. Owing to their sustained character, vowels are generally easily recognizable in telephonic transmission, but the brief duration of the consonant sounds renders it necessary to articulate them with special care (though not with exaggerated emphasis). Noises of brief duration, including the sounds caused by percussion instruments, e.g. drums, and sounds having repetitive but heavily damped wave forms such as those produced by the piano and triangle are often classed together under the term " transients ". Complex wave forms consisting of a fundamental and a succession of harmonics have already been considered in Chapter V, but harmonics are of far greater importance in sound than in ordinary A.C. engineering, because the characteristic quality or timbre of any particular instrument depends upon its harmonic content. Thus the tuning fork and the flute emit waves which are almost truly sinusoidal, whereas the violin emits a wave which is particularly rich in harmonics. The wave form is partly governed by the properties of the sounding board upon which the vibrating body is mounted, e.g. the body of the violin, and certain German physicists contend that the timbre of a musical instrument is partly due to the emission of a heavily damped transient which occurs at the beginning of every cycle of the fundamental oscillation, and is termed a " tone-former ".

4. The intensity of a sound is a purely physical property independent of the ear, or other receiving device, and is proportional to the square of the amplitude of the wave. In practice the terms intensity and loudness are often regarded as synonymous. Loudness, however, is an indication of the degree of sensation produced in the brain of the hearer, and is difficult to define. The pitch of a sound depends upon the frequency of the disturbance, or in the case of a complex wave form upon the fundamental frequency. It has already been stated that the normal human ear will respond to frequencies ranging from about 16 to 20,000 cycles per second. It is useful to fix a mental standard of pitch, by comparison with the keyboard of a piano. Middle C (a white note immediately to the left of two black notes, a little to the left of the lock) has a frequency which for scientific purposes is taken as 256 cycles per second, but which in modern orchestras is given a higher pitch by increasing the tension on the strings; when tuned to concert pitch, middle C has a frequency of 261·65 cycles per second. The seventh white note above is called first upper C, or C^1, and has a frequency twice that of middle C, i.e. 512 on the scientific scale or 523·3 cycles per second in concert pitch, while first lower C, or C_1, has a frequency of 128 and 130·8 on the respective scales. The normal human voice ranges from 80 cycles per second (which is the lowest note usually reached by bass singers) to about 1,200 which is reached by some sopranos. These frequency numbers refer to the fundamental in each case, but harmonics are always present. For good reproduction of human speech in telephony it may be assumed that all frequencies between 200 and 2,000 cycles per second must be retained; frequencies outside this range contribute but little to the intelligibility, although they serve to distinguish the voice of one person from that of another.

The organ of hearing

5. The physiological process which we refer to as " hearing " is only imperfectly understood, but it is generally believed to be principally a phenomenon of resonance. A sound wave gathered by the pinna or external ear enters the ear passage, which is terminated by a membrane commonly though wrongly called the ear drum. The latter is really the cavity at the rear of the membrane. In contact with the membrane is the first of a train of small bones, the last of which communicates with an oval membrane at the other end of the cavity. The real process of hearing appears to commence at this point. The inner structure is very complex; the vibrations which are set up in the outer chamber are communicated to the inner structure, and in particular to the basilar membrane, which consists of a large number of tightly stretched strings like those of a harp, each of which is connected to the brain by a nerve. According to the resonance theory any vibration reaching the basilar membrane throws into vibration just that portion which is

tuned to the particular frequency, and impulses are transmitted through the nervous system to the brain, giving rise to a sensation of sound of the particular pitch, loudness and duration. It will be noted that the wave form is not conveyed to the brain, for the ear resolves a complex wave into its constituent vibrations. As a result of this analysis, the ear is unable to appreciate the relative phase of the harmonics contained in a complex wave form. For example, the waves shown in heavy line in fig. 1 produce the same aural impression, although their shape is entirely different. Both waves consist of a fundamental frequency and its third harmonic ; in fig. 1a the peak value of the harmonic always occurs at an instant when the fundamental is passing through its peak value but of opposite sign. Such a phase relationship gives rise to what is commonly referred to as a flat-topped wave. In fig. 1b, the peak value of the harmonic always occurs at an instant when the fundamental is passing through its peak value of the same sign, giving rise to what is called a peaky wave. The fact that the ear automatically resolves a complex wave form into its constituent frequency components is of considerable importance, in that a sound-reproducing system which does not transmit all frequencies in their correct relative phase may still appear to give distortionless reproduction.

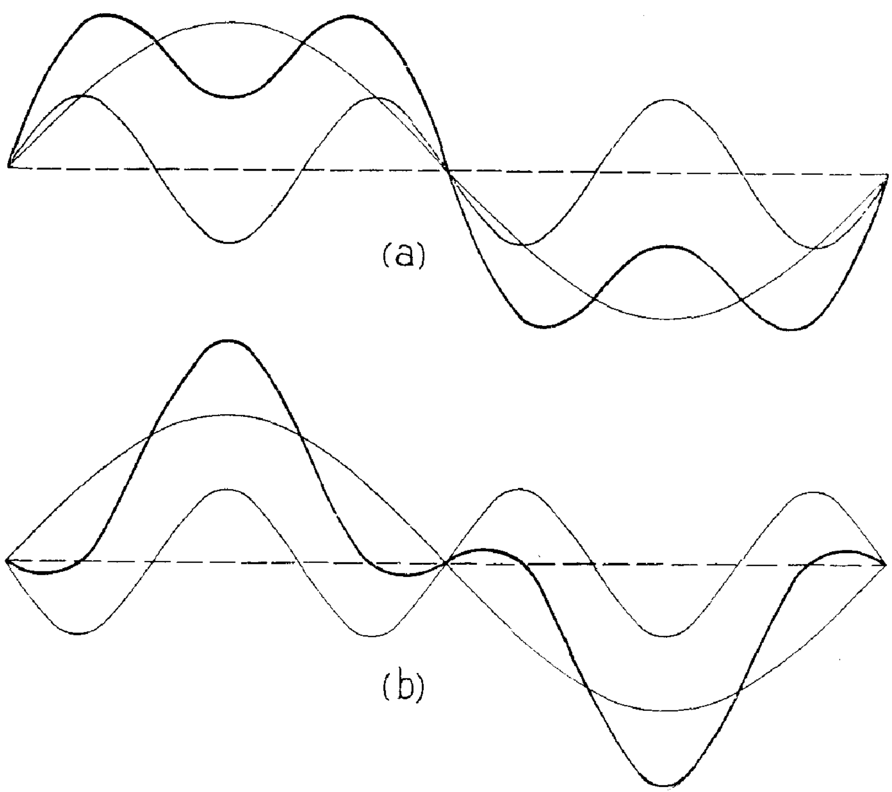


FIG. 1, CHAP. XII.—Waves with positive and negative third harmonic.

MICROPHONES

Pressure and velocity microphones

6. In telephony, the ultimate object is to transmit oral intelligence by electrical means, and the first stage of such a transmission is the conversion of sound waves into electrical impulses of some kind. Instruments designed to this end are called microphones, the original type being the carbon microphone which was briefly described in Chapter I. It will be remembered that this instrument operates by virtue of the change of resistance of the carbon granule pack under the influence of sound waves ; other methods of conversion may be utilized, and the forms of microphone which have been developed may be classified in several ways. Before dealing with these, it should be appreciated that in all forms of microphone, the sound wave is first caused to

set in vibration a light diaphragm, which in turn initiates some electrical action, which differs in the different types. Since sound waves are of the longitudinal type, the diaphragm itself may be set in motion in two different ways. First suppose the diaphragm to be mounted in such a manner as to form one side of a closed box, the remaining sides being very rigid compared to the diaphragm. A sound wave in air consists of a progressive variation of the pressure of the atmosphere (of the order of about 10 dynes per centimetre for ordinary speech, within a few inches from the mouth), and this variation of pressure will set up a corresponding vibration of the diaphragm, because its inner side is not exposed to the pressure variation owing to the rigidity of the other sides of the box. If, however, the same diaphragm is suspended in the air so that the variation of atmospheric pressure affects both sides simultaneously, it cannot be set in vibration by this means. It is thus apparent that when only one side of the diaphragm is exposed to the pressure variation, the diaphragm is set in vibration no matter in what position it is placed relative to the direction of propagation, for fluid pressure is transmitted equally in all directions. A microphone having a diaphragm which is actuated by the alternate compression and rarefaction of the air is called a pressure microphone. To a first approximation it may be said to be non-directional.

7. (i) Now consider the same diaphragm to be mounted in such a manner that its edge is fixed in space and both faces are exposed to the air; owing to its flexibility the centre portions of the diaphragm may still be set in vibration. Suppose a sound wave to be propagated in the vicinity of the diaphragm, then at any instant, although the air pressure may be above or below normal, both sides are affected simultaneously and to the same degree, hence the diaphragm is not set in vibration by the variation of pressure. There is, however, another phenomenon which must be taken into account. Since the vibration is longitudinal, the air particles are also in vibration to and fro along the line of propagation, and if the diaphragm is so placed that its face is perpendicular to the direction of propagation, it will be set in vibration by the actual movement of the air particles, while if its face is parallel to the direction of propagation no vibration will take place. A microphone in which the diaphragm is actuated in this manner is called a velocity microphone. Its characteristic is that it possesses more or less marked directional properties.

(ii) It is obviously desirable that whatever the mechanism of the microphone itself, the waveform of its ultimate electrical output should be a faithful copy of the waveform of the sound by which it is operated. In radio-telephony, the electrical output is usually required in the form of a voltage, current and power being of subsidiary importance. For our purpose then we may say that the ideal pressure microphone should possess such characteristics that the output voltage V_0 is directly proportional to the instantaneous pressure p due to the sound wave, i.e. $V_0 = Kp$, and correspondingly, in the velocity microphone the output voltage should be directionally proportional to the instantaneous velocity u of the air particles, i.e. $V_0 = Ku$. No type of microphone fulfils the above requirement over the whole range of frequency and amplitude, but certain types give a satisfactory approach to the ideal between certain limits. It is however unfortunate, from the service point of view, that the lines of attack appear invariably to lead to greater weight and linear dimensions, to a less robust instrument, and to reduced sensitivity compared with the carbon microphone. The simplicity and sensitivity of this instrument lead to its almost universal adoption for aircraft R/T installations.

Carbon microphones

8. (i) The carbon microphones in general use may be divided into two classes, namely those designed for use on the ground (such as the hand press and breastplate microphones) and those used in the air (mask microphones). The principles are the same in both classes, the objects of the mask microphone being first, to leave the hands free to perform other duties during the act of transmission, and second, to allow the wearer to assume any position these duties require, e.g. he may be required to use the microphone while in a prone position. The main features of the hand press microphone are shown in fig. 2. It consists of a hollow metal cylinder which is made in two parts in order to give access to the interior. The rear end of the cylinder carries ebonite terminal blocks, to which are attached the ends of the flexible cable by which the

microphone is connected in circuit. These terminal blocks carry german silver springs by which contact with the microphone capsule is maintained. The capsule in turn is of aluminium and carries at its wider end a mica diaphragm often called the sounding board. The microphone button is so mounted that one electrode is rigidly held by the rear end of the capsule but is insulated therefrom by fibre washers, while the other electrode, namely that mounted upon the mica diaphragm of the microphone button, is held on the centre of the sounding board by nuts and washers. This electrode is connected to the metal body of the capsule by means of a flat copper strip; in the diagram, however, this connection is shown as a short piece of insulated wire. It will be seen that the lower spring contact makes directly on a threaded extension of the rear electrode, while the upper spring contact is in metallic connection with the front electrode of the microphone button. A short trumpet-shaped extension of ebonite projects to the front of the casing and gathers the speech waves in such a manner that the actual variation of pressure

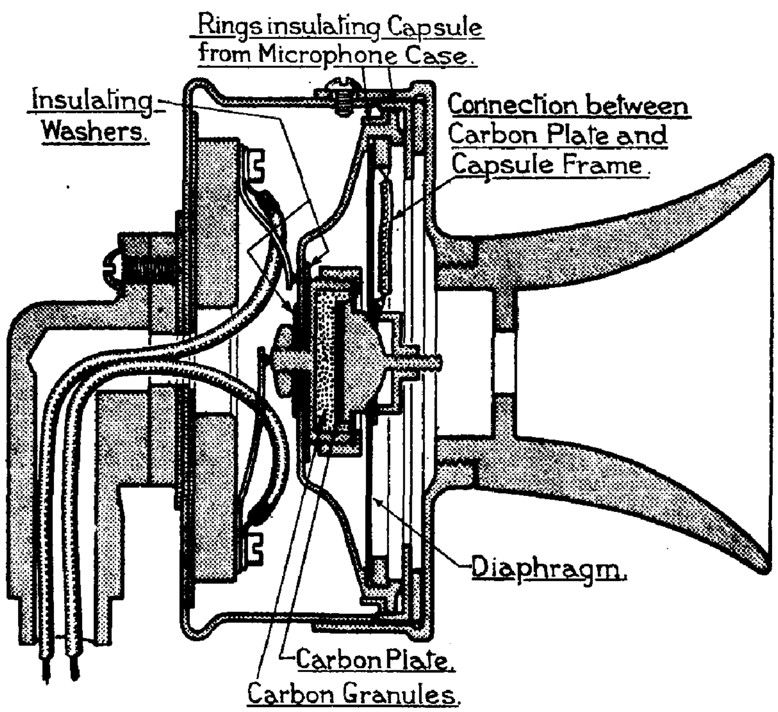


FIG. 2, CHAP. XII.—Head of hand press microphone.

upon the sounding board is greater than the variation of pressure at the open end of the mouth piece. The action of this mouthpiece, in conjunction with the space between its smaller end and the sounding board, is analogous to that of a step-up transformer.

(ii) In the mask microphone no capsule or sounding board is fitted and the button itself differs in design from that shown in fig. 2. Two mica diaphragms are fitted, each carrying a carbon electrode. One diaphragm is mounted on each side of a circular hole in an ebonite plate, the space so enclosed being partly filled with carbon granules. This type of microphone is fully described and illustrated in the appropriate chapter of A.P. 1186, Signal Manual Part IV. With regard to its action, it will be seen that an increase of pressure on both diaphragms simultaneously will lead to a reduction in the resistance of the carbon granule pack, while the reverse effect will be caused by a reduction of pressure. Thus a microphone constructed in this manner is of the pressure type.

9. The mean resistance of both the above forms of microphone button is generally of the order of 50 ohms, but depends to some extent upon the current flowing. It must be borne in mind that if the feed current (i.e. the mean steady current) exceeds about 150 milliamperes,

arcing will occur between the sharp edges of adjacent granules and they tend to become welded together ; the microphone is then said to be caked. This condition is also sometimes caused by the penetration of moisture into the interior of the button. In any event, it is advisable to give the casing of the microphone a sharp tap with the finger before commencing to speak, in order to shake up the granules and so restore the instrument to its most sensitive state. The use of feed currents higher than those recommended for any particular type will soon cause serious deterioration, because the effect of arcing between the granules is to remove their sharp edges, upon which the sensitivity of the microphone depends.

Audio-frequency voltage generated by carbon microphone

10. (i) The action of a carbon microphone depends upon its variation of resistance, and the latter in turn upon the change of pressure upon the diaphragm due to the impressed sound wave. It is desirable to possess some idea of the magnitude of the effect, although an exact numerical treatment is very difficult. It may, however, be assumed that the diaphragm and granules normally exert a slight mutual pressure P, owing to the weight of the granules and the stiffness of the diaphragm. Let the normal resistance of the microphone be R_m ohms. An increase of pressure then causes a decrease of resistance and vice-versa, so that the following law may be expected to hold, viz. :—

$$R_m = \frac{K}{P}$$

Now consider a varying pressure $p = P_o \sin \omega t$ to be applied to the diaphragm. The resistance will then become $R_m + r_m$ where

$$R_m + r_m = \frac{K}{P + p} = \frac{K}{P\left(1 + \frac{p}{P}\right)}$$

Now

$$\frac{1}{1 + \frac{p}{P}} = 1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2 - \left(\frac{p}{P}\right)^3 \dots\dots\dots,$$

and if the maximum value of p, i.e. P_o, is small compared with the normal pressure P, $\left(\frac{p}{P}\right)^2$ will be much smaller than unity, while higher powers will be absolutely negligible, and

$$R_m + r_m = \frac{K}{P}\left\{1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2\right\}$$

$$= R_m \left\{1 - \frac{p}{P} + \left(\frac{p}{P}\right)^2\right\}$$

or

$$r_m = R_m\left\{- \frac{p}{P} + \left(\frac{p}{P}\right)^2\right\}.$$

This then is the magnitude of the variation in resistance of the microphone ; experiment shows that the peak pressure upon the diaphragm during ordinary speech is of the order of 30 dynes per square centimetre, while the effective back pressure P of the diaphragm in a certain design was found to be of the order of 500 dynes per square centimetre. Suppose R_m to be 50 ohms, then

$$r_m = 50\left\{- \frac{30}{500} + \left(\frac{30}{500}\right)^2\right\}$$

$$= - 3 \text{ ohms (nearly)}$$

The feed current to the microphone in an ordinary series circuit will be

$$I_m = \frac{E}{R + R_m + r_m}$$

where R is the D.C. resistance of the external circuit and E is the applied E.M.F.. As r_m is quite small compared to $R + R_m$, however, I_m may be regarded as being constant and equal to $\frac{E}{R + R_m}$. The P.D. between the terminals of the microphone may be considered to consist of two components, namely a constant voltage $V_m = \frac{R_m \; E}{R + R_m}$, with which we are not immediately concerned, and a varying component $v_m = r_m \; I_m$. Then since $p = P_o \; sin \; \omega t$, where $\frac{\omega}{2\pi}$ is the frequency of the sound wave acting on the diaphragm,

$$v_m = I_m \; R_m \left\{ - \frac{P_o \; sin \; \omega t}{P} + \left(\frac{P_o \; sin \; \omega t}{P} \right)^2 \right\}$$

For example, if $I_m = 150$ milliamperes, $R_m = 50$ ohms, $P_o = 30$ dynes per square centimetre, $P = 500$ dynes per square centimetre, we have

$$\begin{aligned} v_m &= \cdot 15 \times 50 \; (- \; 0 \cdot 06 \; sin \; \omega t + \cdot 0036 \; sin^2 \; \omega t) \\ &= - \; 0 \cdot 45 \; sin \; \omega t + \cdot 027 \; sin^2 \; \omega t. \end{aligned}$$

It will be observed that in obtaining a value for r_m the term $\left(\frac{p}{P}\right)^2$ was ignored, but that it has been retained in deriving an approximate numerical value for the effective voltage v_m. The object of keeping it in the second instance is to show the relative amplitude of the second harmonic of the speech frequency. As $sin^2 \; \omega t = \frac{1}{2} \; (1 - cos \; 2 \; \omega t)$, we see that the microphone also generates a second harmonic having an amplitude, in general, equal to $\frac{P_o}{2P}$ times that of the fundamental; in the particular case under consideration, the amplitude will be $\frac{\cdot 027}{2} = \cdot 0135$ volts. Harmonics of higher order will also be present, but provided that P_o is small compared to P, their amplitude will be inconsiderable.

(ii) Summarizing the above, then, it may be said that under the influence of a sound wave the carbon microphone may be considered to act as an alternating current generator, its E.M.F. waveform having a fundamental and a series of harmonics for each frequency contained in the original speech wave; unless the pressure on the diaphragm due to the sound wave is of the same order as the normal pressure, the harmonics introduced by the microphone itself will be negligible. A more serious consequence of its non-linearity, particularly when used in noisy surroundings, is the phenomenon called intermodulation, which is discussed in paragraph 41.

The condenser microphone

11. (i) This consists of an air dielectric condenser, the capacitance of which is varied by the action of sound waves upon one electrode, which is made in the form of a very thin disc, usually of duralumin. The other electrode is a rigid brass disc, and the thickness of the dielectric is of the order of $\cdot 002$ inch. Thus, if the effective diameter of the electrodes is $1 \cdot 5$ inches, the normal capacitance will be about $\cdot 0002 \; \mu F$. When sound waves impinge upon the diaphragm, its movement results in a variation of the thickness of the dielectric and a corresponding change in capacitance. The manner in which this change is caused to develop a varying voltage is shown in fig. 3. Here E is an E.M.F. set up by a battery of some 200 volts, L an audio-frequency choke of several hundred henries, which however, may be replaced by a large non-inductive resistance in some instances, and C the microphone. Under normal conditions the P.D. across the microphone terminals is equal to the E.M.F. of the battery, i.e. E volts.

Let the normal capacitance of the microphone be $C_o = \frac{K}{D_o}$ where K is a constant and D_o the thickness of the dielectric. When the polarizing voltage E is applied to the condenser, the condenser will receive a charge Q, equal to C_oE. Provided the inductance is sufficiently large, this charge will remain sensibly constant even if the capacitance is varied, and as a result we have a varying P.D. at the microphone terminals. Thus if $d = D_o + D \sin \omega t$

$$C = \frac{K}{d} = \frac{K}{D_o\left(1 + \frac{D}{D_o} \sin \omega t\right)} = \frac{C_o}{1 + \frac{D}{D_o} \sin \omega t}.$$

Also $Q = Cv = C_oE$ where v is the P.D. at the microphone terminals.

$$v = \frac{C_o}{C} E$$

$$= C_o\, E \frac{\left(1 + \frac{D}{D_o} \sin \omega t\right)}{C_o}$$

$$= E\left(1 + \frac{D}{D_o} \sin \omega t\right).$$

The variation of voltage at the microphone terminals is thus of the same form as the sound wave, if the required conditions are satisfied. Actually of course the inductance cannot entirely

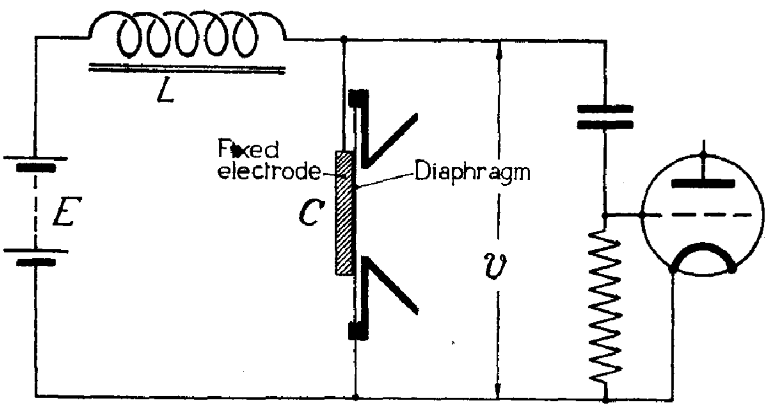


FIG. 3, CHAP. XII.—Condenser microphone.

prevent a variation of charge, and the variation of capacitance will set up a current variation. The result is that v will vary in a more complex manner than the original sound wave, but the additional frequencies introduced will not be of large amplitude, because they are proportional to the second and higher powers of $\frac{D}{D_o}$ and this ratio is always much less than unity.

(ii) In the practical form of condenser microphone the case containing the electrodes is hermetically sealed in order to prevent the entry of moisture and dust, arrangements being made to allow the normal pressure on each side of the diaphragm to be equalized. The microphone possesses a resonant frequency which is usually of the order of from 3 to 5 kc/s, but its response curve is fairly flat, e.g. taking the voltage response at 800 cycles per second as a standard, a typical instrument may give a response of -5 d.b. at 300 cycles per second, and $+10$ d.b. at its resonant frequency, say 4,000 cycles per second. Above this the response falls rapidly, being perhaps -5 d.b. at 8,000 cycles per second. The instrument obviously has a high impedance at speech frequencies and should work into a very high resistance load, such as the input impedance of a carefully designed audio-frequency amplifier. In any event, since the condenser microphone is of very much lower sensitivity than the carbon microphone, at least two extra stages of audio-frequency amplification are required in order to obtain the same output. It is usual to mount the first stage as near to the microphone itself as possible to avoid the shunting effect of the capacitance of a long twin lead or concentric cable. The comparative insensitivity of the condenser microphone outweighs the advantage of its fairly uniform frequency response, except in circumstances where the space and weight of the ancillary apparatus is of no importance.

The moving coil microphone

12. In this instrument the movement of a light coil of wire in a powerful magnetic field produces an E.M.F. conforming in variation with the sound waves which cause the movement of the coil. The magnetic circuit is so arranged that the only air gap is an annular space two to four centimetres in diameter and about ·2 centimetres wide. As shown in fig. 4, the moving coil is wound on a light former, which is carried by a suitable diaphragm in such a manner that the movement of the latter causes a variation of flux though the coil, and the production of a corresponding E.M.F. A fairly uniform response over a wide frequency range is achieved by careful design of the air chamber enclosed by the magnetic system and diaphragm ; taking the

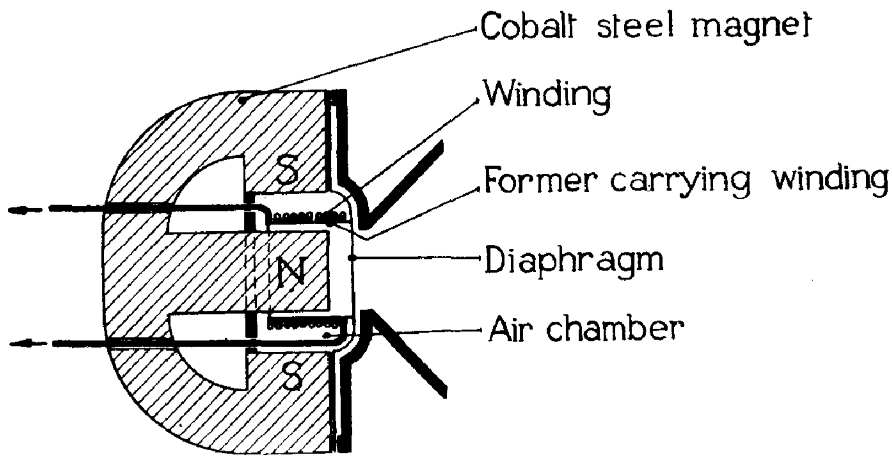


FIG. 4, CHAP. XII.—Moving coil microphone.

response at 800 cycles per second as standard the response of a typical instrument may be —3 d.b. at 50 cycles per second, + 5 d.b. in the neighbourhood of the resonant frequency (about 3,000 cycles per second) and falling rapidly above 10,000 cycles per second. One or more subsidiary resonances may be observed between 3,000 and 10,000 cycles per second. The diaphragm is actuated partly by the change of pressure and partly by the air particle velocity and consequently the instrument is somewhat directional particularly at the higher frequencies, maximum response being obtained when the sound wave impinges perpendicularly upon the diaphragm. The instrument has a low impedance and may be transformer-coupled to the first valve of a speech amplifier. It is rather more sensitive than the condenser microphone, but not nearly so sensitive as the carbon microphone. Both the condenser and moving coil microphones have the advantage of a much lower noise level compared with the carbon microphone, but unless the additional amplifier stages are carefully designed and operated these advantages may be offset by the additional amplifier noise.

MODULATION

Necessity for carrier wave

13. (i) The microphone was originally intended for use in telephonic communication between places connected by wire, and the audio-frequency variations of current resulting from its operation were conveyed from point to point in their original form. It is not possible, however, to radiate power at audio-frequencies over any appreciable distance, for the energy radiated in the form of electro-magnetic waves by an open oscillator is proportional to the square of the

frequency, so that radio communication is only possible by means of frequencies of at least 20,000 cycles per second, while much higher frequencies are to be preferred. It follows then that radio-telephonic communication cannot be accomplished by the use of electrical oscillations having frequencies identical with those of the sound waves to be transmitted. In line telephony, the variations of current caused by the action of sound waves upon the microphone are superimposed upon a steady direct current, which is referred to by telephone engineers as the carrier current. Similarly in R/T practice it is necessary to superimpose the audio-frequency electrical variations corresponding to the sound waves upon a continuous or carrier oscillation of radio frequency. The resulting radiation consists of a continuous or carrier wave, the character of which is varied at audio-frequency; such a wave is said to be modulated. In the course of a complete communication by radio-telephony there are five distinct processes, which are as follows :—

In transmission :

(*a*) Conversion of sound vibrations in the air into corresponding audio-frequency variations of current or voltage (or both). This is effected by the microphone.

(*b*) Modulation, or superimposition of these audio-frequency electrical variations upon a radio-frequency carrier wave.

(*c*) Radiation and propagation, which are identical with the corresponding features of C.W. transmission.

In reception :

(*d*) Detection. This is the process of resolving the modulated currents received by the aerial into radio-frequency and audio-frequency components; it corresponds to rectification in the case of C.W. or I.C.W. reception.

(*e*) Reproduction, which is the process of converting the audio-frequency electrical variations obtained by the process of detection into sound waves in the air.

(ii) In the above processes, except of course the third, it may be necessary to employ suitable methods of amplification, while the carrier wave must be generated by means of a transmitting circuit similar to that used for the production of C.W. It must be particularly noted that after the second process referred to above, the audio-frequency variations of current and voltage cease to have a separate existence. The modulated wave is of radio frequency, and until stage (*d*) of the process is reached, all the circuits possess constants appropriate to the radio frequency in question. After the detection process has been performed, however, audio-frequency currents and voltages once more appear and the circuit constants possess appropriate values.

Modulated oscillations

14. (i) Three forms of modulation are possible; these are referred to as amplitude modulation, frequency modulation and phase modulation respectively. In current practice the first-named is almost universally adopted, and when the word modulation is used without qualification, amplitude modulation is to be understood. Other forms of modulation will be dealt with briefly in due course.

(ii) The simplest form of amplitude-modulated oscillation is one in which the amplitude varies in a sinusoidal manner, as shown in fig. 5, which may be considered to represent the aerial current of an R/T transmitter. The portion OA represents the aerial current in the absence of modulation, i.e. the carrier oscillation. The amplitude of the radiated electro-magnetic wave is proportional to the aerial current, and therefore modulation of the aerial current will result in the radiation of a modulated electro-magnetic wave. When no modulation is taking place, the aerial current may be represented by the equation

$$i_c = \mathcal{I}_c \sin \omega_c t$$

where $\omega_c = 2\pi f_c$, f_c being the frequency of the unmodulated oscillation. During modulation the amplitude of the aerial current varies sinusoidally between the values $\mathcal{I}_c + \mathcal{I}_a$ and $\mathcal{I}_c - \mathcal{I}_a$

and a complete cycle of this variation occurs in a certain time, say T_a seconds. The frequency of this variation is $\frac{1}{T_a}$ cycles per second and may be denoted by f_a, which is the audio-frequency at which modulation is taking place. The amplitude at any instant, say t seconds from the instant A at which modulation commences, is $\mathcal{I}_c + \mathcal{I}_a \sin 2\pi f_a t$, or if $\omega_a = 2\pi f_a$, the amplitude at this instant is $\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t$. Although the amplitude of the current is no longer constant, its frequency is still f_c, and the modulated wave form may be completely represented by the equation

$$i_m = (\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t) \sin \omega_c t \quad \ldots \quad (1)$$

The ratio $\frac{\mathcal{I}_a}{\mathcal{I}_c}$ is called the depth of modulation, or, if multiplied by 100, the percentage of modulation. In the diagram below (fig. 5) $\frac{\mathcal{I}_a}{\mathcal{I}_c}$ is equal to 0·5 and the oscillation is said to be modulated 50 per cent. It is usual to denote the depth of modulation by the symbol K. Hence equation (1) may be written

$$i_m = (1 + K \sin \omega_a t)\, \mathcal{I}_c \sin \omega_c t \quad \ldots \quad (2)$$

(iii) Instead of regarding the modulated waveform as having a single frequency but varying amplitude, it is convenient to regard it as the sum of a number of radio-frequency oscillations

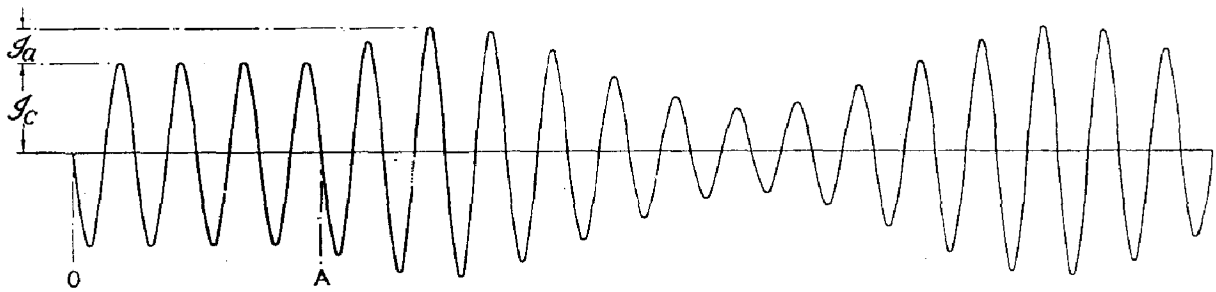


FIG. 5, CHAP. XII.—Amplitude modulated wave.

of various frequencies, each of which is of constant amplitude. There is no subterfuge in such a transformation ; it has been stated repeatedly that no matter how complex a wave-form may be, provided only that it is repetitive, it may be proved to be the sum of a number of sinusoidal components. In the present instance, we may make use of an expression developed in Chapter V, namely

$$2 \sin P \sin Q \equiv \cos (P - Q) - \cos (P + Q)$$

The sign $\equiv$ has been used in order to emphasise that this is not an equation but an identity, that is the two members are always equal no matter what values are allotted to P and Q. The equation representing the modulated oscillation, viz. $i_m = (\mathcal{I}_c + \mathcal{I}_a \sin \omega_a t) \sin \omega_c t$ may be written

$$i_m = \mathcal{I}_c \sin \omega_c t + \mathcal{I}_a \sin \omega_c t \sin \omega_a t$$

By the above identity, therefore,

$$i_m = \mathcal{I}_c \sin \omega_c t + \frac{\mathcal{I}_a}{2} \cos (\omega_c - \omega_a)\, t - \frac{\mathcal{I}_a}{2} \cos (\omega_c + \omega_a)\, t. \quad \ldots \quad (3)$$

The modulated oscillation is thus shown to be the sum of three component oscillations, each of constant amplitude, namely :—

(*a*) The original carrier current, of amplitude $\mathcal{I}_c$ and frequency $f_c = \frac{\omega_c}{2\pi}$.

(*b*) An oscillation of constant amplitude $\frac{\mathscr{I}_a}{2}$, the frequency of which is less than that of the carrier by an amount equal to f_a, the frequency of the modulation. This is called the lower side-frequency.

(*c*) An oscillation of constant amplitude $\frac{\mathscr{I}_a}{2}$, the frequency of which is greater than that of the carrier by an amount equal to f_a. This is called the upper side-frequency.

It will be noted that the modulation frequency f_a is in the audio-frequency range, while f_r is a radio-frequency; the upper and lower side-frequencies being $(f_r + f_a)$ and $(f_r - f_a)$ respectively are both radio-frequencies.

15. In order to prove that the sum of the above three component oscillations gives the modulated oscillation of fig. 5, fig. 6 has been prepared. The two oscillations $\frac{\mathscr{I}_a}{2} \cos (\omega_c + \omega_a) t$ and $\frac{\mathscr{I}_a}{2} \cos (\omega_c - \omega_a) t$ are shown at (*a*), the former oscillation commencing 180° out of phase with the latter in order to comply with the negative sign prefixed to it in equation (3) above.

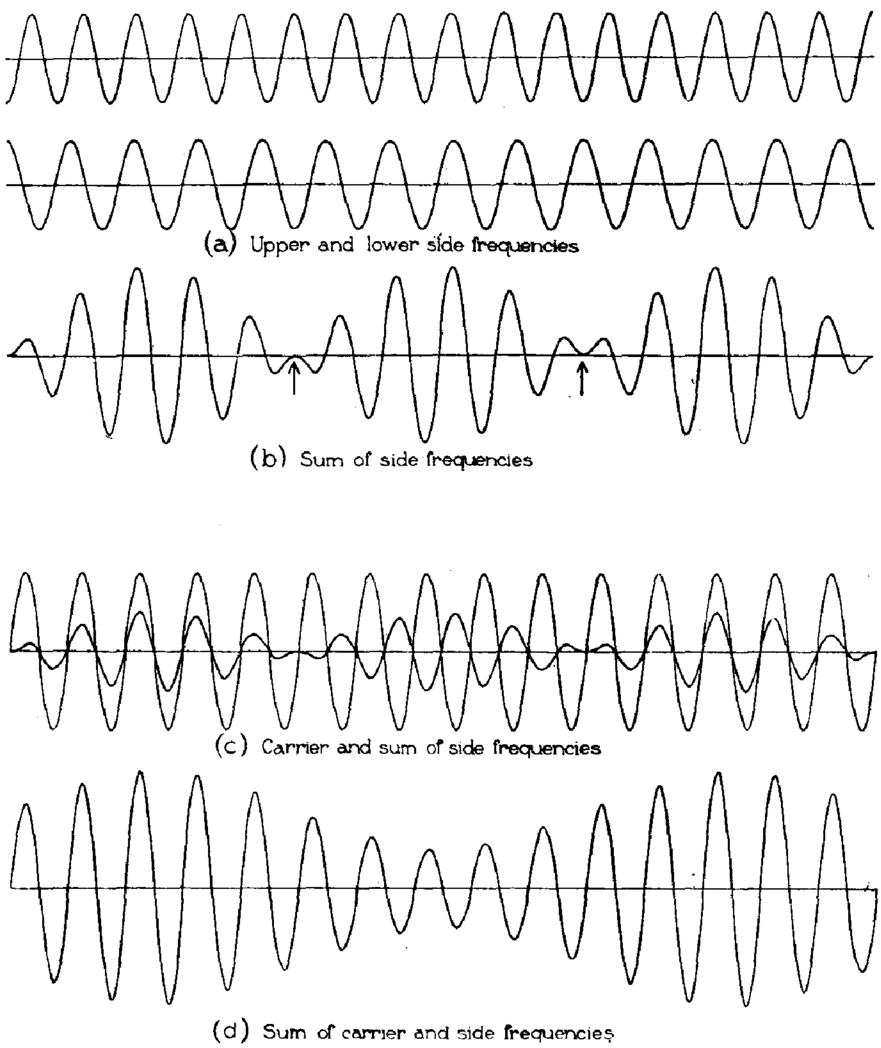


FIG. 6, CHAP. XII.—**Synthesis of modulated wave.**

The sum of these two is shown at (*b*). This result is identical with that obtained in heterodyne reception when the incoming and local oscillations are of equal amplitude. The frequency of the individual waves forming each " beat " is the mean of the two component frequencies, namely $\frac{1}{2\pi}\left\{\frac{(\omega_c + \omega_a) + (\omega_c - \omega_a)}{2}\right\} = \frac{\omega_c}{2\pi} = 2f_c$, i.e. it is equal to that of the carrier. Again, the number of beats per second is equal to the difference between the two frequencies, being equal to $\frac{1}{2\pi}\left\{(\omega_c + \omega_a) - (\omega_c - \omega_a)\right\} = \frac{2\omega_a}{2\pi} = 2f_a$, hence there are two complete beats in the time occupied by one audio-frequency cycle. A point of great importance in the theory of modulation, although of no significance in heterodyne reception, is the change of phase which occurs at the instant at which one beat is concluded and the next is commenced. These instants are indicated by arrows in the figure.

16. In fig. 6c, the beats formed as described above are superimposed upon the carrier oscillation in order to show the effect of this change of phase. During the time taken by the first beat, the carrier and beat oscillations are in phase, and the amplitude

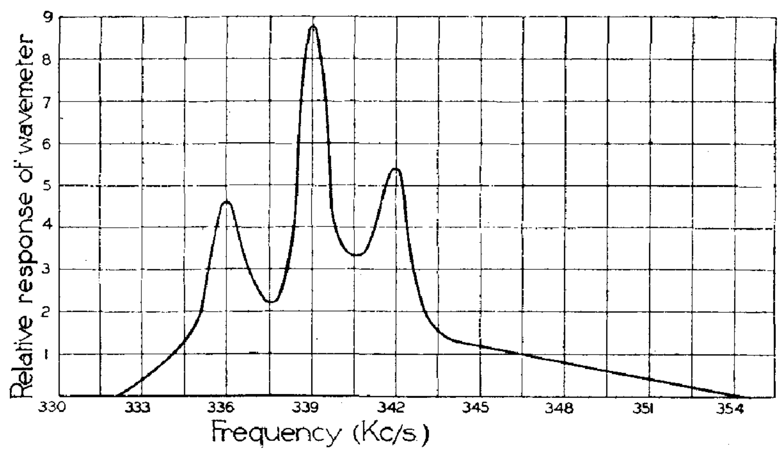


FIG. 7, CHAP. XII.—Response of wavemeter to carrier and side frequencies.

of the resultant, which is initially ϑ_c, gradually rises to a value $\vartheta_c + \vartheta_a$, and then falls again, reaching its original value ϑ_c at the end of the first beat. Owing to the change of phase at this instant, during the next beat the carrier and beat oscillations are 180° out of phase with each other, and the amplitude of their sum falls to $\vartheta_c - \vartheta_a$ at the middle of the beat. The amplitude then increases, and attains the value ϑ_c when the end of the beat is reached, the foregoing cycle of events being then repeated. The sum of the three component oscillations is shown in fig. 6d and it is seen to be identical with the original modulated oscillation in fig.5. When an aerial is carrying an oscillatory current, sinusoidally modulated by a fairly high audio frequency, the carrier oscillation and the upper and lower side-frequencies can be separately identified by a wavemeter, provided that its oscillatory circuit has a high magnification. Fig. 7 shows the response of such a wavemeter in a particular instance, the carrier frequency being 339 kc/s and the modulating audio frequency 3 kc/s. It is seen that, in accordance with the foregoing analysis, the side-frequencies are respectively 3 kc/s above and below the carrier frequency. This experiment was actually performed to demonstrate the objective existence of the side-frequencies. The fact that the wavemeter response at the upper side-frequency is somewhat higher than that at the lower does not mean that the amplitudes of these components are unequal, but merely signifies that the magnification of the wavemeter—and probably that of the aerial circuit also—is slightly greater at 342 kc/s than at 336 kc/s.

17. We have now seen that the modulation of a carrier by a single audio frequency results in the production of two radio-frequency oscillations in addition to the original carrier. Suppose now that the carrier is simultaneously modulated by two audio frequencies. Each of these will give rise to an upper side-frequency and a lower side-frequency and, in general, every audio-frequency oscillation by which the carrier is modulated gives rise to a pair of side-frequencies, those higher than the carrier frequency being collectively referred to as the upper side-band, and those lower than the carrier frequency being called the lower side-band. For example, if the carrier frequency is 100 kc/s and it is simultaneously modulated by frequencies of 250, 480 and 1,200 cycles per second, the modulated oscillation would contain all the following components :—

(i) The carrier frequency	1,000,000 cycles per sec.
(ii) The upper side-band	$1{,}000{,}000 + 250 = 1{,}000{,}250$
	$1{,}000{,}000 + 480 = 1{,}000{,}480$
	$1{,}000{,}000 + 1{,}200 = 1{,}001{,}200$
(iii) The lower side-band	$1{,}000{,}000 - 250 = 999{,}750$
	$1{,}000{,}000 - 480 = 999{,}520$
	$1{,}000{,}000 - 1{,}200 = 998{,}800$

The complex vibrations of speech and music can be resolved into a number of sinusoidal variations, and when a carrier oscillation is modulated in accordance with the frequencies of these sound waves, the side-bands consist of a pair of radio-frequency oscillations for each constituent (sinusoidal) audio-frequency component of the sound wave. When a carrier oscillation is subjected to modulation by several audio-frequencies simultaneously, some care is needed in speaking of the depth or percentage of modulation, in fact these terms can only be used with accuracy when dealing with sinusoidal modulation by a single frequency. The point is best illustrated by actual examples, such as are shown in figs. 8, 9 and 10, to which further reference will be made.

Power in carrier and side-bands

18. The total energy contained in a modulated wave is the sum of the energies carried by the components of different frequencies, so that if, for example, we have an aerial circuit whose radiation resistance is R ohms, and I_c is the R.M.S. carrier current, the power radiated by the carrier alone is I_c^2R watts. Now suppose the wave to be modulated to a depth of unity, the R.M.S. value of each side component will be $\frac{I_c}{2}$, and the total power radiated by the side-frequencies will be $\left(\frac{I_c}{2}\right)^2 R \times 2 = \frac{I_c^2R}{2}$. The total power radiated in the completely modulated wave is therefore $(1 + \frac{1}{2})\ I_c^2R = \frac{3}{2}\ I_c^2R$, of which only one-third is carried by the side-frequencies and two-thirds by the carrier frequency. Now instead of a completely modulated wave, for which $K = 1$, let us consider one modulated to a depth of less than unity. The carrier will radiate I_c^2R watts, as before, and each side-frequency, having an R.M.S. value $\frac{K\ I_c}{2}$, will radiate $\left(\frac{K\ I_c}{2}\right)^2 R$ watts, the total side-band power being therefore $\frac{K^2\ I_c^2\ R}{2}$ watts. The total power contained in the wave is now equal to $I_c^2R\left(1 + \frac{K^2}{2}\right)$ so that the power in the side frequencies is $\frac{K^2}{2 + K^2}$ of the total power. Since the required intelligence is conveyed entirely by the side-frequencies, the effective signalling range is rapidly reduced by a reduction in the depth of modulation. From this point of view it would appear desirable to modulate to a depth of unity, but in practice it is usual to aim at not more than about 80 per cent. This is partly on

account of certain difficulties which arise in reception, partly because provision must be made for a considerable variation in the intensity of the modulating sound, but chiefly because in practice the latter is rarely of sinusoidal wave-form. For distortionless transmission, the envelope of the radio-frequency oscillation must be identical with the wave-form of the audio-frequency variation. This will not be so, even with sinusoidal modulation, if the variation of amplitude exceeds the amplitude of the carrier. Thus in fig. 8a, the variation of amplitude is equal to $1 \cdot 5$ $\mathcal{I}_c$; the envelope of the resulting radio-frequency oscillation is far from sinusoidal, and after detection would be found to contain a pronounced second harmonic. In practice however the wave-form shown in fig. 8a cannot be produced by ordinary methods of modulation. Under the conditions indicated, the usual result is a total cessation of oscillation during the "overlap" period, as in fig. 8b. Such a wave is said to be over-modulated.

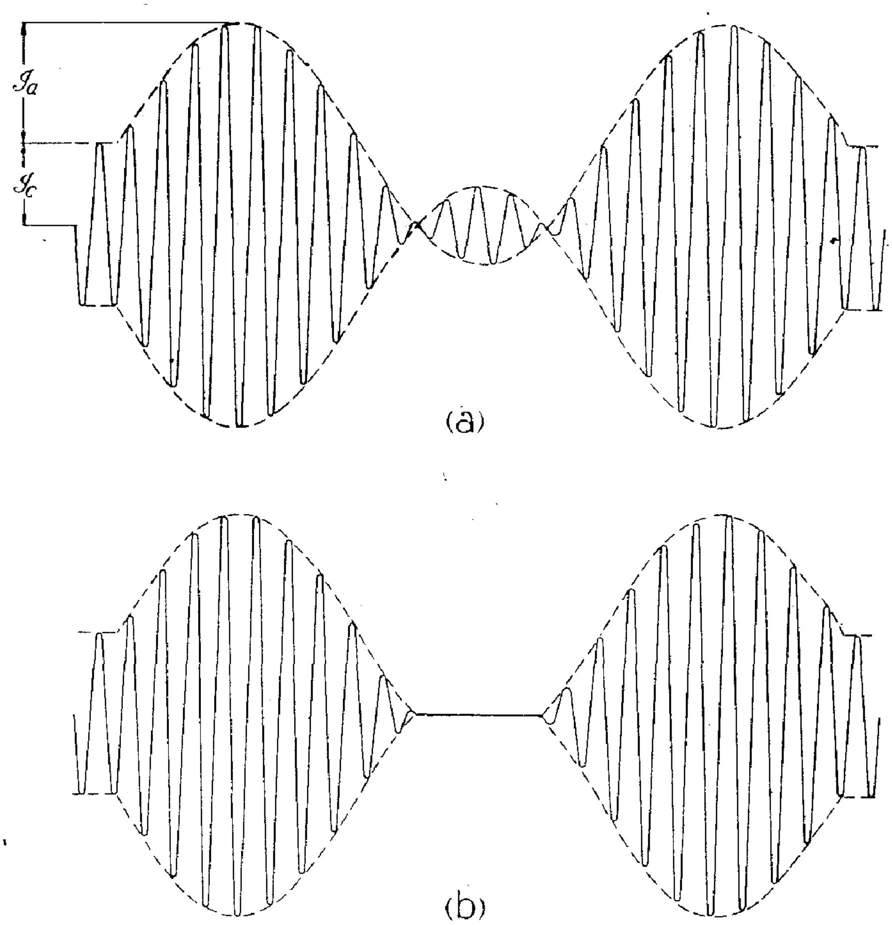


FIG. 8, CHAP. XII.—Over-modulation.

19. Assuming that a wave is completely modulated, more energy is radiated in the side-bands of a flat-topped wave than in one of peaky form. Figs. 9 and 10 show the components and resultant modulated wave when modulation is caused by a fundamental audio frequency and its third harmonic, the relative phase in fig. 9 corresponding to fig. 1a, and in fig. 10 to that of fig. 1b. Taking fig. 9 first, the carrier wave has an amplitude of $\mathcal{I}_c$, say, and the fundamental side frequency $\mathcal{I}_1$ is equal to $\frac{\mathcal{I}_c}{2}$. It can be shown that a third harmonic having an amplitude $\mathcal{I}_3 = \frac{2}{5}\mathcal{I}_1 = \frac{1}{5}\mathcal{I}_c$ can be added without over-modulation. Then if the radiation resistance of

the aerial is R as before, the power in the carrier is $\frac{\mathcal{I}_c^2 R}{2}$, in the fundamental side-frequencies $2\left(\frac{\mathcal{I}_c^2 R}{2 \times 2^2}\right)$, and in the third harmonic side-frequencies $2\left(\frac{\mathcal{I}_c^2 R}{2 \times 5^2}\right)$. Thus the power in the side-bands is to the power in the carrier as $2\left(\frac{1}{4}+\frac{1}{25}\right)$ is to 1 or $\cdot 58$ to 1. This may be compared with $\cdot 33$ to 1 for sinusoidal modulation (paragraph 18). If, however, the phase of the third harmonic is reversed, giving rise to the peaky wave form shown in fig. 10, and the third harmonic is still to have an amplitude equal to $\frac{2}{5}\,\mathcal{I}_1$, the amplitude of the fundamental itself must be less than $\frac{\mathcal{I}_c}{2}$, otherwise over modulation will occur. A simple calculation shows that for complete modulation, $\mathcal{I}_1 = \frac{5}{14}\mathcal{I}_c$ and $\mathcal{I}_3 = \frac{2}{5}\,\mathcal{I}_1 = \frac{1}{7}\,\mathcal{I}_c$, so that the power in the side-bands is to the power in the carrier as $2\left\{\left(\frac{5}{14}\right)^2 + \left(\frac{2}{14}\right)^2\right\} : 1$ or $\cdot 296$ to 1. From these two examples it may be deduced that a perfectly square-topped wave would have the greatest ratio of side-band to carrier power; it can be shown that if such a wave is completely modulated the side-bands contain the same amount of power as the carrier.

Increase of aerial current during modulation

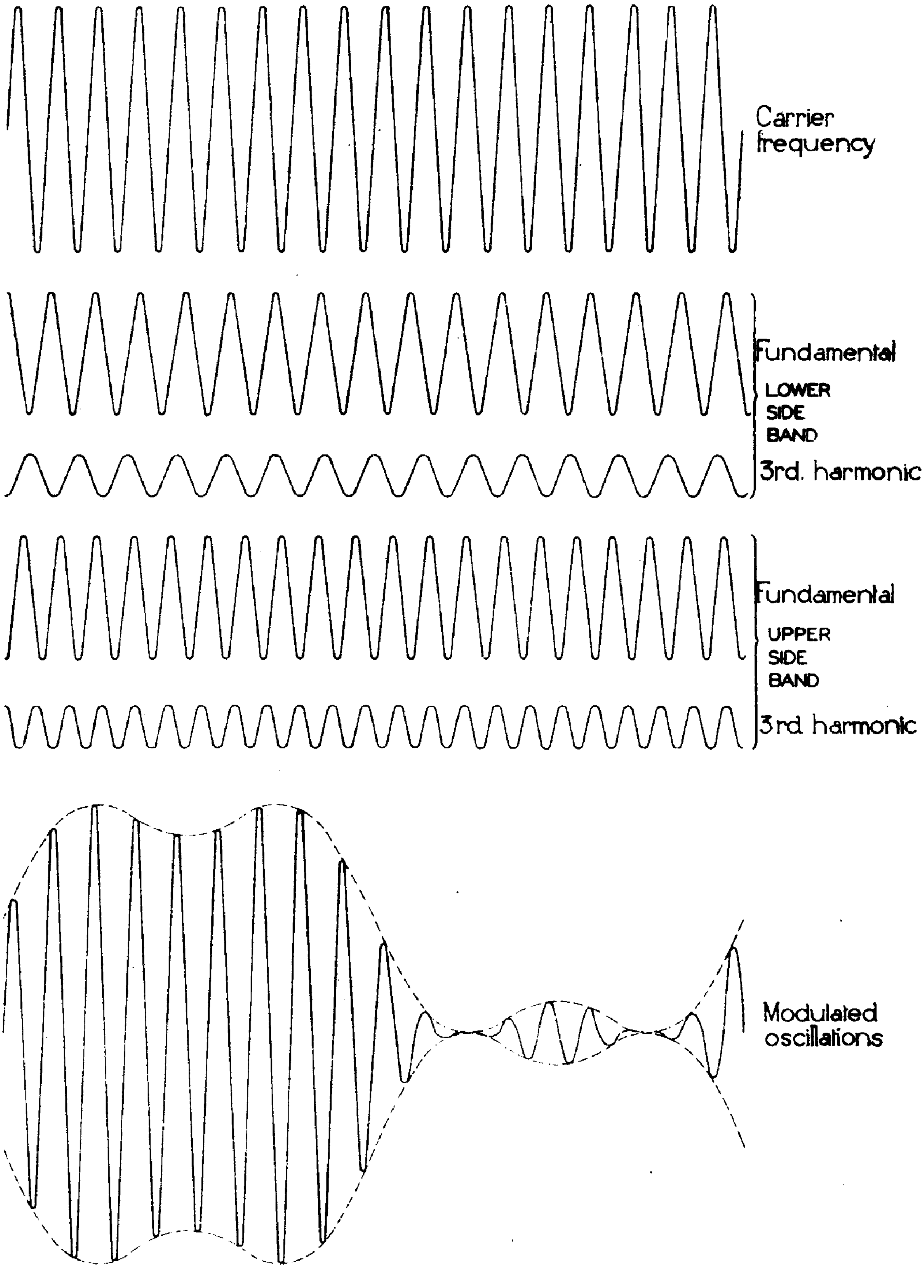
20. Since during modulation the power in the aerial circuit is continually varying, the aerial current must also vary. In paragraph 18 it was shown that the power in the aerial circuit is proportional to $I_c^2\left(1+\frac{K^2}{2}\right)$ where I_c is the unmodulated R.M.S. aerial current. It follows therefore that during modulation the R.M.S. aerial current will rise to some value I_M where

$$I_M = I_c\sqrt{1+\frac{K^2}{2}}$$

Thus if $K = 1$, the aerial ammeter will indicate an increase of aerial current amounting to only $22 \cdot 5$ per cent; for lower depths of modulation the increase in aerial current will be still less marked. The formula given above is strictly applicable only when the wave is sinusoidally modulated. If the modulation is caused by a flat-topped audio-frequency wave, complete modulation will give rise to a 40 per cent. increase in aerial current. During an actual R/T transmission, however, the aerial current should never show even a momentary increase of this order, for the depth of modulation and audio-frequency wave-form are never constant over periods comparable with the "lag" of the aerial ammeter. We may estimate the probable increase of aerial current, under correct working conditions, as follows. Since it is desirable to limit the peak modulation to 80 per cent., the mean depth of modulation may be taken as 40 per cent. Taking the average depth over a period of several seconds, as distinct from the mean depth, we must allow for intervals between words, say 20 per cent. of the total time, thus the average depth will be about $\cdot 8$ of the mean depth. Inserting $K = \cdot 32$ in the formula, we find

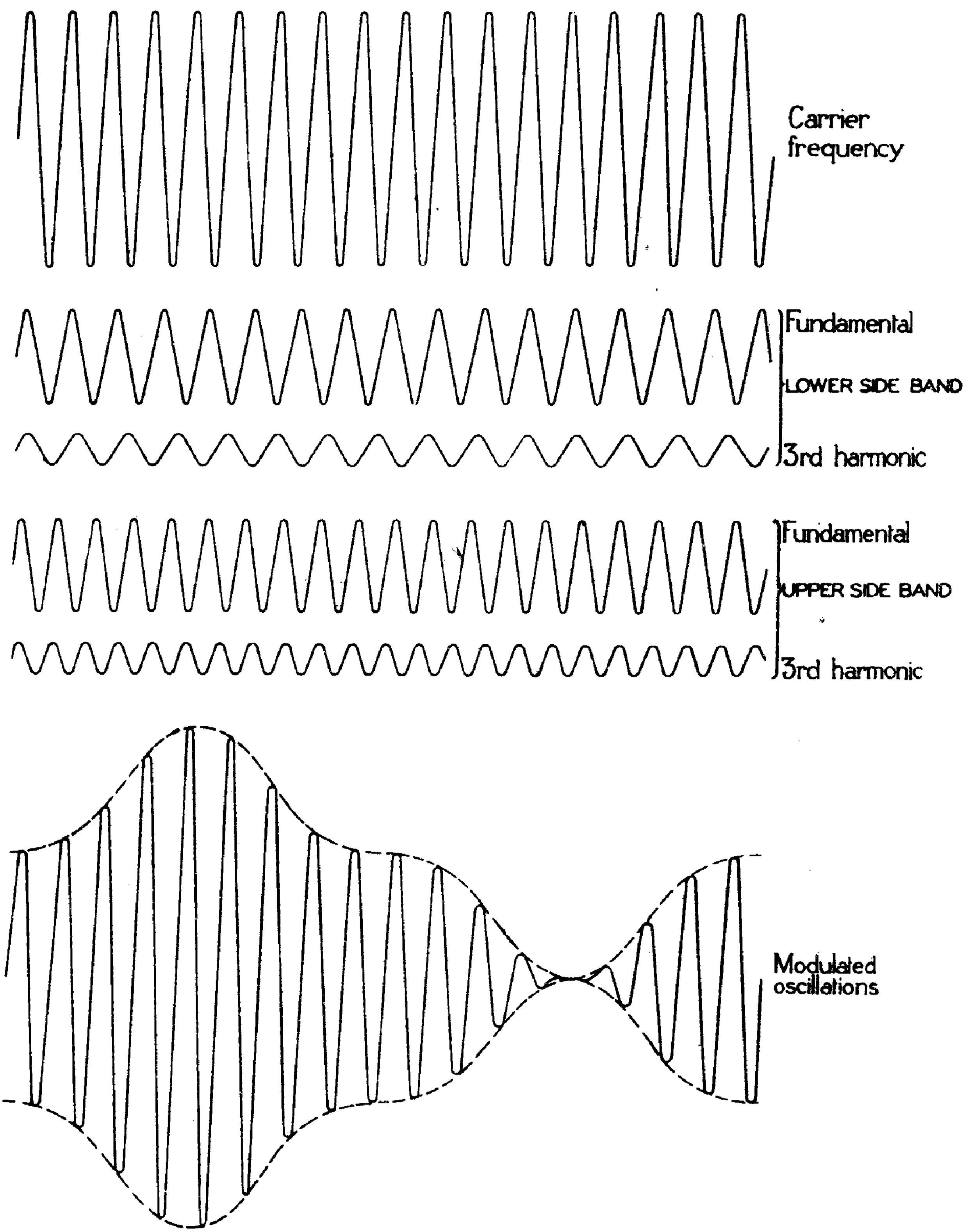
$$I_M = I_c\sqrt{1+\frac{\cdot 32^2}{2}}$$
$$= 1 \cdot 05\, I_c.$$

Although obtained by a very approximate method, this figure agrees with the results obtained on a correctly adjusted transmitter during speech modulation. A 5 per cent. increase of aerial current is not noticeable in a transmitter fitted with the usual thermo or hot-wire ammeter; during modulation, the pointer of such an instrument may be in a state of barely perceptible vibration, but the occurrence of violent movements indicates that the aerial oscillations are



WAVE MODULATED BY FUNDAMENTAL AND +ve. THIRD HARMONIC

FIG. 9
CHAP. XII



WAVE MODULATED BY FUNDAMENTAL AND $-^{ve}$ THIRD HARMONIC

FIG. 10
CHAP. XII

intermittent. This in turn shows that serious over-modulation is taking place, as will be seen on reference to fig. 8b. Similar considerations apply to the modulated stages of frequency-controlled transmitters, without regard to the manner in which the modulation is performed.

Methods of modulation

21. Assuming that we have at our disposal a generator of radio-frequency oscillations, such a simple C.W. transmitter in which the key terminals have been bridged by a conducting link, the radiated wave may be modulated by several different methods. In discussing these it will be assumed that a carbon microphone is used to convert the sound waves into electrical impulses in some portion of the circuit.

(i) *Variation of aerial resistance.*—The simplest method of modulation is to connect the microphone directly in series with the aerial or in the earth lead as shown in fig. 11a. When no speech is taking place, a continuous wave is radiated from the aerial, but on speaking into the microphone, the resulting variation in its resistance causes a corresponding

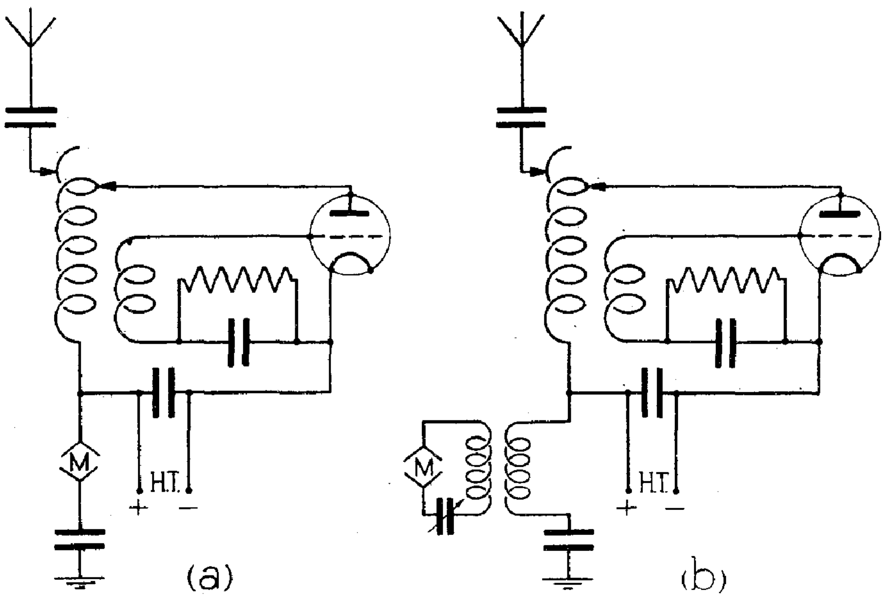


FIG. 11, CHAP. XII.—Modulation by variation of aerial resistance.

variation in aerial current and therefore of the power radiated. The amplitude of the aerial current will vary in accordance with the wave-form of the sound waves impressed upon the microphone and after detection at the receiver, an audio-frequency current of the same wave-form will be obtained. This will of course actuate the reproducing device, i.e. telephones or loud speaker, and so the original speech wave-form will be reproduced. This method is of little practical importance for several reasons. First, if an ordinary carbon microphone is used, its resistance is added to that of the aerial circuit, and this will in itself necessitate a large increase in power input to the transmitter if the same carrier power is required as is obtained in the absence of the microphone. Second, the microphone must be capable of carrying the whole aerial current, and this prohibits the use of a carbon microphone of ordinary design, which will only handle a feed current of a few hundred milliamperes without overheating. Third, the variation of microphone resistance is small compared with its mean resistance, and the variation of aerial current will be small compared with its mean value, hence only a small depth of modulation is obtainable. Some improvement is obtainable by coupling the microphone to the aerial circuit indirectly, (fig. 11b), but even then the utility of the method is limited to low powers and low depths of modulation.

(ii) *Variation of anode potential.*—As an alternative, the microphone may be caused to vary the anode supply voltage to the oscillator valve ; a simple circuit by which this may be achieved is given in fig. 12, in which a self-oscillatory transmitter provides the carrier frequency aerial current. In series with the H.T. supply is included the secondary winding of an iron-core transformer, the primary circuit of which contains a microphone and battery. When the microphone is quiescent the amplitude of the aerial current depends upon the H.T. supply voltage, and if the latter is constant the carrier wave will be of unvarying amplitude. On speaking into the microphone, the variations in its resistance cause variations of current in the primary winding of the transformer and consequent variation of magnetic flux. A corresponding E.M.F. will be induced in the secondary winding, the wave-form of which will closely resemble that of the original sound wave. The supply voltage to the oscillator valve is no longer constant, but consists of this induced E.M.F. superimposed upon the H.T. supply voltage. As the amplitude of the aerial current is proportional to the anode-filament P.D. it is apparent that the radiated wave will be modulated at the speech frequency. This method is of no practical importance owing to the limited depth of modulation obtainable ; as a first approximation it may be

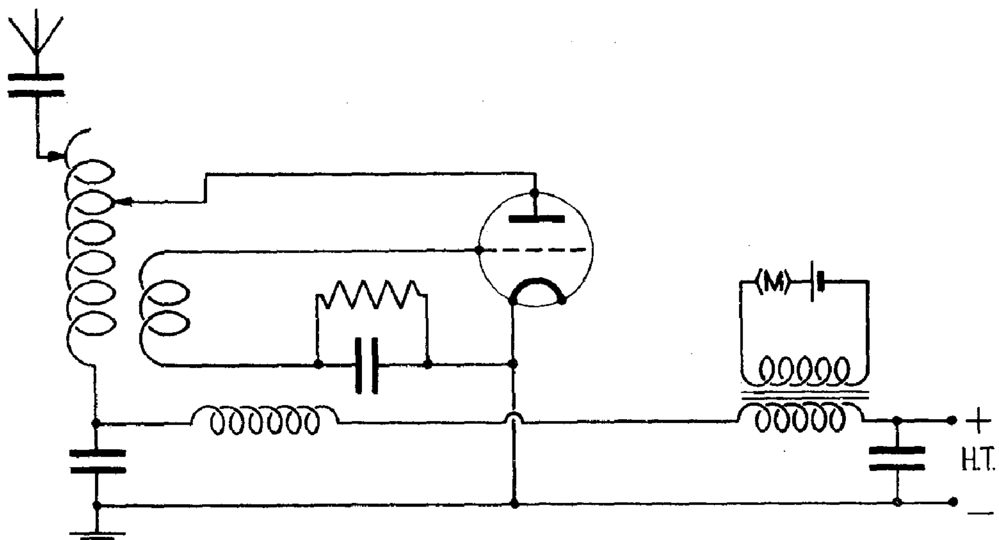


FIG. 12, CHAP. XII.—Modulation by variation of anode voltage.

assumed that in order to attain 100 per cent. modulation the peak value of the transformer secondary E.M.F. should be equal to the constant voltage derived from the H.T. supply. This is impossible with oscillators having a supply voltage higher than about 50 volts, but the method forms the basis of the system known as choke control modulation.

(iii) *Choke control modulation.*—This system is similar in principle to that just described but may be utilized in transmitters of the highest powers used in service or commercial practice.

(iv) *Grid bias modulation.*—This system is suitable for use in low power transmitters, particularly where the primary function is C.W. or I.C.W. telegraphy, and R/T is only occasionally required.

The two latter are in extensive use in service R/T transmitters.

Action of choke control modulator

22. (i) The circuit of a simple form of transmitter using this form of modulation is given in fig. 13, in which O is the oscillator valve and M is called the modulator valve. The iron-core inductance L_3, which is called the speech choke, is in series with the anode of the modulator

valve and also in series with the anode of the oscillator valve. The microphone with its transformer is arranged to supply speech-frequency voltages between grid and filament of the modulator valve, and the action may be outlined as follows. When the microphone is quiescent, the anode-filament P.D. of the oscillator valve is constant, and the oscillator valve maintains an unmodulated aerial current. When sound vibrations impinge upon the microphone diaphragm, the grid-filament P.D. of the modulator valve will vary at the frequency of the sound waves. These changes in grid-filament P.D. result in corresponding variations in the anode current of the modulator valve, and consequently of similar but amplified variations of P.D. between the terminals of the iron-core inductance L_3. As this choke is also in series with the H.T. supply to the oscillator valve, the anode-filament P.D. of the latter will also vary at speech frequency. The amplitude of the aerial current will then vary in like manner, i.e. will be modulated at the frequency of the original sound wave impressed upon the microphone.

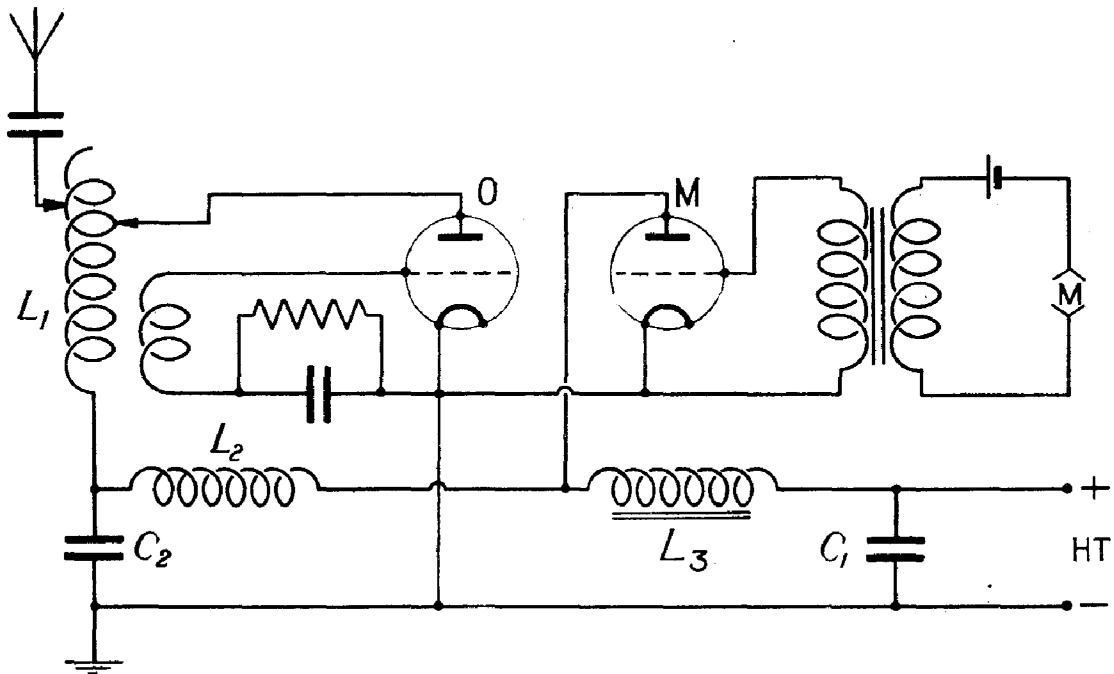


FIG. 13, CHAP. XII.—Choke control modulation.

(ii) Although in this simple explanation of the principle of choke control modulation it is stated that the function of the modulator valve is to set up variations of voltage across the speech choke, in reality the modulator valve is called upon to vary the power supply to the oscillator. It must therefore function as a power amplifier, and the depth of modulation depends upon the amount of power which the modulator valve is able to deliver. It is desirable that the relation between the amplitude of the radio-frequency oscillation and the anode supply voltage to the oscillator valve shall be a linear one, as in fig. 14, in which the amplitude of the oscillatory current in the aerial circuit is plotted against the anode-filament P.D. of the oscillator valve. The " curve " connecting these quantities is a straight line through the origin. If, as in the diagram, the anode-filament P.D. varies sinusoidally at speech frequency, the amplitude of the aerial oscillations will vary in the same manner and will possess a sinusoidal envelope. An approach to this ideal relationship is achieved by operating the oscillator with a large negative grid bias obtained by the condenser and leak method. Under these conditions the amplitude of the oscillatory anode-filament P.D. is only slightly less than the anode supply voltage, so that the desired linear relationship is closely approached.

Load on modulator

23. The load into which the modulator valve must deliver power is best appreciated by re-arranging fig. 13. So far as audio-frequency currents are concerned the inductances L_1 and L_2 offer no appreciable impedance, and may be neglected. The output circuit of the modulator then consists of the speech choke L_3, in parallel with which is a capacitance consisting of the mains condenser C_1 and earth condenser C_2 in series. The anode A.C. resistance of the oscillator valve is, in effect, in parallel with the condenser C_2 and acts as a purely resistive impedance at audio-frequencies, hence the equivalent circuit becomes that shown in fig. 15a. As the mains condenser C_1 is always very large compared to the earth condenser C_2 the circuit may be further simplified, becoming that of fig. 15b. The load impedance into which the modulator valve is called upon to supply power is therefore a very flatly-tuned parallel resonant or rejector circuit, the flat tuning being of course due to the damping imposed by the anode A.C. resistance, r_a, of the oscillator valve.

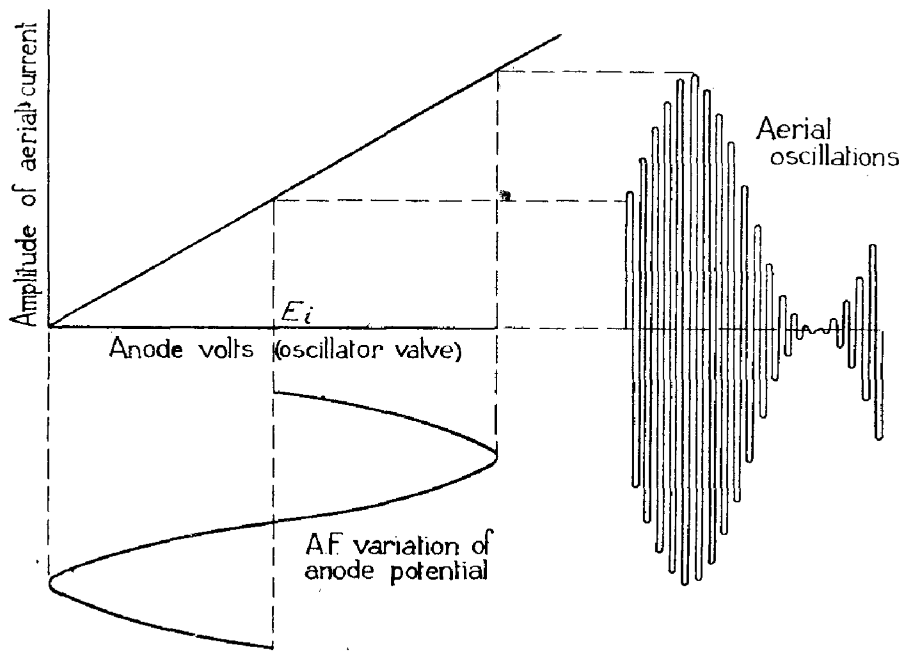


FIG. 14, CHAP. XII.—Ideal relation between anode voltage and aerial oscillations.

Power in carrier and sidebands

24. The power relations for operation under the ideal conditions shown in fig. 14 may be summarized in the statement that the power required to generate the carrier wave is delivered directly to the oscillator valve by the source of H.T. supply, while the power required to generate the side-bands of the modulated wave (although derived from the source of H.T. supply) is delivered in the form of an audio-frequency power output by the modulator valve. When no alternating E.M.F. is generated in the speech choke the power supplied to the oscillator valve is $P_i = E_i I_i$ where E_i is the voltage of the H.T. supply and I_i the average anode current. If the oscillator is working at an efficiency of η_o the output power will be $P_o = \eta_o P_i$. During modulation the modulator valve must supply to the oscillator valve the power which is required to vary the anode voltage of the latter ; for 100 per cent. sinusoidal modulation the peak value of the voltage induced in the speech choke must be equal to the steady component of the anode voltage E_i. For this depth of modulation the carrier power is twice the power carried by the side-bands and consequently the power output of the modulator valve must be equal to one-half the input power to the oscillator under normal (i.e. unmodulated) conditions. With a lower depth of modulation ($K < 1$) the power output of the modulator will be proportional to K^2.

Thus if the oscillator takes 60 watts from the source of supply during quiescent periods, the modulator will give 100 per cent. modulation if its output power is 30 watts. If, however, the modulator valve has a power output of only 7·5 watts, it will modulate the carrier to a depth of only $\sqrt{\frac{7 \cdot 5}{30}} = \cdot 5$ or 50 per cent.

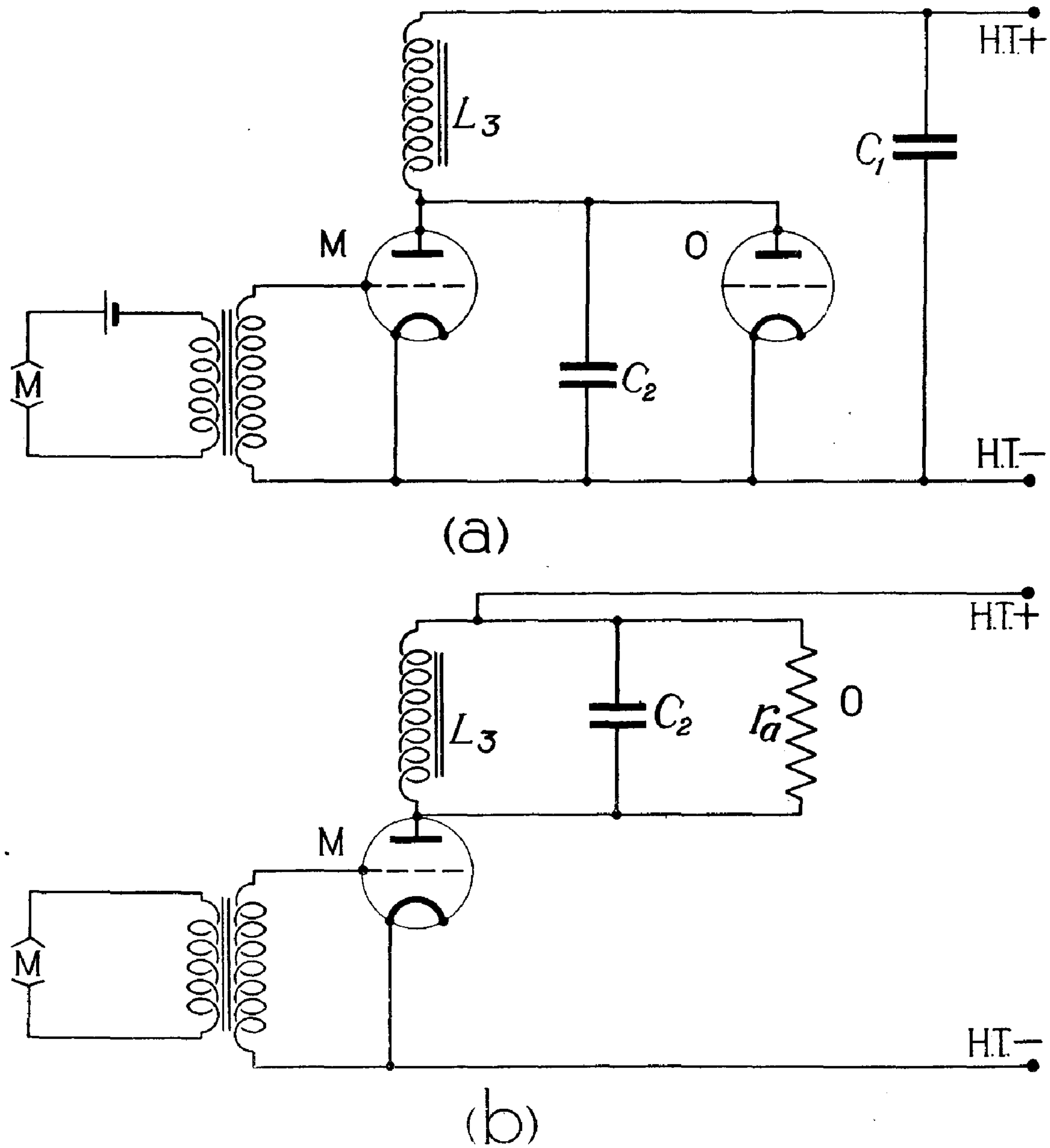


FIG. 15, CHAP. XII.—Equivalent circuit of choke control modulator.

25. If the output is to be completely modulated, the oscillator valve must be operated at an anode voltage E_i somewhat below that permissible for the generation of undamped oscillations, and the permissible power dissipation is also somewhat smaller. The reduction of anode voltage is necessitated by the fact that during complete modulation the input power is 50 per cent. greater than during quiescent periods, and the anode dissipation increases in the same proportion. Let

the permissible dissipation of the oscillator valve be P_L watts, and its input and output powers P_i and P_o watts respectively, then

$$P_o = \eta_o P_i$$

$$P_L = P_i - P_o$$

$$\therefore \quad P_i = \frac{1}{1 - \eta_o} P_L$$

and

$$P_o = \frac{\eta_o}{1 - \eta_o} P_L \quad .$$

If this is the power output of the oscillator valve when completely modulated, the power P_c in the carrier will be two-thirds of this and therefore

$$P_c = \frac{2\eta_o}{3(1 - \eta_o)} P_L \quad .$$

and the input power to the oscillator is

$$P_i = \frac{P_c}{\eta_o} = \frac{2}{3(1 - \eta_o)} P_L \quad .$$

For complete sinusoidal modulation, the output P_m of the modulator valve is one-half this or

$$P_m = \frac{1}{3(1 - \eta_o)} P_L$$

and if η_m is the efficiency of the modulator valve the input to the latter is $\frac{1}{\eta_m}$ times its output. The input to the modulator must therefore be $P_M = \frac{P_m}{\eta_m}$, or

$$P_M = \frac{1}{3\eta_m (1 - \eta_o)} P_L.$$

It is important to remember that if a power amplifier is to operate without appreciable distortion its theoretical efficiency cannot exceed 25 per cent., while in practice it is more likely to be only about 20 per cent. If modulator and oscillator valves are of the same type, it is necessary to use at least three valves in parallel in the modulator valve in order to modulate to a depth of unity the carrier generated by a single valve.

Example :—Using valves having a permissible dissipation P_L of 100 watts, if the oscillator efficiency is 60 per cent., and the modulator efficiency 20 per cent., find the maximum power which can be generated in the carrier, and the power input and output of the modulator in order to modulate the carrier sinusoidally to a depth of unity.

$$\text{Carrier power} = \frac{2}{3} \times \frac{\cdot 6}{1 - \cdot 6} \times 100$$

$$= \frac{2}{3} \times \frac{6}{4} \times 100$$

$$= 100 \text{ watts.}$$

$$\text{Modulator output} = \frac{1}{3} \times \frac{1}{1 - \cdot 6} \times 100$$

$$= 83 \cdot 3 \text{ watts.}$$

The modulator efficiency being 20 per cent., the modulator input will be five times this or 416·5 watts. During periods in which modulation is not actually occurring, the modulator valves are required to dissipate the whole of the input, and therefore, even if four modulator valves are connected in parallel, there will be a slight tendency to overheating.

Variation of load impedance with frequency

26. It has been stated, with reference to figs. 13 and 15, that the anode load impedance, into which the modulator delivers power, is a flatly tuned rejector circuit, and its numerical value and power factor will vary with the frequency. The variation in numerical value, in a particular instance, is shown by the full line curve of fig. 16. This curve was calculated by an approximate graphical method, assuming that $L_3 = 10$ henries, $C_2 = \cdot 007\ \mu F$, r_a (oscillator valve) $= 20{,}000$ ohms. At first glance it might appear that the power output of the modulator valve would also vary considerably with the frequency, but this is not so, because the effect of the power factor is not apparent from this curve. The maximum power output will be obtained when the anode A.C. resistance r_a of the oscillator valve is equal to that of the modulator valve, which will be denoted by r_m. Maximum undistorted output, however, is obtained when $r_a = 2r_m$ (*see* Chapter XI). No matter what the ratio of r_a to r_m may be, the greatest output is always obtained at the resonant frequency of the circuit $L_3\,C_2$, and the output falls off rapidly at frequencies below resonance owing to the shunting effect of the speech choke upon the load resistance r_a. At frequencies above resonance, however, the choke has little effect, and the shunting effect of the condenser C_2 causes a slight reduction in output which, however, is not serious except at the very highest audio-frequencies. The curve shown in dotted line in fig. 16 has been calculated for the

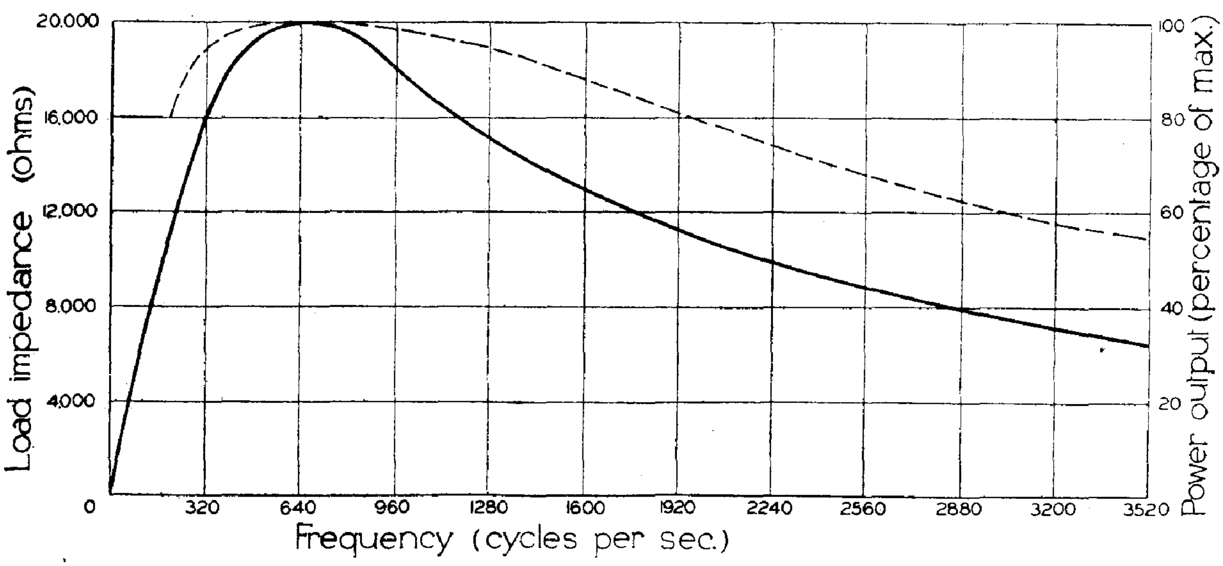


FIG. 16, CHAP. XII.—Load impedance and output power of modulator at various frequencies.

constants already used, and assuming $r_a = 2\,r_m$. It shows the power output at various frequencies as a percentage of the output at the resonant frequency. An increase in the ratio $\frac{r_a}{r_m}$ will give more uniform frequency response at the expense of a reduction of the peak output, while a decrease in this ratio will give a greater peak output with a corresponding increase in both frequency and amplitude distortion. It should be appreciated that the departure from even response at different frequencies is a measure of the frequency distortion of the circuit. The increase of amplitude distortion which follows a reduction in the ratio $\frac{r_a}{r_m}$ is caused by the increased curvature of the dynamic characteristic of the modulator valve owing to the reduction of load resistance. When the respective values of r_m and r_a are such that the desired matching conditions are not satisfied, either a two-coil audio-frequency transformer, or a suitable auto-transformer may be employed as a coupling between the modulator valve and its output circuit, as shown in figs. 17a and 17b respectively.

Modulation of frequency-controlled transmitter

27. With the present-day requirements of constant carrier frequency, the simple choke control modulator circuit with direct aerial excitation has been largely supplanted by circuits in which some form of master-oscillator is employed, the aerial receiving its excitation through one or more stages of radio-frequency power amplification. The modulation may be introduced either in the master-oscillator itself or in one of the amplifier stages. Fig. 18 gives the skeleton diagram of a transmitter in which the master-oscillator stage is modulated. Oscillations of the desired carrier frequency are maintained by the triode T_1. The speech-frequency voltage from the microphone transformer is first amplified by the sub-modulator valve T_3 which is resistance-capacitance coupled to the modulator valve T_2. The speech choke L_3 is common to the anode circuits of both the master-oscillator and the modulator valves, thus setting up an audio-

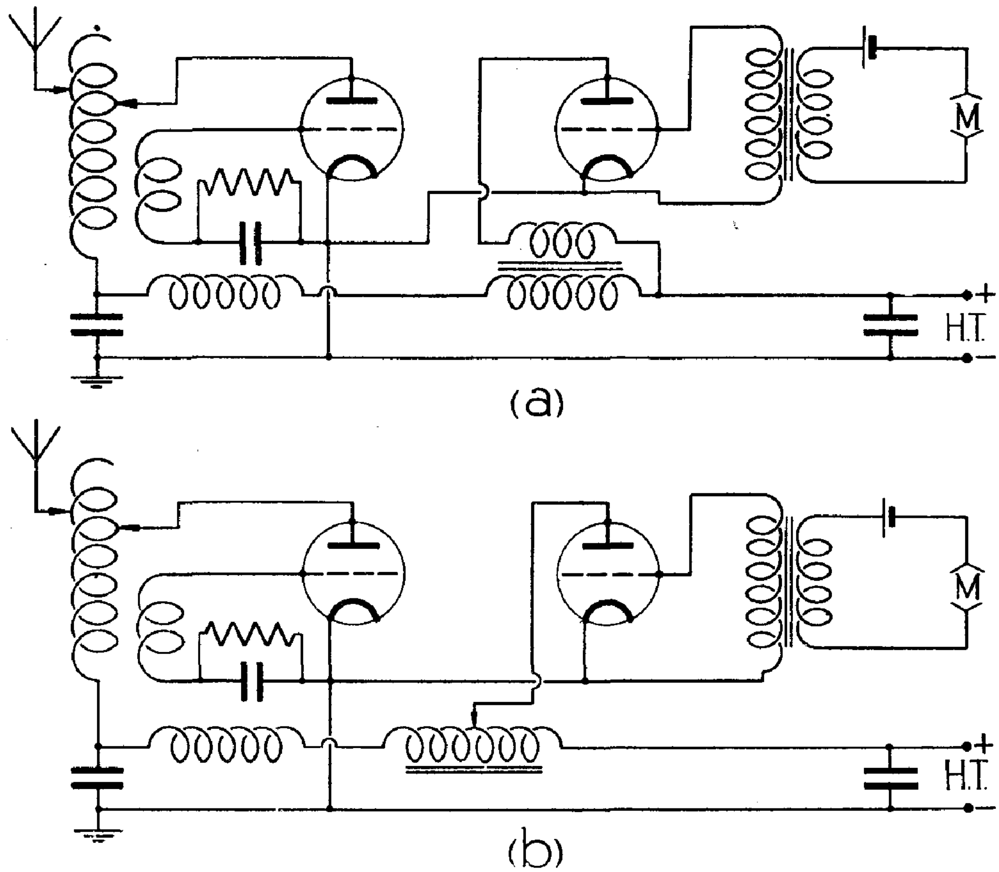


FIG. 17, CHAP. XII.—Alternative circuit arrangements of choke control modulator.

frequency power variation of the radio-frequency oscillation generated in the circuit L_1 C_1. The resistance R is fitted in order that the mean operating anode voltage of the master-oscillator shall be below that of the modulator valve. We have already seen that the audio-frequency voltage generated in the speech choke may be nearly but not quite equal to the applied H.T. voltage of the modulator valve, in practice usually about 85 per cent. If then the anode voltage of the master-oscillator is only 85 per cent. of that of the modulator valve, the speech voltage across L_3 will cause the radio-frequency oscillation in L_1 C_1 to be modulated to a depth of unity. The master-oscillator derives its grid bias from a condenser and leak resistance. As stated in Chapter IX, this method of biasing an oscillator valve tends to give a constant input conductance and is therefore conducive to frequency stability. In the power amplifier, however, the mean grid bias must be maintained at a constant value irrespective of the instantaneous input grid swing, otherwise the valve will be overheated during periods of low depth of modulation, and therefore, in this particular circuit, battery bias is used. The frequency stability of a well-designed circuit of this type is considerably better than when direct aerial excitation is used, but where space, weight and cost are not of primary importance, it is capable of further improvement.

28. For the highest degree of frequency stability, it is desirable to maintain the anode voltage of the master-oscillator at a very steady value, modulating at some later stage. Even then it is possible that the variation in input impedance of the modulator valve may react on the master-oscillator in such a manner as to cause frequency variation. This possibility may be removed by the interposition of a buffer or isolator stage between the modulated stage and the master-oscillator, and the circuit diagram of an R/T transmitter of this kind is shown in fig. 19. It comprises a master-oscillator stage, driven by the valve T_1, and supplying grid excitation to the buffer valve T_5. The speech voltages are amplified by the submodulator valve T_3 and are applied to the modulator valve T_2 as in the circuit previously discussed. The buffer valve in turn supplies excitation to the modulated amplifier valve T_6, the speech choke being common to the anode circuits of the valves T_2 and T_6. The modulated amplifier T_6 is operated under Class C conditions, although if provision is made for complete modulation by a flat-topped audio-

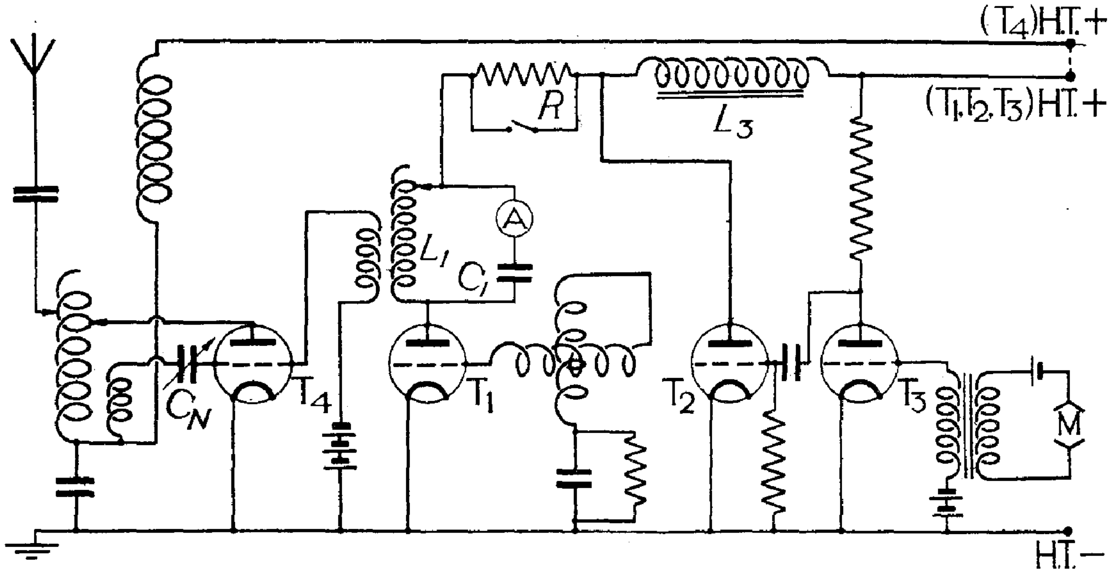


FIG. 18, CHAP. XII.—R/T transmitter with choke control of master oscillator.

frequency wave the average efficiency of this stage will be less than 40 per cent. The modulator must be capable of supplying an amount of power equal to the output of the modulated amplifier, and also of dissipating the whole of its direct current input during periods when no modulation is taking place. Hence the valve T_2 must be capable of dissipating considerably more power than any of the valves immediately associated with it. The modulated output of the valve T_6 is raised to the desired power level by linear amplifiers, which are operated under B class conditions but without running into the grid current region. Two such stages are shown in the diagram. The efficiency of each stage is comparatively low, being only about one-half of that obtainable in a C.W. amplifier. With the valves rated as shown in the diagram, the output of the final linear amplifier would be only about 100 watts. Higher efficiency can be obtained only by operating under conditions which give rise to appreciable distortion.

Grid bias modulation

29. This method is not suitable for high power transmitters but possesses an advantage over choke control modulation in that no separate modulator valve is required. It is possible

to employ grid bias modulation in a self-oscillatory transmitter, provided that the mean grid bias is maintained at a constant value, but this application is of little practical importance. The action will be explained with reference to a transmitter controlled by a valve master-oscillator, the essential features of the circuit being shown in fig. 20. In this diagram T_1 is the master-oscillator valve and T_2 the power amplifier. Oscillations are generated in the circuit L_1 C_1 and radio-frequency variations of grid-filament P.D. are applied to the amplifying valve via the adjustable coupling between L_1 and L_2. These voltage variations cause variations of anode current in the amplifying valve, and consequent impulses of voltage in the aerial circuit, the latter being maintained in oscillation by energy drawn from the H.T. supply. The grid-filament potential of the amplifier valve is also varied at speech frequency by means of the microphone, the latter being included in the primary circuit of a suitable step-up transformer. When speech is not taking place, a steady current is established in the circuit comprised by the microphone battery, the microphone itself and the primary winding of the transformer,

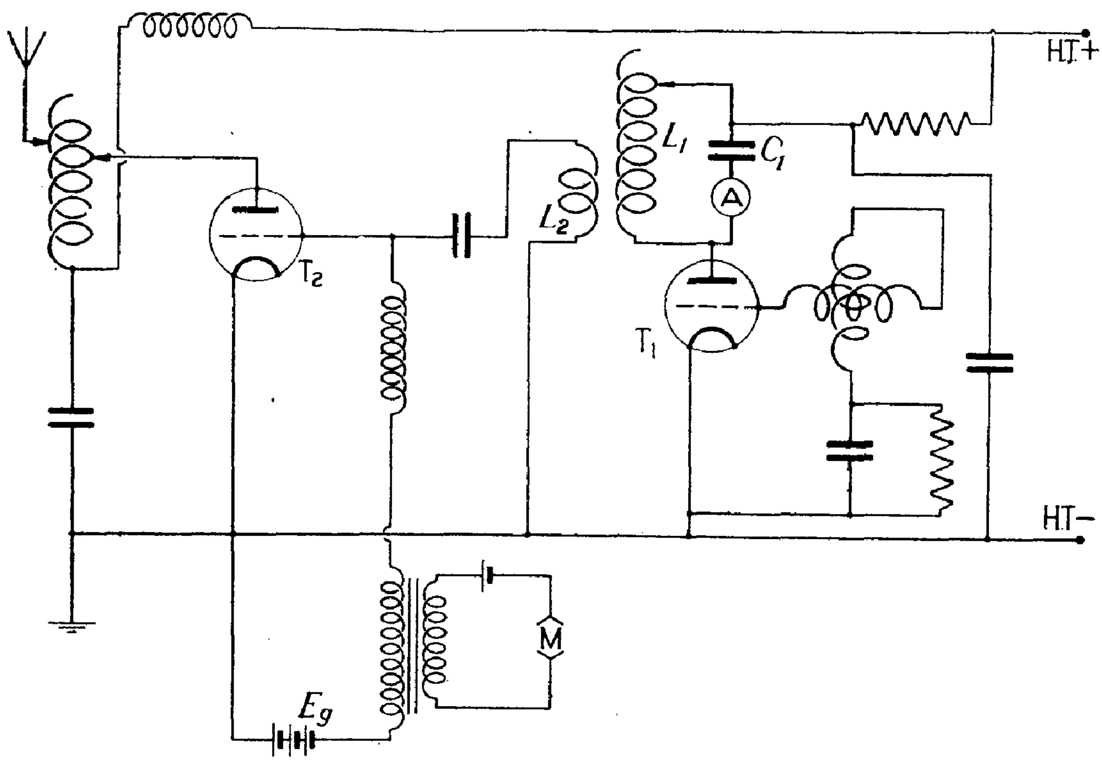
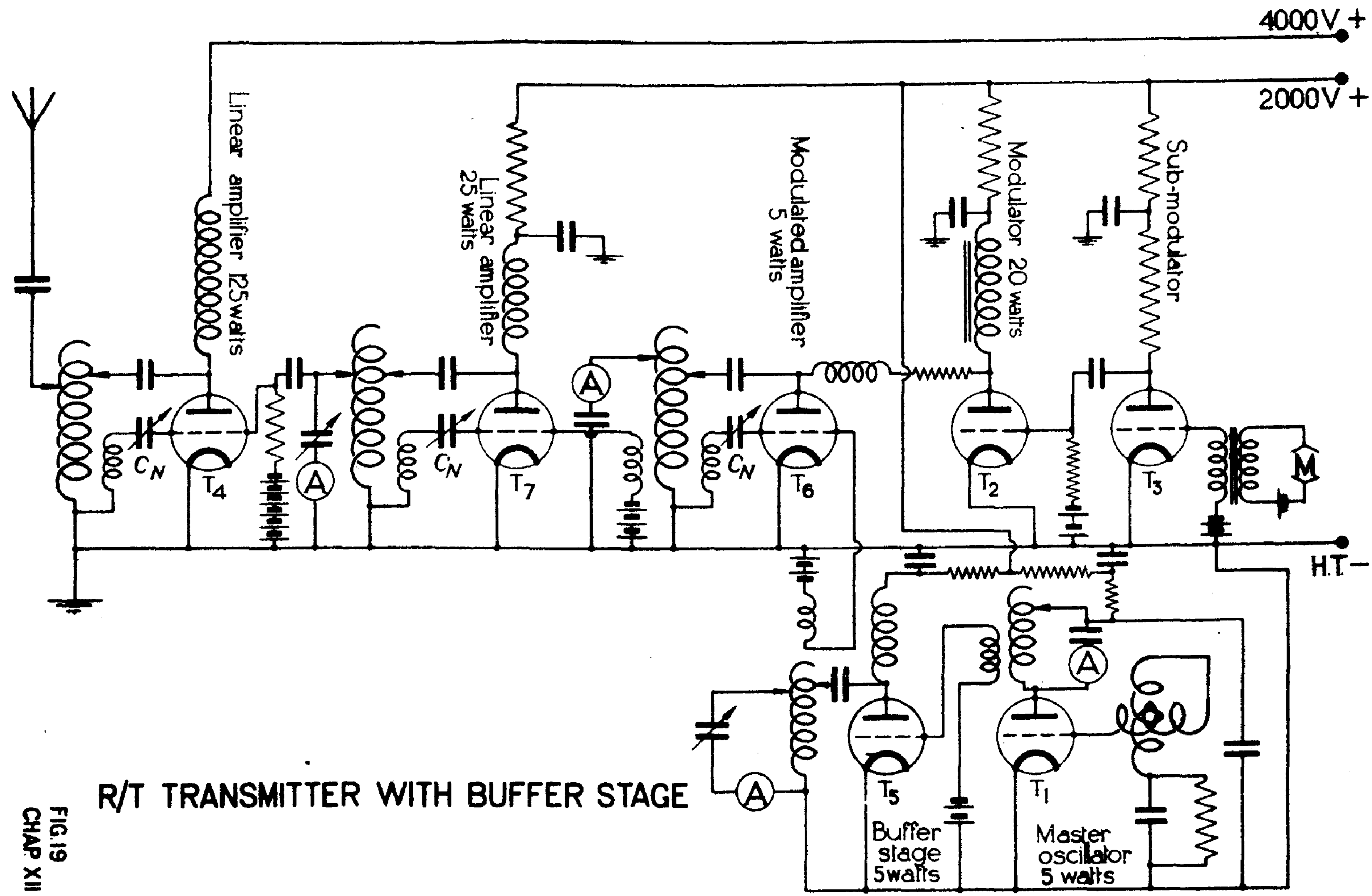


FIG. 20, CHAP. XII.—R/T transmitter with grid bias modulation.

but of course no secondary E.M.F. is generated by this steady current. On speaking into the microphone, however, the variation of resistance causes corresponding changes in the current, which in turn set up a varying flux in the core of the transformer, and an E.M.F. is generated in the secondary winding, the wave form of the secondary E.M.F. closely resembling that of the speech applied to the microphone. As the secondary winding is connected between grid and filament of the amplifier valve, the grid-filament P.D. of the latter is varied at the speech frequency as well as at radio frequency.

30. The effect of the speech-frequency grid-filament P.D. upon the anode current of the valve depends upon (i) the curvature of the $I_a - V_g$ characteristic and (ii) the chosen mean operating point on the curve. Suppose the operating conditions to be as shown in fig. 21, in which the $I_a - V_g$ curve is assumed to be perfectly straight, and the excursions of grid voltage to be limited. The grid-filament voltage consists of three components, namely (i) the steady



R/T TRANSMITTER WITH BUFFER STAGE

FIG.19
CHAP. XII

This Page Intentionally Blank

bias voltage E_g, (ii) the audio-frequency speech voltage of amplitude $\mathcal{V}_a$, (iii) the radio-frequency voltage of amplitude $\mathcal{V}_r$ due to the master-oscillator. Over the limited range of grid voltage shown, the characteristics may be represented by the equation $i_a = g_m v_g$.

As
$$v_g = E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t$$
$$i_a = g_m [E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t].$$

Thus the anode current is merely the sum of a steady component, an audio-frequency component and a radio-frequency component and is not of modulated wave-form. We may therefore conclude that if the excursions of grid voltage and anode current are confined to the straight portion of the characteristic, modulation will not be effected.

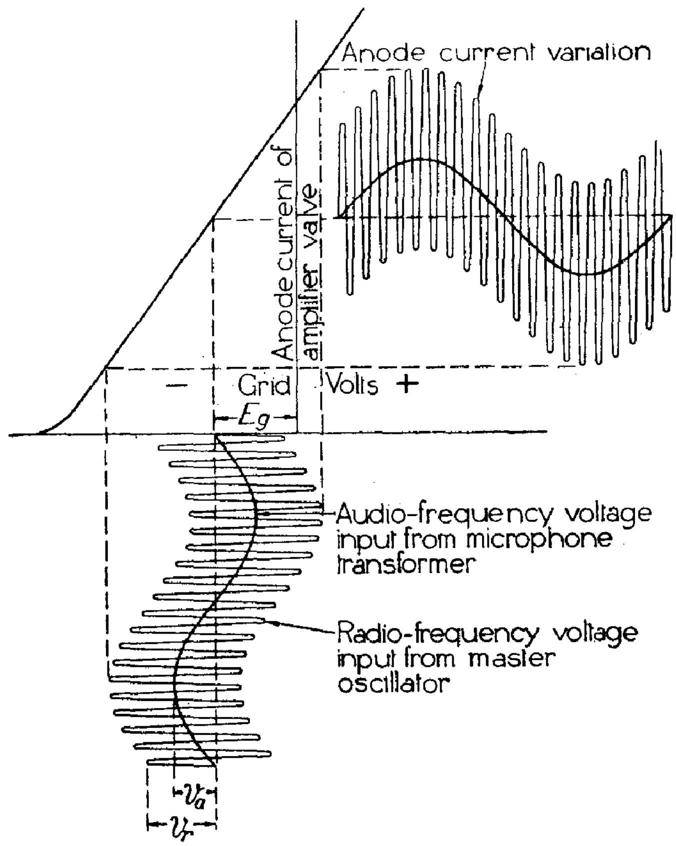


FIG. 21, CHAP. XII.—Operating conditions failing to produce modulation.

"Square-law" modulation

31. The $I_a - V_g$ characteristics of most triodes are approximately parabolic over that portion of the grid voltage lying in the region of negative grid-filament voltage. A fairly complete study of the modulation obtainable with a characteristic of this type is of great assistance in understanding the general principles both of modulation and the detection of modulated waves. The characteristic shown in fig. 22 is a close approximation to the static $I_a - V_g$ curve of a V.T. 25 valve and may be represented by the equation

$$i_a = 360 + 6v_g + \frac{v_g^2}{40} \qquad (1)$$

During modulation, the total voltage applied to the grid is

$$v_g = E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t$$

as before ; in the diagram, $E_g = -50$ volts, $\mathscr{V}_a = 25$ volts and $\mathscr{V}_r = 20$ volts. Equation (1) may be expressed in a more general form as

$$i_a = I_o\left(1 - \frac{2}{E_o}\, v_g + \frac{1}{E_0^2}\, v_g^2\right) \quad \ldots \quad (2)$$

in which I_o is the anode current at $V_g = 0$ (360 milliamperes in the given curve) and E_o is the "cut-off" grid voltage, i.e. the grid voltage necessary to reduce the anode current to zero. E_o must be allotted its correct sign in numerical work. Labour is also economized by writing $b = -\dfrac{2}{E_o}$ and $c = \dfrac{1}{E_o^2}$, so that equation (2) becomes

$$i_a = I_o\,(1 + b\, v_g + c\, v_g^2) \quad \ldots \quad (3)$$

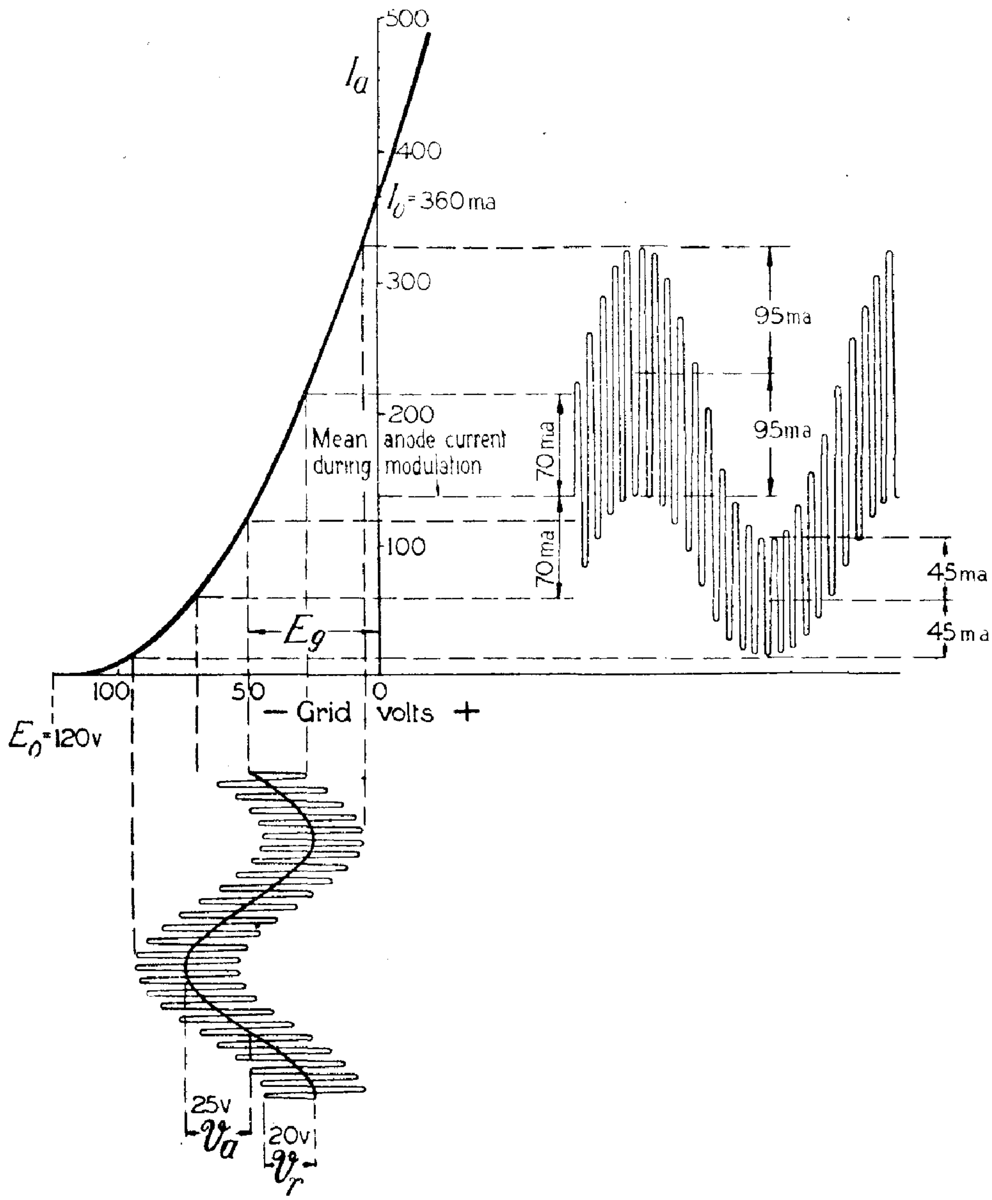


FIG. 22, CHAP. XII.—"Square law" grid bias modulation.

Inserting the above value for v_g, equation (3) becomes

$$\begin{aligned} i_a &= I_o \left[1 + b \left\{ E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t \right\} \right. \\ &\qquad \left. + c \left\{ E_g + \mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t \right\}^2 \right] \\ &= I_o \left[1 + b\, E_g + b\, \mathcal{V}_a \sin \omega_a t + b\, \mathcal{V}_r \sin \omega_r t \right. \\ &\qquad + c\, E_g^2 + 2c\, E_g \left(\mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t\right) \\ &\qquad \left. + c \left(\mathcal{V}_a \sin \omega_a t + \mathcal{V}_r \sin \omega_r t\right)^2 \right] \\ &= I_o \left[1 + b\, E_g + b\, \mathcal{V}_a \sin \omega_a t + b\, \mathcal{V}_r \sin \omega_r t \right. \\ &\qquad + c\, E_g^2 + 2\, c\, E_g \mathcal{V}_a \sin \omega_a t + 2\, c\, E_g \mathcal{V}_r \sin \omega_r t \\ &\qquad + c\, \mathcal{V}_a^2 \sin^2 \omega_a t + 2\, c\, \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t \\ &\qquad \left. + c\, \mathcal{V}_r^2 \sin^2 \omega_r t \right]. \end{aligned}$$

32. This somewhat complicated expression gives the instantaneous value of the anode current and may now be resolved into a steady component and varying components. Only those containing "$\sin \omega_r t$" can possibly cause radio-frequency impulses in the anode circuit. Ignoring the common factor I_o, the terms of interest are therefore $b\, \mathcal{V}_r \sin \omega_r t$, $2\, c\, E_g \mathcal{V}_r \sin \omega_r t$, $2c\, \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t$ and $c\, \mathcal{V}_r^2 \sin^2 \omega_r t$. The latter term is equal to $\frac{c\, \mathcal{V}_r^2}{2}\,(1 - \cos 2\, \omega_r t)$. It is responsible for a second harmonic variation of anode current but will not affect the aerial circuit appreciably. Hence the important terms are $(b + 2\, c\, E_g)\, \mathcal{V}_r \sin \omega_r t$ and $2\, c\, \mathcal{V}_a \mathcal{V}_r \sin \omega_a t \sin \omega_r t$. The radio-frequency variation of anode current is in fact represented by the equation

$$i_r = I_o \left[(b + 2\, c\, E_g) + 2c\, \mathcal{V}_a \sin \omega_a t\right] \mathcal{V}_r \sin \omega_r t$$

or

$$i_r = (1 + K \sin \omega_a t)\,(b + 2c\, E_g)\, I_o \mathcal{V}_r \sin \omega_r t, \quad \ldots \quad (4)$$

which will be remembered as the equation of a modulated wave, the carrier amplitude being $(b + 2\, c\, E_g)\, I_o \mathcal{V}_r$, and the depth of modulation K, where

$$\begin{aligned} K &= \frac{2\, c\, \mathcal{V}_a}{b + 2\, c\, E_g} \\ &= \frac{\frac{2}{E_o^2} \mathcal{V}_a}{-\frac{2}{E_o} + \frac{2\, E_g}{E_o^2}} \\ &= \frac{\mathcal{V}_a}{-E_o + E_g} \end{aligned}$$

Bearing in mind that both E_o and E_g are negative and must be allotted their correct signs in numerical calculation this may be written

$$K = \frac{\mathcal{V}_a}{|\, E_o \,| - |\, E_g \,|}$$

the upright lines denoting that the numerical value of the enclosed quantity is to be inserted without regard to its sign.

In fig. 22 $|\, E_o \,| = 120$ volts, $|\, E_g \,| = 50$ volts, $\mathcal{V}_a = 25$ volts, hence the depth of modulation is $\frac{25}{120 - 50} = \frac{25}{70} = \cdot 3575$ or $35 \cdot 75$ per cent.

33. It must be remembered that the parabolic equation $i_a = I_o\,(1 + b\, v_g + c\, v_g^2)$ only holds if the grid bias and grid swing are so adjusted that the total grid voltage never reaches a negative value exceeding the "cut-off" voltage. The total grid swing is $2\,(\mathcal{V}_a + \mathcal{V}_r)$. The grid potential base line available to accommodate the negative half of this swing is the difference

between E_0 and E_g, and so the maximum permissible grid swing is such that $\mathscr{V}_a + \mathscr{V}_r = | E_0 | - | E_g |$. With the mean bias fixed at -50 volts, as in the diagram (fig. 22) $| E_0 | - | E_g | = 70$, and $\mathscr{V}_a + \mathscr{V}_r$ may be 70 volts but not more. Provided that operation takes place wholly within the parabolic region, therefore, the depth of modulation with the maximum permissible grid swing is

$$K = \frac{\mathscr{V}_a}{\mathscr{V}_a + \mathscr{V}_r}$$

which approaches unity as the amplitude of the radio-frequency component of input voltage approaches zero. The amplitudes of the radio-frequency components of anode current will then of course also approach zero. For maximum modulated output we require that the modulation term, $2\,c\,\mathscr{V}_a\,\mathscr{V}_r$ shall be as large as possible consistent with the maximum permissible value of $\mathscr{V}_a + \mathscr{V}_r$. It is apparent that the maximum modulated output will occur when $\mathscr{V}_a$ and $\mathscr{V}_r$ are equal, and the depth of modulation is then only 50 per cent. The amplitude of the carrier-frequency component of anode current has been shown to be $(b + 2c\,E_g)\,\mathscr{V}_r\,I_0$. In the given example this is $\left(\frac{1}{60} - \frac{2 \times 50}{14{,}400}\right) \times 20 \times 360 = 70$ milliamperes, which agrees with the value shown in the diagram. The amplitude of each side-band component is $\frac{2\,c\,\mathscr{V}_a\,\mathscr{V}_r}{2}$ or $12 \cdot 5$ milliamperes,which however is not immediately obvious from an examination of the diagram because the large amplitude of the audio-frequency variation of anode current tends to mask the amplitude of radio-frequency variation. The value can be found as follows. The total variation of anode current, at the positive peak of the audio-frequency cycle, is 190 milliamperes, and at the negative peak is 90 milliamperes. The amplitude of each side-band is

$$\frac{190 - 90}{8} = 12 \cdot 5 \text{ milliamperes.}$$

34. It will be observed that the above calculations have been made with the aid of the assumed static $I_a - V_g$ characteristic. The dynamic characteristic must not be used because the anode circuit load is not the same for all the component variations of effective anode circuit voltage. If, as is the case in all practical circuits, the anode load is an oscillatory circuit tuned to the frequency $\frac{\omega_r}{2\pi}$, its dynamic resistance at any radio frequency $\frac{\omega}{2\pi}$ near but not necessarily equal to $\frac{\omega_r}{2\pi}$ will be $R_d = \frac{\omega^2 L^2}{R}$ ohms. At radio frequencies considerably larger than this, e.g. of the order of the second harmonic of $\frac{\omega_r}{2\pi}$, the anode circuit impedance will be negligible owing to the low reactance of the aerial capacitance at these frequencies, while at audio-frequencies the anode circuit simply offers a few ohms resistance. The amplitude of the carrier and side band variations of anode current will therefore be less than that calculated above in the ratio $\frac{r'_a}{r'_a + R_d}$, where r'_a is the average anode A.C. resistance of the valve for the particular grid swing and is analogous to the conversion resistance of a frequency-changing valve (Chapter XI). As these currents act as the supply to a rejector circuit, the circulating current in the latter will be approximately $\frac{\omega L}{R}$ times as great as the supply current. For example, assuming that the anode load R_d is equal to the average A.C. resistance r'_a of the valve, the carrier-frequency component in our example will be not 70 but 35 milliamperes, and if the magnification $\frac{\omega L}{R}$, of the anode circuit is 50, the aerial current will be 50×35 milliamperes or $1 \cdot 75$ amperes, in the absence of modulation.

" Linear " grid modulation

35. The method of operation just described is of very low efficiency and is rarely if ever used in radio transmitters although it has a useful sphere of employment in radio-frequency transmission on conductive circuits. It has already been shown that if the mean grid potential of an amplifier is maintained at a highly negative value, higher efficiency is possible, and such an amplifier may be subjected to grid bias modulation, comparatively high efficiency being obtainable with little amplitude distortion. It is again convenient to study this mode of operation with respect to a master-oscillator-controlled transmitter, e.g. fig. 20, and to neglect the effects of curvature of the characteristic, the operating conditions being shown in fig. 23. Here the point A is situated at the theoretical " cut-off " voltage and it is assumed to be permissible to allow a small grid current to flow, the maximum positive grid potential being represented

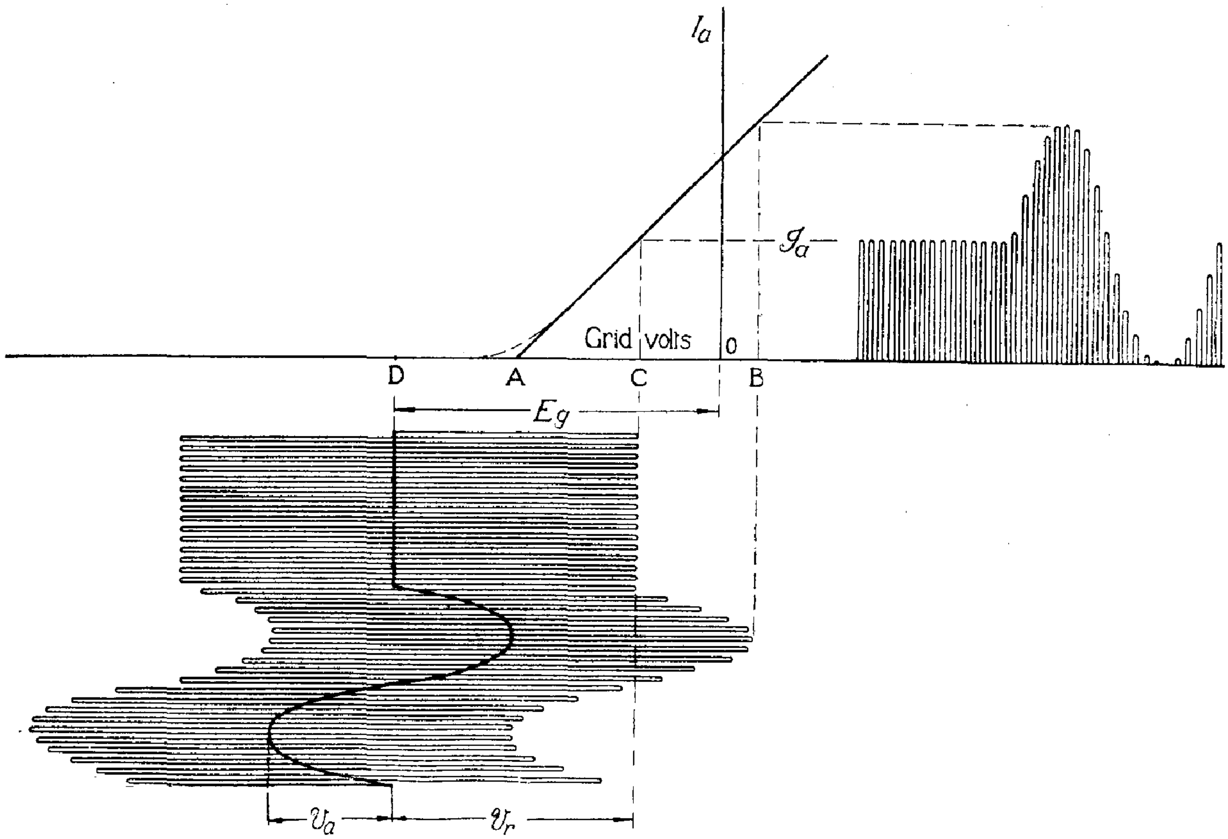


FIG. 23, CHAP. XII.—Linear grid bias modulation.

by the point B while the point C is midway between A and B. The grid bias is sufficiently negative to bring the mean grid potential to the point D where AD = AC, and the maximum permissible grid swing is then 2BD. The amplitude of the radio-frequency input voltage, which is derived from the oscillatory circuit of the oscillator valve, is equal to CD, and the maximum audio-frequency amplitude, derived from the secondary of the microphone transformer, is equal to AD. During periods of peak modulation, the input voltage variations and corresponding changes of anode current are as shown. The audio-frequency envelope of the anode current is a faithful replica of the audio-frequency voltage ; the radio-frequency variation of anode current is not sinusoidal, but contains a whole series of radio-frequency components which are not necessarily in harmonic relationship. Owing to the presence of the oscillatory anode circuit, however, only those components which are near the resonant frequency produce appreciable aerial currents. The amplitude of the radio-frequency variation of i_a is $\mathcal{I}_a\ (1 + \sin\ \omega_a\ t)$ and the total variation will be of the form

$$\mathcal{I}_a\ (1 + \sin\ \omega_a t) \left\{ A_1 \sin\ (\omega_1 t + \varphi_1) + A_2 \sin\ (\omega_2 t + \varphi_2) + A_3 \sin\ (\omega_3 t + \varphi_3) \cdots\cdots \right\}$$

where no restrictions need be placed on the value of A_1, A_2, ω_1, ω_2, etc. The important point to observe is that all the frequencies $\frac{\omega_1}{2\pi}$, $\frac{\omega_2}{2\pi}$, etc., are modulated to a depth of unity, because $(1 + \sin\ \omega_a t)$ is a common multiplier of all these terms. The first of these, $A_1 \sin\ (\omega_1 t + \varphi_1)$ is nearly equal to $\frac{1}{2} \sin\ \omega_r t$, so that the principal effect of the variation of grid voltage is to cause an anode current variation

$$i_m = \frac{r'_a}{r'_a + R_d}\ \frac{\mathcal{I}_a}{2}\ (1 + \sin\ \omega_a t)\ \sin\ \omega_r t,$$

the factor $\frac{r'_a}{r'_a + R_d}$ being introduced to allow for the presence of the dynamic resistance in the anode circuit. Thus i_m is modulated to a depth of unity, at the audio-frequency $\frac{\omega_a}{2\pi}$, and the oscillatory current in the aerial circuit will be modulated in like manner.

36. The effect of a slight curvature at the foot of the characteristic will be to introduce some slight degree of amplitude distortion and, for an input voltage ratio $\frac{\mathcal{V}_r}{\mathcal{V}_a} = 2$, to reduce the depth of modulation slightly below unity. An increase in $\mathcal{V}_r$, so that this ratio slightly exceeds 2, may restore the depth of modulation, but may also increase the amount of amplitude distortion. It is important to observe that this method depends for its effectiveness upon the maintenance

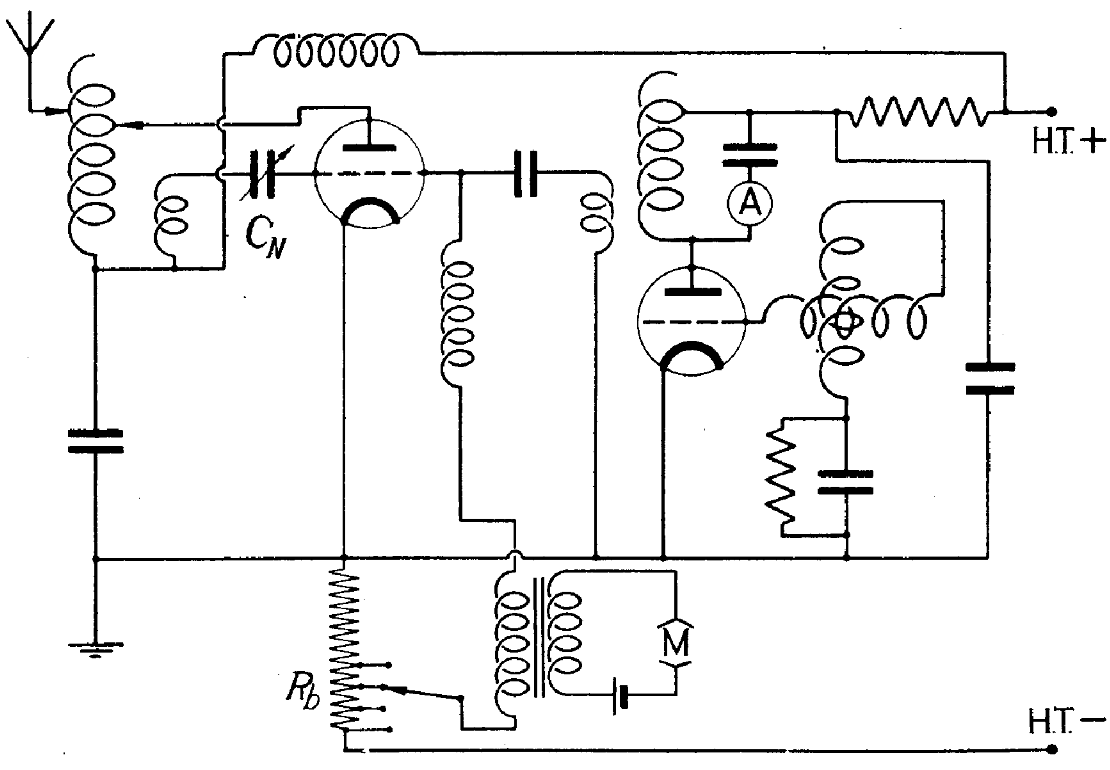
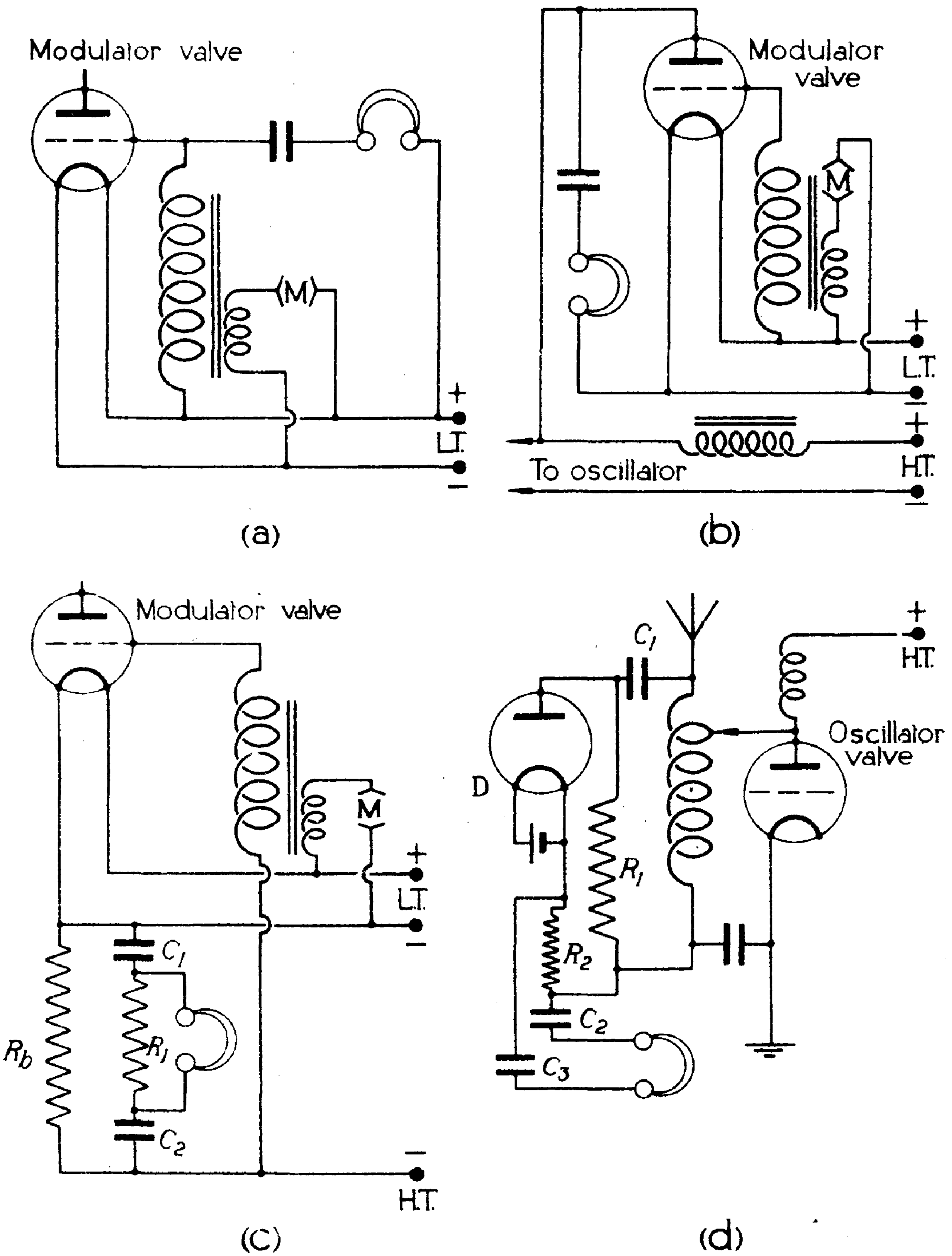


FIG. 24, CHAP. XII.—R/T transmitter with automatic amplifier bias.



SIDE TONE DEVICES

FIG. 25
CHAP. XII

This Page Intentionally Blank

of the mean grid bias at a constant value, irrespective of the instantaneous input grid swing and output power. This requirement forbids the use of a grid condenser and lead for this purpose, for with this device the grid bias varies with the oscillatory output.

37. (i) The circuit diagram of a typical R/T transmitter in which modulation is achieved by the above method is given in fig. 24, in which the neutralizing arrangements for the power amplifying stage are also included. The circuit also shows the method of obtaining automatic grid bias. The resistance R_b is fitted in the H.T. negative conductor and is common to the anode circuits of both valves. A constant P.D. E_g is developed between the ends of this resistance; if I_{pa} and I_o are the mean anode currents of the power amplifier and oscillator valves, respectively, $E_g = R_b\ (I_{pa} + I_o)$. The direction of current is such that the end of the resistance remote from the filament is at negative potential with respect to the latter, and a number of tappings are provided so that the mean bias may be varied within certain limits. The radio-frequency choke in the grid circuit of the power amplifier maintains the input impedance of this valve at a high value; in its absence the input impedance would be low owing to the comparatively large self-capacitance of the microphone transformer.

(ii) In considering the action of the grid modulator it was assumed that the amplifier valve may be allowed to pass a small grid current. If this is permitted, the efficiency of the amplifier stage will be higher, but the frequency stability of the master-oscillator will be impaired, because the passage of grid current implies a variation in input impedance during each audio-frequency cycle and a corresponding change in the effective resistance of the tuned circuit of the master-oscillator. The bias is therefore generally adjusted to such a value that the grid of the power amplifier is never allowed to reach a positive potential with respect to the filament, and the efficiency under these conditions is only about 40 per cent.

Side tone

38. (i) In order that the operator may have some indication of the correct performance of the transmitter, arrangements are usually made to reproduce his speech in the telephone receivers. This reproduction is called side tone, and its utility depends largely upon the manner by which it is achieved. Thus, if the telephones are connected to the secondary winding of the microphone transformer as in fig. 25a (a very small series condenser being inserted to reduce the secondary load and to limit the telephone current) the side tone serves to indicate first that the microphone and transformer are working properly, and second that the speech level is correct. The latter point is of great importance in transmission from aircraft. This method, however, gives no indication as to the performance of the modulator valve.

(ii) In transmitters using choke control modulation, therefore, it is preferable to connect the telephone receiver and series condenser in parallel with the speech choke; this arrangement serves as a complete check upon the audio-frequency operation of the transmitter. In practice the combination of telephone receiver and series condenser is usually connected directly between the anode and filament of the modulator valve (fig. 25b).

(iii) In transmitters employing automatic grid bias, the arrangement shown in fig. 25c, may be adopted. Here R_b is the resistance which provides the automatic grid bias and corresponds with R_b in fig. 24. The telephones are in effect connected in parallel with the resistance R_b. The condensers C_1, C_2 are fitted to confine the steady component of anode current to the bias resistance, and by choice of suitable values, in conjunction with the series resistance R_1, the side tone is maintained at the desired level. The acutal values of R_1, C_1, C_2 depend upon the power of the transmitter.

(iv) The scheme outlined in fig. 25d serves as a check upon all circuits in the transmitter, for side tone is only produced when the aerial is actually radiating a modulated wave. Here D is a diode rectifier, C_1 a high-voltage condenser of small capacitance, and R_1 a suitable resistance. A small radio-frequency current flows in the path C_1, R_1 which is parallel with the aerial, and the oscillatory voltage across R_1 causes a rectified current to flow in the resistance R_2, with which the telephones are effectively in parallel. The condensers C_2, C_3 serve to isolate the telephones from the aerial circuit. The insulation of the condenser C_1 is of great importance, for this component must withstand the sum of the full oscillatory voltage across the aerial coil and the steady H.T. voltage of the supply.

Frequency distortion

39. In all R/T transmitters, irrespective of the method of modulation, since the side-band frequencies differ slightly from the resonant frequency of the tuned circuits associated with the modulator and amplifier stages, some frequency distortion is bound to occur. The amount of this distortion depends upon the relative magnification of the tuned circuits at the side-band and carrier frequencies ; if any circuit has a high magnification the carrier frequency will receive greater amplification than the side-bands and the depth of modulation will be reduced. The problem of avoiding frequency distortion is more difficult on low and medium than on high carrier frequencies, and in any event the side-bands corresponding to high audio-frequencies are attenuated to a greater extent than those corresponding to low audio-frequencies. The obvious remedy is to use tuned circuits having only a low magnification so that the resonant curves are sensibly flat over the frequency band which it is desired to radiate. It is also highly desirable that each resonant curve shall be symmetrical on either side of the resonant frequency so that corresponding side-bands, say 5 kc/s above and below the carrier shall be of equal amplitude.

Effect of unsymmetrical resonance curve

40. (i) Problems involving the relative magnitude and phase of side-bands and carrier are conveniently visualized by means of vector diagrams. The principle will first be exhibited as applied to an oscillatory current of carrier amplitude $\mathcal{I}_c$ and frequency f_c modulated sinusoidally to a depth of unity. Let $\mathcal{I}_H$ be the amplitude of a side-frequency $f_c + f_a$ and $\mathcal{I}_L$ be the amplitude of the corresponding side-frequency $f_c - f_a$. Then the carrier may be represented by a vector rotating in the positive direction f_c times per second, the higher side-frequency by a vector rotating $f_c + f_a$ times per second and the lower side-frequency by a vector rotating at $f_c - f_a$ times per second. With reference to the carrier vector $\mathcal{I}_c$, therefore, $\mathcal{I}_H$ may be considered to rotate at a frequency $+f_a$ and $\mathcal{I}_L$ at the frequency $-f_a$. The negative sign may now be considered to signify contrariety of direction and if the carrier is considered to be stationary, $\mathcal{I}_H$ rotates at the frequency f_a in the counter-clockwise (positive) direction, while $\mathcal{I}_L$ rotates at the frequency f_a in the clockwise (negative) direction. The amplitude of the vector sum of the two side-frequencies may be denoted by $\mathcal{I}_s$ and the amplitude of the modulated oscillation at various intervals during a single audio-frequency cycle may then be derived as shown in fig. 26b. It is seen that if the side-bands are of equal amplitude and in phase opposition as is normally the case, $\mathcal{I}_s$ is always either in phase or in anti-phase with $\mathcal{I}_c$ and the amplitude of the resultant modulated oscillation at any instant during the audio-frequency cycle is obtained by algebraic addition of $\mathcal{I}_c$ and $\mathcal{I}_s$.

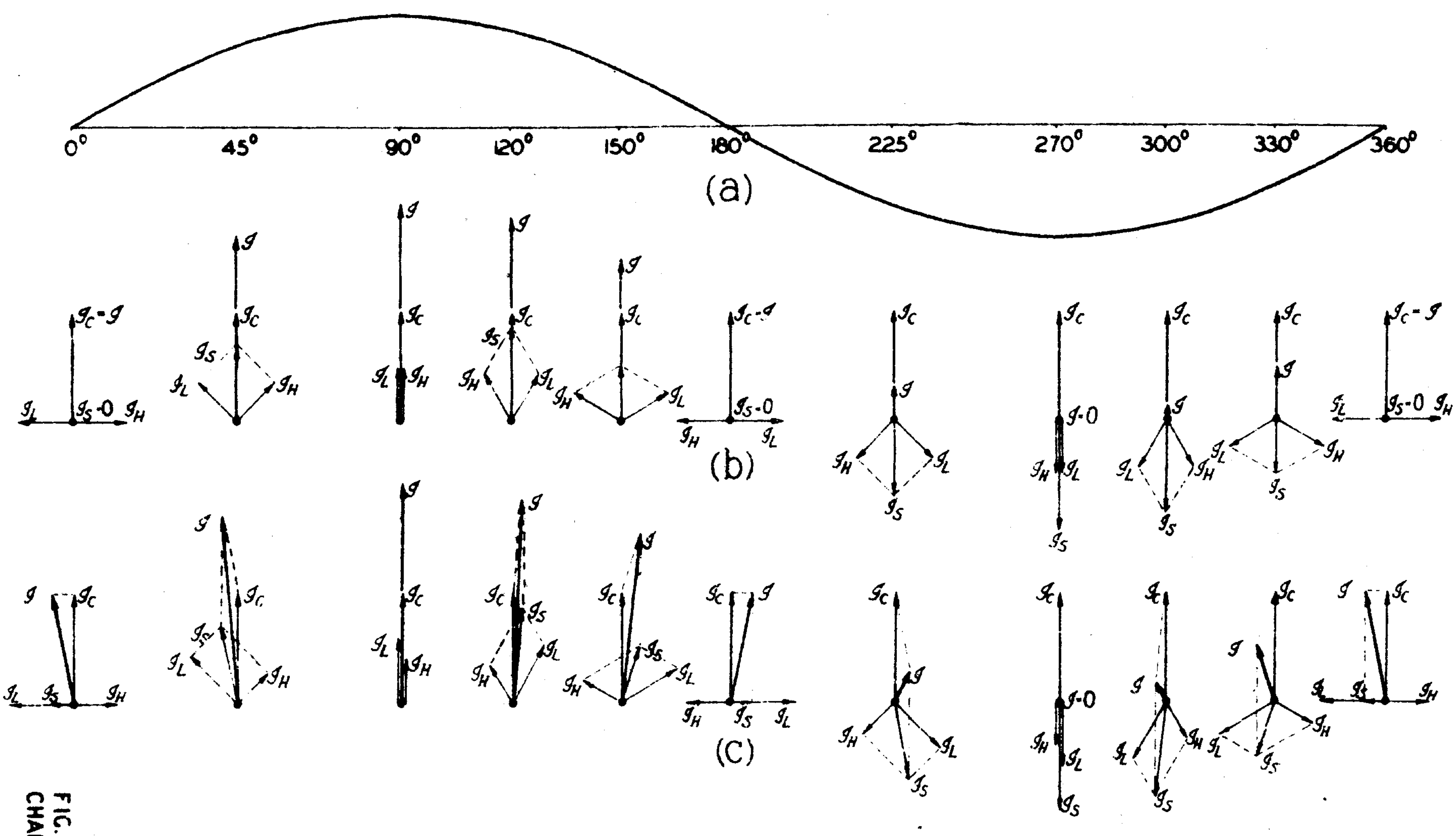
(ii) Now let us consider the effect of unequal amplification of upper and lower side-bands. Suppose that $\mathcal{I}_H$ is not equal to $\mathcal{I}_L$ although they are still in phase opposition. The conditions are now as illustrated in fig. 26c where it is seen that the vector sum of $\mathcal{I}_H$ and $\mathcal{I}_L$ varies in phase with respect to the carrier $\mathcal{I}_c$, the resultant, $\mathcal{I}$, being the vector sum of $\mathcal{I}_s$ and $\mathcal{I}_c$ and not the algebraic sum. The vector $\mathcal{I}_s$ sometimes leads and sometimes lags upon $\mathcal{I}_c$, so that with respect to the carrier frequency, the resultant oscillation undergoes a cyclical change of phase, and is modulated both in phase and amplitude.

Intermodulation by non-linear microphone

41. When a single sinusoidal sound wave represented by $p = P_o \sin \omega t$ impinges upon the diaphragm of a carbon microphone, it was shown in paragraph 10 that the equivalent voltage generated is

$$v_m = -I_m R_m \left\{ \frac{P_o}{P} \sin \omega t \quad \left(\frac{P_o}{P}\right)^2 \sin^2 \omega t \quad \ldots\ldots\ldots \right\}$$

It is of interest to examine the effect produced when two or more sinusoidal sound waves are



VECTORS REPRESENTING CARRIER & SIDE-BANDS

FIG. 26
CHAP. XII

This Page Intentionally Blank

applied simultaneously. Let two sinusoidal sound waves, of amplitudes P_1 and P_2 and frequencies $\frac{\alpha}{2\pi}, \frac{\beta}{2\pi}$ be simultaneously applied, then the total sound pressure on the diaphragm will be $P_1 \sin \alpha t + P_2 \sin \beta t$. The voltage generated will therefore be proportional to

$$\frac{P_1 \sin \alpha t + P_2 \sin \beta t}{P} - \frac{(P_1 \sin \alpha t + P_2 \sin \beta t)^2}{P^2}$$

or

$$\frac{P_1}{P} \sin \alpha t + \frac{P_2}{P} \sin \beta t - \left(\frac{P_1}{P}\right) \sin^2 \alpha t -$$

$$\frac{2P_1 P_2}{P^2} \sin \alpha t \sin \beta t + \left(\frac{P_2}{P}\right)^2 \sin^2 \beta t \ .$$

The terms containing $\sin^2 \alpha t$ and $\sin^2 \beta t$ represent second harmonic distortion as already shown. It will be observed however that the voltage now contains a modulation product $\frac{2P_1P_2}{P^2} \sin \alpha t \sin \beta t$,

$$= \frac{P_1 P_2}{P^2} \cos (\alpha - \beta) t - \cos (\alpha + \beta) t.$$

The result is in fact that owing to the non-linear response of the microphone the two sound waves modulate each other. When they are merely two speech waves of different frequencies and of small amplitudes so that $\frac{P_1 P_2}{P^2}$ is of the same order as $\left(\frac{P_1}{P}\right)^2$ and $\left(\frac{P_2}{P}\right)^2$ the result of this intermodulation is no more serious than the introduction of second harmonic distortion, but if one of them, say P_2, represents a very loud interfering noise such as that caused by the airscrew in an aeroplane, it may be many times greater than the average pressure due to speech, and the sum and difference terms will give rise to considerable loss of intelligibility. The actual effect is, of course, much more complex than when only two sinusoidal components are considered, but in general the result at the receiver is to impart to the speaker's voice a peculiar tremolo character which is unmistakable. It must be noted also that this effect is additional to the interference caused by the transmission of the frequency $\frac{\beta}{2\pi}$ in the ordinary manner. Intermodulation can only be completely overcome by the use of a microphone having a perfectly linear characteristic. Either moving coil or condenser microphones would probably effect a considerable improvement, but only at the cost of a considerable increase in the weight, volume and complexity of the installation. The latter considerations dictate the use of carbon microphones in most aircraft installations, but intermodulation by engine noises is reduced to a minimum by careful design of the mask microphone.

Suppressed carrier and single side-band telephony

42. It has been shown that in the ordinary amplitude-modulated R/T transmitter, the power carried by the carrier frequency is at least equal to and may be very much greater than the power carried by the side-band frequencies, although only the latter convey the desired intelligence. The carrier frequency is required in the first instance to produce the side-bands by interaction with the audio-frequency input, and it is necessary at the receiver in order to avoid distortion. Once the desired side-bands have been obtained there is however no reason why the carrier frequency should not be eliminated in the amplifier stages of the transmitter, and reintroduced at the receiver. The transmitter is then required to handle considerably less power and may also operate at a higher efficiency. One method of producing the desired side-bands and eliminating the carrier is by means of the balanced modulator shown in fig. 27 which is based upon the circuit used for grid modulation. The two triodes T_1, T_2 are of similar (preferably identical) characteristics, and the input and output circuits are arranged symmetrically with respect to them. In the absence of an audio-frequency input the grid-filament potential is varied in accordance with the radio-frequency input derived from the master-oscillator. To

this voltage the grid-filament paths of the valves are in parallel and the output circuit is in push-pull, so that the radio-frequency voltages developed in the two halves, L_1, L_2, of the inductance in the output circuit are in opposition. No power is developed in the oscillatory circuit L_a C_a under these conditions. To the audio-frequency voltage supplied by the microphone transformer, both input and output circuits are in push-pull, and the grid-filament voltages of the two valves are in antiphase. The result of the simultaneous application of audio-frequency and radio-frequency input voltages is to develop, in the output circuit, an oscillatory E.M.F. corresponding to the beats between side-bands. Thus, if the audio-frequency input voltage is of sinusoidal wave-form, the output voltage is as shown in fig. 6c, taking the form of heterodyne beats, and it must be remembered that two beats are formed for each audio-frequency cycle. This voltage may now be amplified as found desirable, the final output circuit being the transmitting aerial.

43. Neglecting any distortion due to the medium of propagation, the aerial current at the receiver will have precisely the same wave form, and on rectification this would give an audio-frequency output of double the original modulating frequency; if this disadvantage is to be

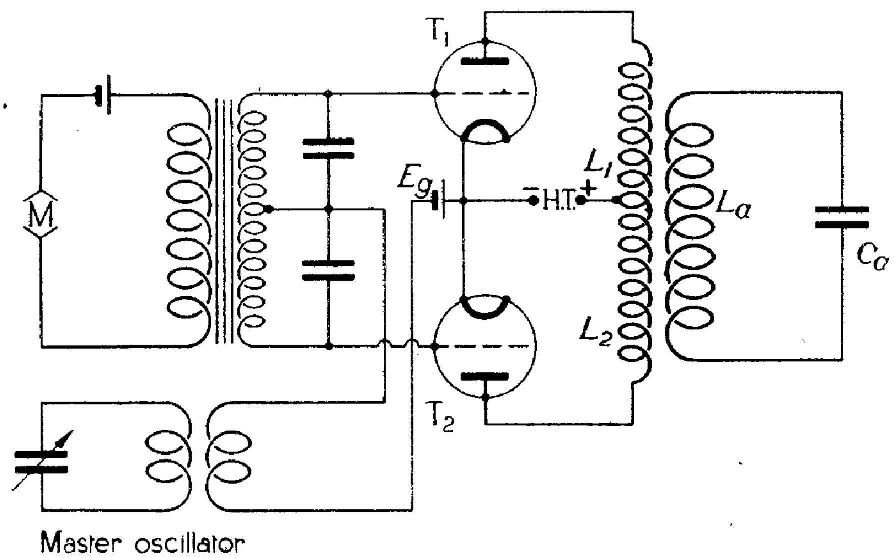


FIG. 27, CHAP. XII.—Balanced modulator.

avoided, the carrier must be reintroduced before the wave is rectified. The mere replacement of the carrier, however, is not in itself sufficient to ensure that the detector output will possess a wave-form corresponding with the original. Referring again to fig. 6, it is seen that the sinusoidal envelope of the amplitude modulated wave is entirely dependent upon the change of phase between adjacent beats. If the carrier is not reintroduced in its correct phase, the rectified output of the detector will be distorted. Fig. 28a has been developed to show what happens in an extreme case, when the reintroduced carrier is 90° out of phase with respect to the original carrier. A single " beat " is drawn in light line and the replaced carrier in heavy line. The respective amplitudes are such that if the relative phase were correct the resultant would be modulated to a depth of unity. Actual point-to-point addition of the two waves, as in fig. 28b, shows however that the resultant of the two waves is modulated to a depth of only about 15 per cent. and the shape of the envelope indicates that rectification would still give an output at double the original frequency. Thus the reintroduction of the carrier with a phase displacement of this magnitude would reduce the strength of the rectified signal without reducing the distortion. From the practical point of view, this rules out the possibility of replacing the carrier merely by using, as part of the receiving equipment, a separate heterodyne tuned to the carrier frequency, although with the introduction of automatic tuning control there is some hope of a solution of the problem. At the present time, however, suppressed carrier working is rarely if ever adopted.

44. Since each of the two side-bands contains the whole of the intelligence to be transmitted it is possible to convey the desired signal by means of only one side-band, eliminating the carrier by means of the balanced modulator and the unwanted side-band by means of a band-pass filter of the Campbell-Zobel type. The carrier is then reintroduced at the receiver by means of a separate heterodyne, which beats with the frequencies embraced by the single side-band. After rectification the resulting difference frequencies reproduce the original audio frequencies, and although some distortion is introduced, the effect of the relative phase shift of replaced carrier and single side-band is not so serious as when the carrier only is suppressed, provided that the amplitude of the heterodyne oscillation is much larger than that of the received signal. Single side-band working is not more economical than suppressed carrier working, but it reduces the frequency band occupied by a given transmitter by one-half. The additional complexity

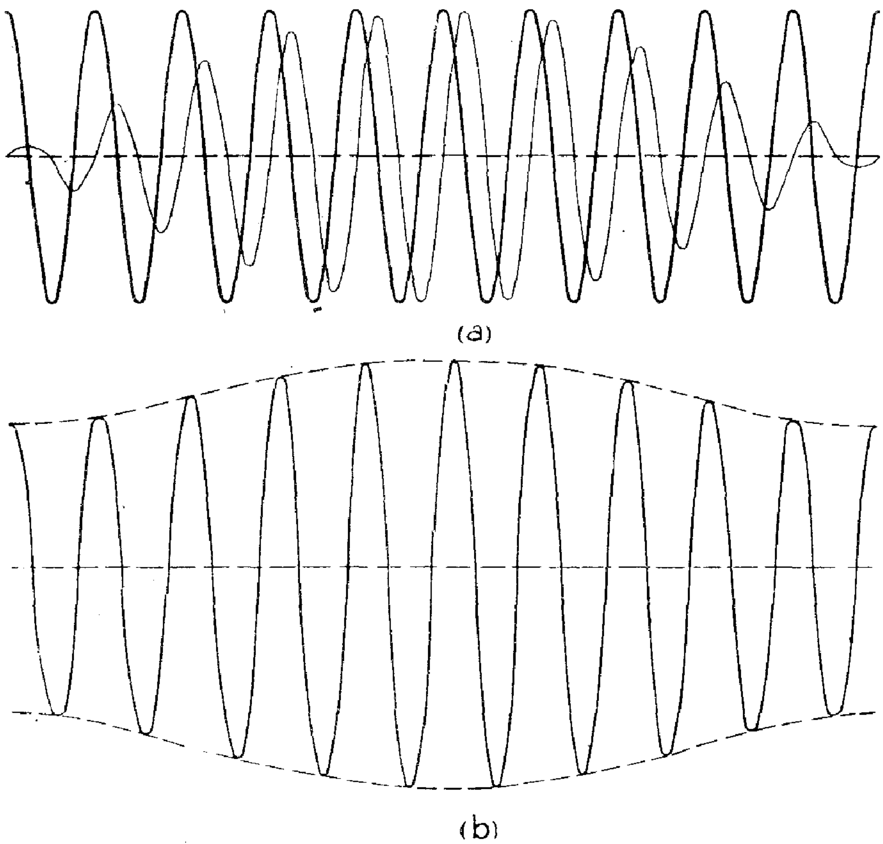


FIG. 28, CHAP. XII.—Effect of phase displacement of carrier.

renders it suitable only for ground-to-ground communication between high-power stations operating on a single fixed-frequency channel. The stringent operating requirements are more easily met on low and medium than on high frequencies, although progress is being made in its application to the latter.

R/T RECEPTION

45. In principle, the circuits used for reception of amplitude-modulated radio-telephonic signals are very similar to those used for telegraphic reception. The arrival of modulated electromagnetic waves at the receiving aerial sets up an induced E.M.F. of identical wave form, a corresponding radio-frequency current being established in the aerial circuit. The process of detection in this case is the reverse of the modulation process at the transmitter, for the detector

is required to give an audio-frequency output having a wave-form identical with that of the envelope of the modulated signal. Radio-frequency amplification may be and in practice generally is employed in order to achieve the highest degree of selectivity possible (space, weight and cost of the receiver being taken into account), and also in order that the detection process shall be carried out efficiently. After this, audio-frequency amplification is employed to bring the final reproduction to a level sufficiently high to be audible over the local noise, either head telephones or loud speaker being employed to convert the electrical power output of the final amplifier stage into sound waves. Whereas for C.W. reception the sensitivity of the receiver is of chief importance, neither amplitude, phase or frequency distortion being detrimental to efficient reception, in an R/T receiver fidelity is an important consideration, i.e. the audible output of the receiver must be as faithful a reproduction as possible of the original sound vibrations.

Selectivity

46. (i) To a first approximation, distortion of the radio-frequency wave-form is of no importance provided the original shape of the modulation envelope is preserved ; to achieve this the radio-frequency circuits preceeding the detector stage must give equal magnification of the

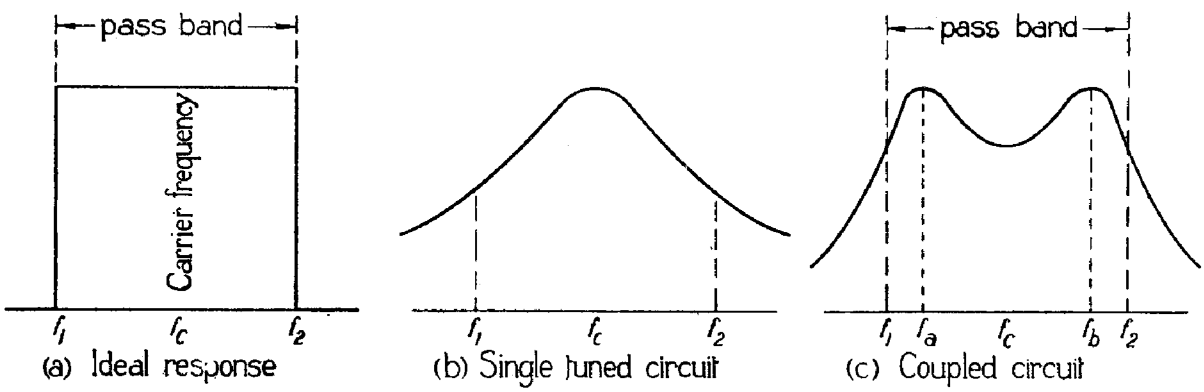


Fig. 29, Chap. XII.—Ideal and practicable response characteristics.

whole range of frequency embraced by the carrier and side-bands, and an ideal response characteristic for this portion of the receiver would have the form given in fig. 29a, in which all frequencies between f_1 and f_2 receive equal amplification while frequencies outside this range are attenuated to zero. The frequency range between f_1 and f_2 is called the " pass-band "

(ii) If the receiver contains only a single tuned circuit, its frequency response will be governed by the resonance curve of the tuned circuit and will resemble fig. 29b. Since all frequencies above and below the carrier frequency f_c are attenuated to some extent, the receiver cannot strictly be said to possess any pass-band whatever. An approach to the ideal may be obtained by the use of coupled circuits of the types described in Chapter VI, which have a frequency response similar to that shown in fig. 29c. The pass-band of such a receiver is usually taken as the region within which the height of the curve is equal to or greater than its height at the carrier frequency f_c. This is 1·414 times the frequency separation of the " peaks " of the resonance curve. For telephony of commercial quality a pass band of some 5 kc/s is sufficient, but for faithful reproduction of music and particularly of noises such as the rattling of keys, all frequencies up to about 15 kc/s are required and the pass-band of the receiver should be some 30 kc/s wide. The shape of the resonance curve of two coupled circuits depends upon the coupling factor, k, and upon the circuit magnification of each member. If the magnification of the primary circuit is χ_p, that of the secondary circuit χ_s, and both are tuned to the frequency f_c, it is shown in Chapter VI that provided $k^2 > \frac{1}{\chi_p \chi_s}$ the resonance curve possesses two peaks, the frequencies at which they occur being

$$f_a = \frac{f_c}{\sqrt{1+k}}$$

$$f_b = \frac{f_c}{\sqrt{1-k}}$$

while a trough occurs at the frequency f_c. It must be observed that f_a does not correspond with f_1, nor f_b with f_2, in fig. 29c. The peak separation is $f_b - f_a$ and the pass-band is $f_2 - f_1 = \sqrt{2}\,(f_b - f_a)$.

47. The relative magnification at the frequencies f_a and f_c (or f_b and f_c) depends upon the circuit magnification. In order to obtain an approach to the ideal response characteristic, fig. 29a, it is obvious that the peaks of the resonance curve must be nearly but not quite suppressed. The width of the pass-band depends chiefly on the coupling factor k, while the uniformity of response within the band depends upon k, χ_p, and χ_s. The degree of discrimination against frequencies just outside the pass-band also depends upon k, χ_p, and χ_s and is not under

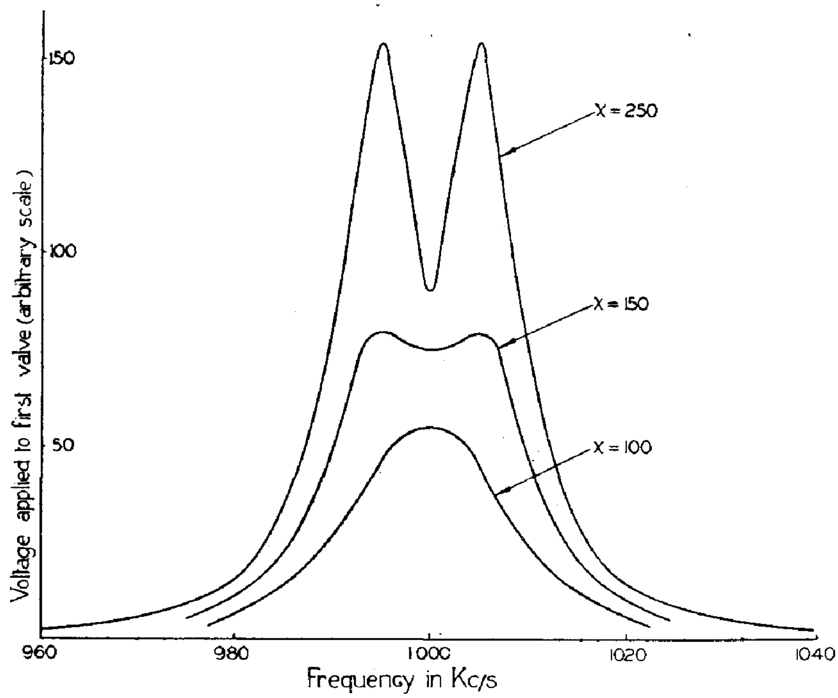


FIG. 30, CHAP. XII.—Effect of varying resistance of coupled circuit.

separate control. In showing the effect of varying the magnification and coupling factor it is convenient to assume that $\chi_p = \chi_s = \chi$. The effect of varying k, while keeping χ constant, has been shown in Chapter VI. In fig. 30 the coupling factor is maintained at a constant value ($k = \cdot 01$) and the circuit magnification is varied by adjustment of the resistance. It is seen that if a high value of χ is adopted, the response characteristic possesses pronounced double peaks, so that after rectification the higher audio-frequencies will be exaggerated and the reproduction will appear to be highly pitched. On the other hand, very low values of χ cause the lower audio-frequencies to receive greater magnification than the higher, and the reproduction will appear " woolly ". The curve marked $\chi = 150$ represents a very close approach to the ideal, giving a very even response over a band width of about 13 kc/s and a high degree of discrimination against the frequencies outside the pass-band.

48. The effect of simultaneously increasing k, in order to obtain a wider pass-band, while at the same time reducing χ (by an increase of resistance) in order to suppress the double peaks, may be seen from fig. 31. A wide band having an even response can only be achieved at the expense of a reduction in output voltage. The effective magnification of the whole arrangement is $\frac{1}{2}\sqrt{\chi_p \chi_s}$, e.g. if $\chi_1 = \chi_2$ the voltage gain is only one-half that which would be given by a single circuit of the same magnification. This is the price which must be paid in return for the "band-pass" effect. The curves of figs. 30 and 31 were calculated on the assumption that the circuits are coupled by mutual inductance, but the above deductions are applicable to all forms of reactive coupling. Resistance coupling is seldom if ever deliberately adopted in receiving circuits. The term "band-pass filter" is used in broadcast receiver practice to describe an arrangement of two coupled circuits in which both members are simultaneously adjusted to the same frequency, e.g. by using equal inductances in each member and inter-connecting the

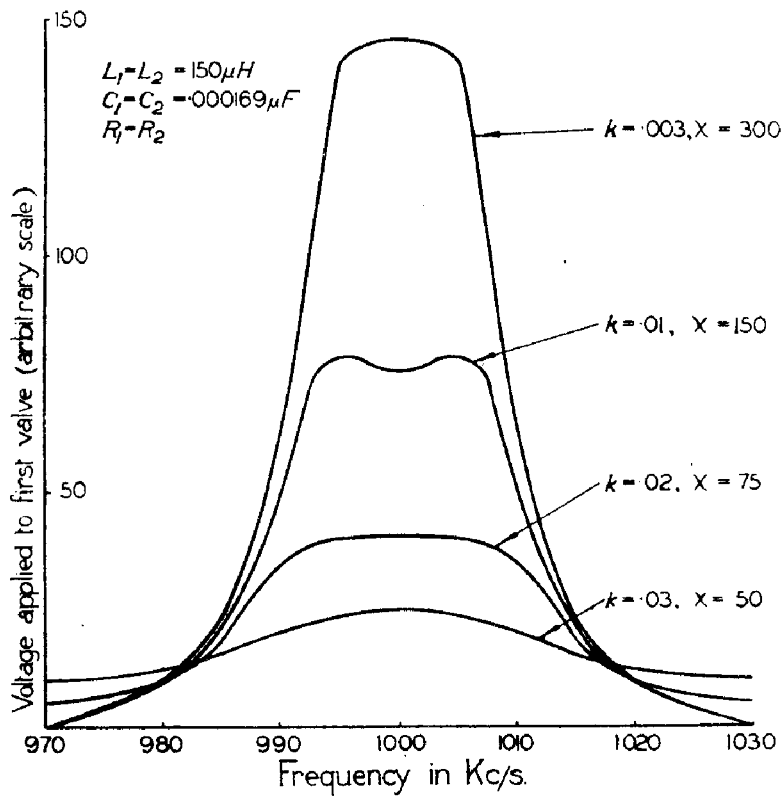


FIG. 31, CHAP. XII.—Effect of simultaneous variation of magnification and coupling factor.

tuning condensers in such a manner that both are adjusted to the desired capacitance by a single tuning control. In choosing the type of coupling reactance an important consideration is the variation in width of pass-band as the tuning is varied. Thus if mutual inductive coupling is employed, since the width of the pass-band is $f_2 - f_1 = \sqrt{2}\,(f_b - f_a) = \sqrt{2}\,f_c\,k$, the pass-band increases with the frequency to which the individual circuits are tuned. Assuming that the maximum capacitance of the condenser is ten times the minimum (including of course the distributed capacitance of the coil and winding) it is seen that if the band width is adjusted to 10 kc/s at the lowest frequency of operation, for which the tuning condenser is adjusted to its maximum value, the band width at the highest frequency will be $\sqrt{10} \times 10$ kc/s or $31 \cdot 6$ kc/s. This follows from the relation $f_c = \frac{1}{2\pi\sqrt{LC}}$, because if L is constant $f_c \propto \frac{1}{\sqrt{C}}$. On the other hand, if auto-capacitive coupling is employed, $k = \frac{C}{C + C_m} \doteqdot \frac{C}{C_m}$, because C_m is always very much larger than C. The band width is therefore directly proportional to $\sqrt{C}$ and increases as the operating frequency is decreased. In practical circuits of this type an attempt is frequently made to overcome this disadvantage by a combination of different forms of coupling.

Amplitude distortion

49. (i) While the radio-frequency circuits may cause some degree of frequency distortion, the first source of appreciable amplitude distortion is the rectifying valve, by which the modulated wave is resolved into its radio-frequency and audio-frequency components and the former discarded. It has already been shown that for small amplitudes of input voltage all practical rectifiers give an output which is proportional to the square of the input (except in the case of heterodyne reception). The "square law" rectification of a completely modulated wave by means of a diode will result in an anode current variation somewhat as depicted in fig. 32, in

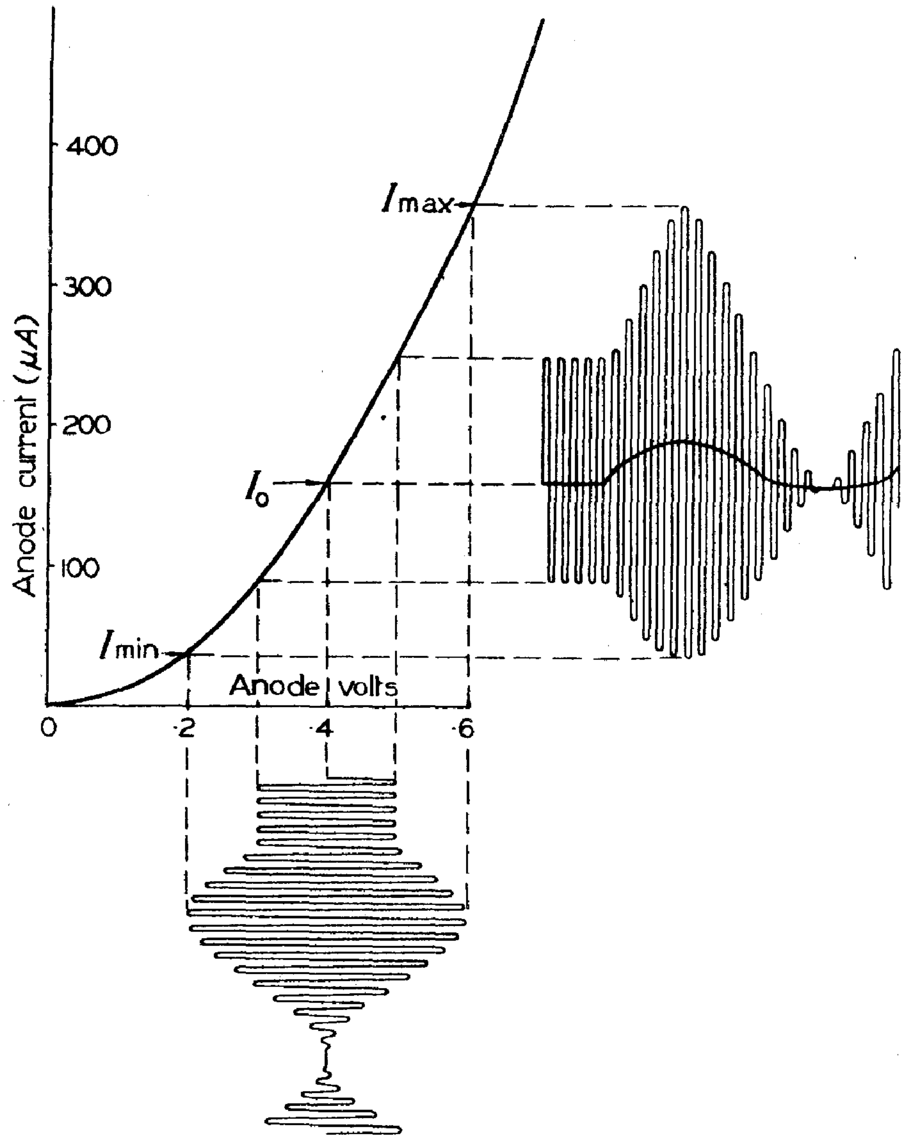


FIG. 32, CHAP. XII.—Square law rectification of modulated wave.

which the wave form of the audio-frequency component is shown by a heavy line; its shape obviously differs from that of the modulation envelope of the applied voltage, i.e. amplitude distortion has taken place.

(ii) The degree to which distortion is introduced is easily calculated from the known equation of the diode characteristic.

Let
$$i_a = g\, v_a^2$$
and
$$v_a = E_a + \mathcal{V}_r\,(1 + \sin \omega_a t)\sin \omega_r t$$

Here E_a is a steady anode voltage and $\mathcal{V}_r\,(1 + \sin \omega_a t)\sin \omega_r t$ a radio-frequency voltage modulated to a depth of 100 per cent. Then

$$\begin{aligned} i_a &= g\,\{E_a + \mathcal{V}_r \sin \omega_r t\,(1 + \sin \omega_a t)\}^2 \\ &= g\,\{E_a^2 + 2\,E_a\,\mathcal{V}_r \sin \omega_r t + \mathcal{V}_r^2 \sin^2 \omega_r t \\ &\quad + [2\,E_a\,\mathcal{V}_r \sin \omega_r t \sin \omega_a t] \\ &\quad + [2\,\mathcal{V}_r^2 \sin^2 \omega_r t \sin \omega_a t] \\ &\quad + [\,\mathcal{V}_r^2 \sin^2 \omega_r t \sin^2 \omega_a t]\,\} \end{aligned}$$

Only the terms containing $\omega_a t$ are able to contribute to the audio-frequency output and these are enclosed in square brackets in the final equation. They are

(*a*) $2E_a\,\mathcal{V}_r \sin \omega_r t \sin \omega_a t$.

This is equivalent to

$$E_a\,\mathcal{V}_r\,\{\cos(\omega_r - \omega_a)\,t - \cos(\omega_r + \omega_a)\,t\}$$

and represents anode current components corresponding to the side-band frequencies.

(*b*) $2\,\mathcal{V}_r^2 \sin^2 \omega_r t \sin \omega_a t$.

This can be written

$$\mathcal{V}_r^2\,(1 - \cos 2\,\omega_r t)\sin \omega_a t$$

and yields a component of frequency $\frac{\omega_a}{2\pi}$ and amplitude $g\,\mathcal{V}_r^2$ together with other components of frequency $\frac{2\omega_r \pm \omega_a}{2\pi}$, which are of no immediate interest.

(*c*) $\mathcal{V}_r^2 \sin^2 \omega_r t \sin^2 \omega_a t$.

This is equivalent to

$$\begin{aligned} &\frac{\mathcal{V}_r^2}{4}\left(1 - \cos 2\omega_r t\right)\left(1 - \cos 2\omega_a t\right) \\ &= \frac{\mathcal{V}_r^2}{4}\left\{1 - \cos 2\omega_r t - \cos 2\omega_a t + \cos 2\omega_a t \cos 2\omega_r t\right\} \end{aligned}$$

yielding certain radio-frequency components, together with an audio-frequency component of amplitude $g\,\frac{\mathcal{V}_r^2}{4}$ and frequency $\frac{2\,\omega_a}{2\pi}$. The latter component, being of twice the frequency of modulation, is referred to as second harmonic distortion.

(iii) The total audio-frequency variation of anode current, to which the telephones will respond, is therefore

$$i_a' = g\left\{\mathcal{V}_r^2 \sin \omega_a t - \frac{\mathcal{V}_r^2}{4}\cos 2\omega_a t\right\},$$

and the square law rectification of a completely modulated wave is shown to result in the introduction of second harmonic distortion, the amplitude of the harmonic being 25 per cent. of the fundamental component. For any other depth of modulation, K, the percentage of second harmonic distortion is $\frac{K}{4} \times 100$. The mean increase of anode current during rectification is also equal to $g\,\frac{K\,\gamma_r^2}{4}$ so that if the mean increase is found the amplitude of the second harmonic is known. Applying the method given in Chapter IX, therefore, the percentage of second harmonic distortion is found from the formula.

$$\left.\begin{array}{r}\textit{Percentage of second}\\ \textit{harmonic distortion}\end{array}\right\} = \frac{\frac{1}{2}\,(I_{\text{max.}} + I_{\text{min.}}) - I_0}{I_{\text{max.}} - I_{\text{min.}}} \times 100$$

the notation used being shown in fig. 32.

This formula may be used to estimate the second harmonic distortion introduced by the audio-frequency stages, even though the valve characteristic is not truly parabolic. It is generally accepted that if the second harmonic distortion is less than 5 per cent. it is hardly perceptible to the ear, that in musical reproduction, up to 10 per cent. may be tolerated, and that for commercial or service telephony up to 15 or 20 per cent. is permissible.

Triode rectification—general

50. For a small input grid swing, say up to about ·5 volt, a triode operates approximately as a square law rectifier, no matter whether the curvature of the $I_g - V_g$ or $I_a - V_g$ curve is employed, and the sound output is considerably distorted unless the average depth of modulation is low. When possible, however, the detector of an R/T receiver is operated with a grid swing of at least 3 volts (unmodulated carrier), i.e. an input voltage of about 1 volt R.M.S. Under these conditions, provided that the depth of modulation does not exceed about 80 per cent., the audio-frequency output current is practically proportional to the input voltage and the rectifier is said to be linear.

Anode circuit linear rectification

51. (i) The dynamic characteristic of a typical detector valve, with an anode load of 100,000 ohms, is given in fig. 33. Assuming that provision is made for operation with a carrier amplitude not exceeding 4·5 volts, and 100 per cent. modulation, the H.T. voltage in this particular case has been so chosen that the anode current cut-off occurs at a grid voltage of —9 volts, and the mean bias is adjusted to this value in the usual manner. The input voltage shown has a carrier amplitude of 4·5 volts but is modulated to a depth of only 80 per cent. The variation of anode current consequently occurs over that portion of the characteristic which is practically straight; a rectifier operated under conditions such as these is usually referred to as a linear anode bend detector. The envelope of the anode current variation is to all intents and purposes a copy of the envelope of the input voltage, while the audio-frequency component of anode current, shown in heavy line, closely follows the wave form of the original sound wave.

(ii) Under actual operating conditions, the anode circuit load resistance is shunted by a condenser having a comparatively low reactance at the carrier and side-band frequencies. If this condenser is omitted, the detector valve will impose a heavy damping upon the input circuit owing to the miller effect. On the other hand, if its reactance at the carrier frequency is very small compared with the anode A.C. resistance of the valve, the damping due to the valve will be negligible, but the radio-frequency variations of anode current will correspond with the static and not with the dynamic characteristic. Since the former has a greater curvature than the latter the distortion will be more pronounced. This condenser is, in effect, the reservoir condenser of the rectifier, and the determination of its capacitance C_0 is subject to the somewhat conflicting considerations mentioned above.

52. Stated more fully, these considerations are :—

(i) the reactance, X_o, of the reservoir condenser at the highest audio-frequency must be very much larger than the load resistance, R_o,

(ii) its reactance at carrier frequency must be small compared with the load resistance,

(iii) the time constant, C_oR_o of the anode circuit must be small in order that the output may follow the variation of input voltage as closely as possible. As R_o must be large in order to give the desirable linearity of dynamic characteristic, this requirement calls for a small value of C_o.

(iv) the input admittance of the valve must be as low as possible.

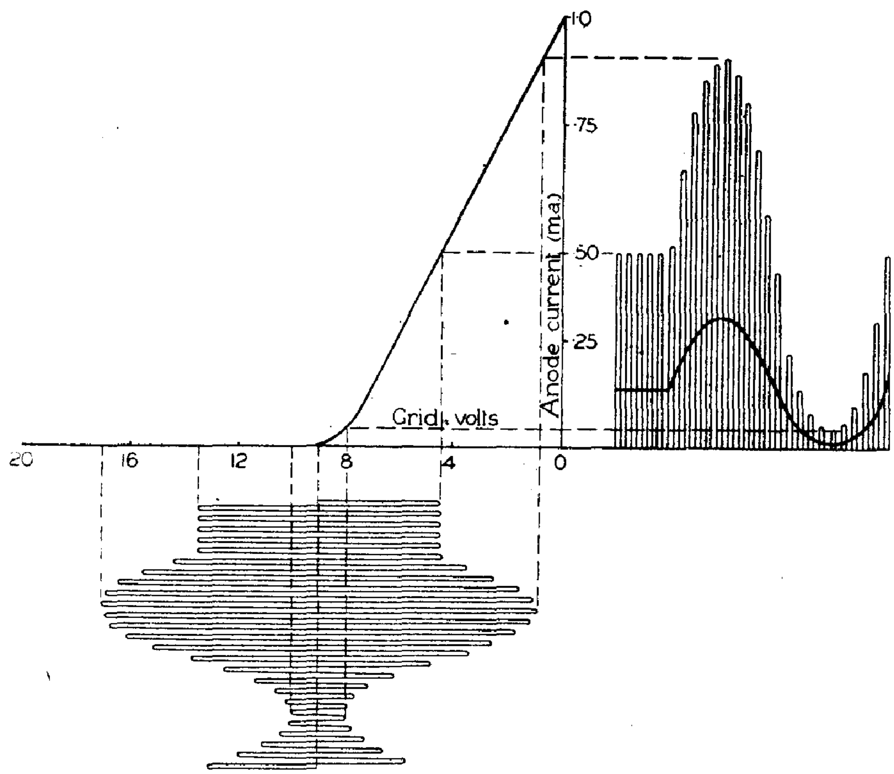


FIG. 33, CHAP. XII.—Anode circuit linear rectification.

Applying these considerations to a practical example, let the highest audio-frequency be 8,000 cycles per second ($\omega = 5 \times 10^4$) and let us assume that to meet requirement (i) a ratio $\frac{X_o}{R_o} = 5$ will be satisfactory. Then if $R_o = 100{,}000$ ohms, $X_o = 5 \times 10^5$,

$$\frac{1}{\omega C_o} = 5 \times 10^5$$

$$C_o = \frac{1}{5 \times 10^4 \times 5 \times 10^5} \text{ farad.}$$

$$= \cdot 00004 \ \mu\text{F.}$$

To meet requirement (ii) the reactance of the condenser C_o, at the carrier frequency, must be small compared to R_o. For example, if the carrier frequency be 796 kc/s ($\omega = 5 \times 10^6$) then the reactance of a capacitance of $\cdot 00004 \ \mu F$ is 5,000 ohms which is satisfactory. The time constant will be $C_oR_o = \cdot 00004 \times 10^{-6} \times 10^5 = \cdot 000004$ second. Since this time is not very much greater than the duration of one radio-frequency cycle, the envelope of the anode current variation will follow that of the grid voltage very closely and requirement (iii) is satisfied. The effect of the time constant is shown qualitatively in fig. 34. For simplicity, the grid-filament voltage is assumed to vary in amplitude in an abrupt manner (fig. 34a). If the time constant is infinitely small, the anode current will vary in such a manner that the telephone current will reproduce exactly the envelope of the grid filament voltage, taking the form indicated in fig. 34b. With a finite time constant, the anode current will not vary in precisely the same manner as the envelope of the grid-filament voltage. Provided C_0R_0 is comparatively small, the anode current will vary somewhat as in fig. 34c, while a further increase in C_0R_0 will result in the anode current variation shown in fig. 34d. For a given load resistance R_0, therefore, it is obvious that the value of C_0 must be a compromise between the values indicated by requirements (i) and (iv) since the latter indicates the largest practicable value of C_0.

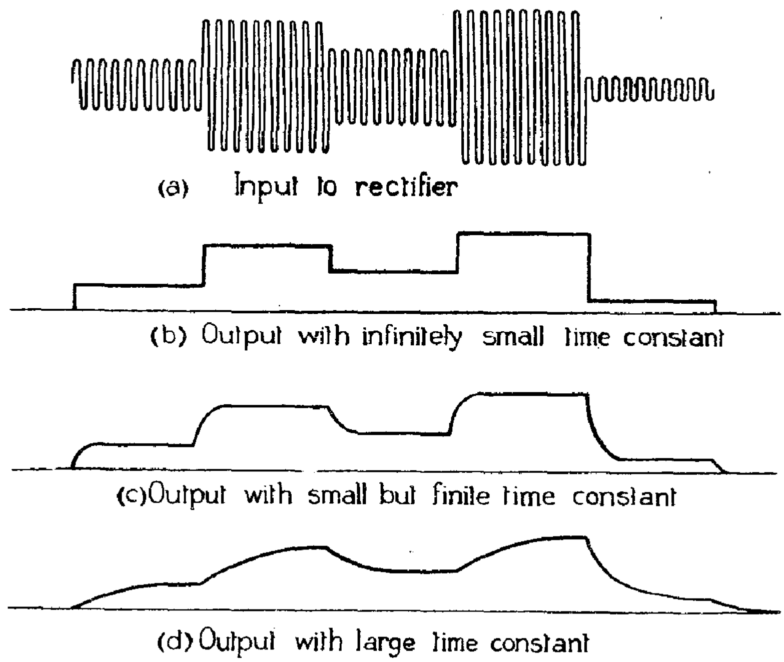


FIG. 34, CHAP. XII.—Effect of detector time constant.

53. The operating potentials of the detector valve are specified on the assumption that a definite grid swing is available. The radio-frequency stages preceding the detector must be designed to give this input voltage at the detector, with a given field strength, and the appropriate aerial. Thus if the amplitude of the grid-filament voltage at the detector is to be 4·5 volts, and the appropriate aerial has an effective height of 2 metres, while the receiver is to give the specified output with a field strength of 45 micro-volts per metre, the total voltage gain of the radio-frequency stages must be $\frac{4 \cdot 5}{90 \times 10^{-6}} = 5 \times 10^4$. Three stages, each having a gain of about 37, are therefore required, or assuming that a circuit magnification of 20 may be achieved before applying the signal to the first valve, two stages each giving a voltage amplification of 50. This gain can easily be realized at frequencies below about 1,000 kc/s. Signals of considerably lower field strength will also be received, though at reduced strength. If however the field strength is sufficiently great to give a detector input greatly exceeding 4·5 volts (with the valve represented in fig. 33), the detector will be overloaded and will give rise to distortion. The remedy for this is to provide some form of gain control in the radio-frequency amplifier stages and thus limit the detector input to that for which it is designed.

Grid circuit (linear) rectification

54. This method of detection is used to a greater extent than anode circuit rectification because, for a given input, it is capable of delivering a somewhat greater output with less distortion. The circuit used is identical with that used for the cumulative grid rectification of C.W. or I.C.W. signals, and is given in Chapter IX. The values of grid condenser and grid leak resistance are, however, primarily chosen with a view to avoiding distortion, rather than for maximum signal strength. The detection of a modulated signal is very similar to that of a C.W. signal with separate heterodyne and may be explained with reference to figs. 35 and 36. The action in the grid circuit is shown in fig. 35; the input signal voltage is assumed to be similar to that of fig. 33, and its original envelope is shown in heavy line. If the grid were connected

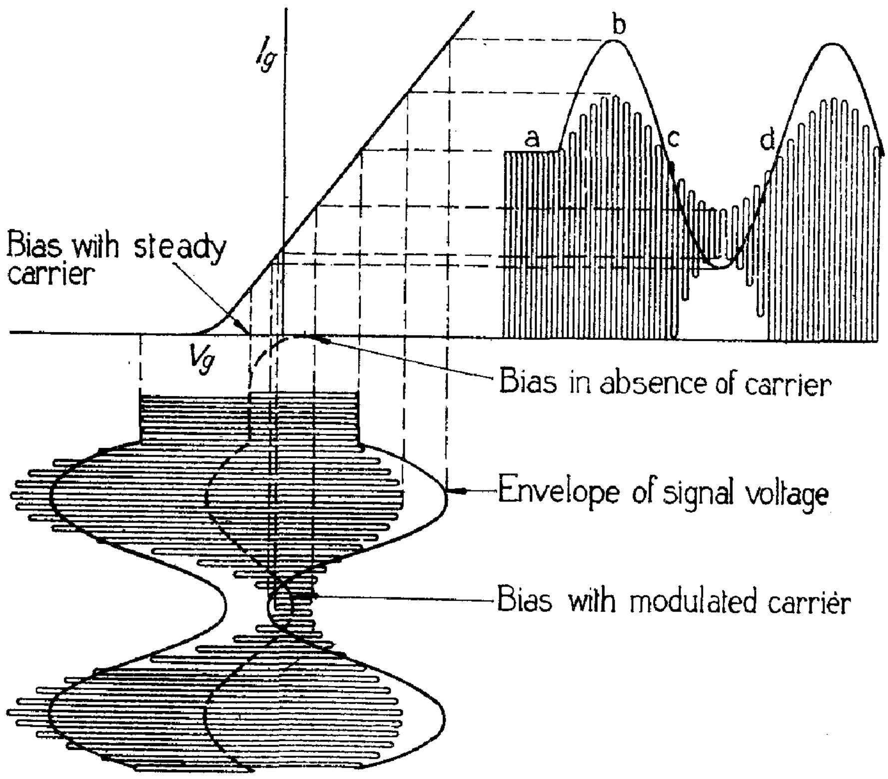


FIG. 35, CHAP. XII.—Grid circuit linear rectification; action at grid.

directly to the filament, the envelope of the resulting variation of grid current would be as indicated by the curve *a, b, c, d*, Since, however, the grid current must return to the filament via the grid leak resistance, the grid voltage varies in a manner corresponding with the changes of anode current. The actual variations of radio-frequency input voltage and grid current are therefore as shown in light line, and the audio-frequency variation in grid bias in dotted line. The latter has a wave-form similar to that of the original modulation envelope. The effect of the total variation of grid voltage upon the anode current is shown in the following diagram (fig. 36). The anode current varies at the radio frequency, and its mean value also varies in accordance with the mean variation of grid bias, i.e. at audio frequency. The radio-frequency variations are by-passed by a suitable condenser, and are not of immediate concern. The audio-frequency variation may be caused to operate the telephone receivers either directly or, as is more usual, after one or more stages of audio-frequency amplification.

Values of grid leak and condenser

55. In a grid circuit rectifier the grid condenser is in effect the reservoir condenser of the rectifier and the actual rectifier load is the grid leak resistance. The voltage across the rectifier condenser can only decrease at the rate allowed by the load resistance, and, as in anode circuit rectification, the condenser voltage can follow the modulation envelope only if the time constant is small. The fluctuation of the modulation envelope is obviously more rapid for high audio frequencies than for low, and for deep than for shallow modulation; for any given audio frequency $\frac{\omega_a}{2\pi}$, it can be proved that the grid condenser voltage will closely follow the modulation envelope provided that

$$C_g R_g > \frac{\sqrt{1 - K^2}}{\omega_a K}$$

The time constant $C_g R_g$ may of course be reduced by a reduction of either or both of its factors, but C_g must be large in order to avoid attenuation of the radio-frequency signal, while C_g should be small and R_g large in order to reduce the damping imposed upon the input circuit by the grid leak, and by the input resistance of the valve itself, taking the Miller effect into account. In practice the capacitance of the grid condenser is first decided, say from five to ten times the input capacitance of the valve, e.g. $C_g = 150$ to $300\ \mu\mu F$, and the grid leak is then given such a

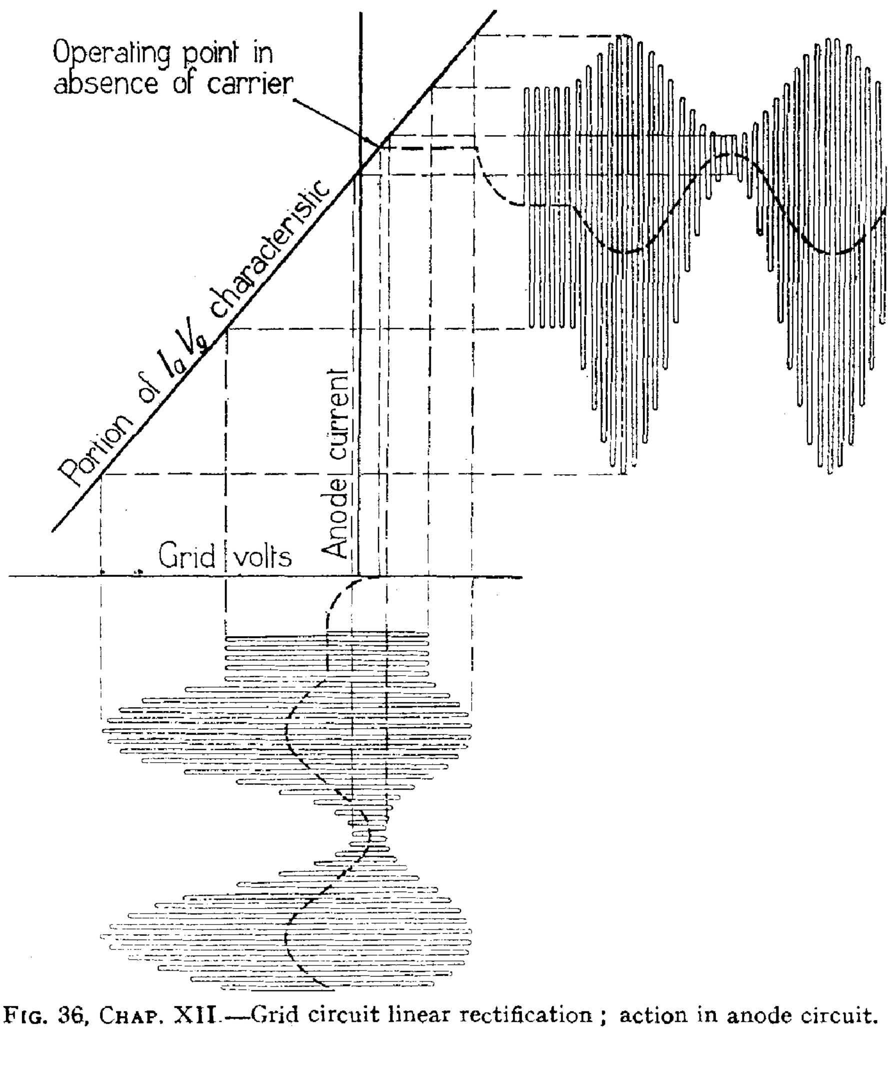


FIG. 36, CHAP. XII.—Grid circuit linear rectification; action in anode circuit.

value that $C_g R_g$ is of the order of 20 microseconds. The grid leak resistance is therefore of the order of only ·1 megohm, which is considerably less than is used for the detection of C.W. or I.C.W. signals. The signal strength is of course somewhat less than with a leak of higher resistance.

56. The maximum permissible input voltage of the grid circuit rectifier is limited by the fact that unless the total anode current variation is confined to the straight portion of the $I_a - V_g$ curve, rectification will take place in the anode circuit and will give rise to distortion. Since the radio-frequency variation is much larger than the audio-frequency variation, it is impossible to obtain a large undistorted audio-frequency power output. The greatest efficiency is achieved by operating with a high H.T. voltage, e.g. 100 to 150 volts, and since the valve must dissipate the whole of the D.C. input during periods of no modulation, a small power valve is preferably used. Even so the H.T. voltage must be considerably less than when the valve is used for power amplification, since under the latter conditions it is operated with considerable negative bias, while as a detector the bias voltage is rarely more than a fraction of a volt negative and may be slightly positive. Where it is possible to provide an adequate input swing to the detector, it is preferable entirely to separate the functions of detection and audio-frequency amplification.

Linear rectification by diode

57. Although for small input voltages the diode acts as a square law rectifier, it was shown in Chapter X that if the input voltage is not allowed to fall below a certain value, depending upon the particular type of diode, the peak anode current is proportional to the anode-filament P.D. Provided therefore that the average depth of modulation is below about 70 per cent.,

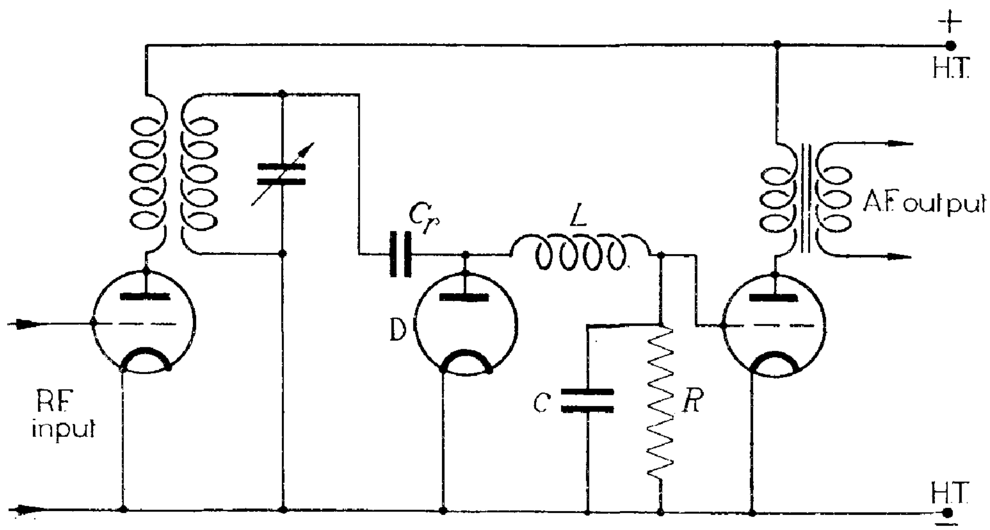


FIG. 37, CHAP. XII.—Diode used as linear rectifier.

and the carrier amplitude is of the order of 10 volts, the rectified output of the diode is practically proportional to the input voltage. In R/T receivers embodying a high degree of radio-frequency amplification, the diode is extensively used as a detector. A typical circuit is given in fig. 37, in which C_r is the reservoir condenser and R the load resistance. The audio-frequency voltages set up across the latter are applied to the grid and filament of the triode which acts as an audio-frequency amplifier. The radio-frequency choke L and condenser C are inserted in order to reduce the amplitude of the radio-frequency voltages applied to the triode. The action of the diode rectifier and the following audio-frequency stage are therefore exactly the same as that of the linear grid rectifier, except that the grid swing on the triode valve is greatly reduced owing to the separation of radio-frequency and audio-frequency functions.

Necessity for radio-frequency gain control

58. It has been shown that in order to reduce distortion to a minimum the input to the detector must lie between certain limits. The radio-frequency amplification required to give this detector input depends upon the minimum field strength upon which the receiver is called upon to operate. The interference level on the receiving site must be taken into account in this respect. For example, in broadcast reception in an industrial centre where the noise level due to electrical devices is very high, programmes of entertainment value can only be received from transmitters giving a field strength of about 10 milli-volts per metre, while in rural areas where electrical interference is negligible a higher degree of amplification can be used and equally good reception obtained from transmitters giving a field strength of only about ·5 milli-volts per metre. Service R/T communication must be performed with transmitters giving a much weaker field than this, in the presence of an interference level comparable with that of an industrial area, in addition to a high level of local (i.e. non-electrical) noise. In cases where the field strength is so low that the receiver must be operated at its maximum sensitivity, a considerable noise level must be accepted, whereas if the initial field strength is high a reduction in overall gain will effect a considerable improvement in the signal-noise ratio. This constitutes one reason for the incorporation of a gain control in the pre-detector stages. The second object of this control is to avoid overloading the detector valve. To take a numerical example, suppose the detector to operate linearly over the range ·5 to 5 volts and the receiver to be used on an aerial having an effective height, h, of 5 metres and a magnification χ, between aerial proper and the grid of the first valve, of 5. If the sensitivity of the receiver, i.e. the minimum field strength, $\hat{\Gamma}$, which will give ·5 volt input to the detector, is 4 micro-volts per metre, the input voltage to the first valve must be $\chi h \hat{\Gamma}$ volts, or $5 \times 5 \times 4 \times 10^{-6}$ volts. The amplifier is therefore required to give a gain of 5,000, which is easily attained by two tuned stages using screen-grid valves or radio-frequency pentodes. The receiver will only rarely be required to operate at its maximum sensitivity, but may often be called upon to deal with field strengths as high, say 5 milli-volts per metre. The maximum permissible detector input, 5 volts, can then be obtained with a voltage gain of 40, and if this is exceeded the detector will be overloaded. Hence it is necessary to provide a smooth control of the voltage gain between these limits.

Methods of gain control

59. (i) When the radio-frequency stages embody screen-grid valves or radio-frequency pentodes, a certain amount of control may be obtained by variation of either the grid bias voltage or the screen potential. The first method operates by virtue of the fact that the curvature of the $I_a - V_g$ characteristic is considerably greater than that of the triode, i.e. the mutual conductance g_m varies with the mean grid bias, as shown in fig. 38, which is the $g_m - V_g$ curve of a typical radio-frequency pentode. Since for a given dynamic resistance R_d the voltage gain is approximately equal to $R_d\ g_m$, it is easy to find the variation in gain for various bias voltages, e.g. if the anode load is 50,000 ohms, the gain varies from 80, for values of bias in the region of $-$·5 volt, to about 25 when the bias is -3 volts. At the latter point, however, the mean operating point is situated very low down upon the $I_a - V_g$ curve where the curvature is very pronounced, the curvature being proportional to the slope of the $g_m - V_g$ curve. The objection to this method of control is now easily seen. The larger the input voltage is, the greater is the necessity for a long straight portion of the $I_a - V_g$ curve, whereas an increase of negative bias shifts the operating point into a region where the curvature is more pronounced. The result is that although the amplification is reduced, the envelope of the anode current variation is considerably distorted. In addition, cross-modulation will be introduced if an interfering signal is also present. The latter phenomenon will be dealt with later.

(ii) Control of amplification by variation of screen potential suffers from similar disadvantages. A reduction of screen potential reduces the slope of the $I_a - V_g$ curve and therefore reduces the amplification, but the grid base line available for the input voltage swing is reduced in a corresponding degree, the curve moving to the right with a decrease of screen voltage.

Thus the operating point is shifted into the curved portion of the characteristic as before. In addition, if the input swing is increased without a corresponding increase of negative grid bias, grid current may flow during a portion of each positive half-cycle.

60. With either of the above methods, it is generally found that operation of the gain control affects the tuning of the receiver to some extent. This raises an important point in practical manipulation. The standard practice is to tune aircraft receivers to the desired frequency on the ground, by means of a suitable heterodyne wavemeter. If the latter is placed very near the receiver, the operator will first pick up the emission (modulated C.W.) with the gain control at maximum sensitivity, and may then proceed to reduce the gain and trim up the tuning simultaneously. This procedure is incorrect, for on returning the gain control to the position of maximum sensitivity the receiver will be slightly de-tuned, and may be incapable of dealing with a very weak signal. The correct method is to pick up the signal with maximum gain, and progressively to move the wavemeter away from the receiver while trimming up,

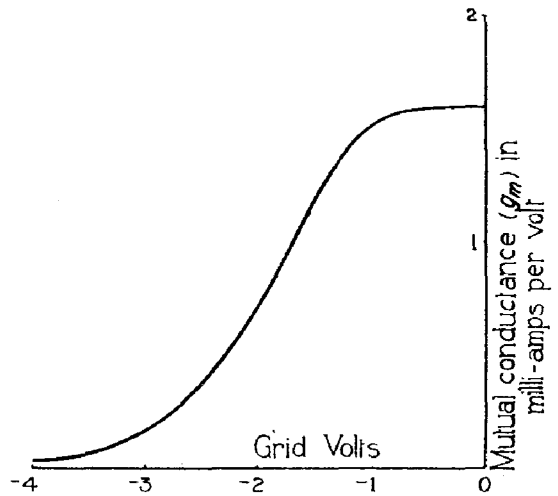


FIG. 38, CHAP. XII.—g_m — V_g curve of R.F. pentode.

operating always with maximum gain. The receiver will then be taken into the air fully capable of dealing with the weakest signals, and will be slightly de-tuned only when very strong signals are being handled.

Cross-modulation

61. Cross-modulation is the name given to a form of interference peculiar to radio-telephony, and will be described with reference to a receiver using screen-grid valves in the radio-frequency stages. If the receiver is tuned to a station giving a high carrier level, and the gain reduced by negative grid bias, the operating point is situated in the curved region of the I_a — V_g characteristic. Any other transmitter which is operating on an adjacent frequency channel may give an appreciable input to the first valve, and the instantaneous grid swing on the latter is then equal to the sum of the two swings. Owing to the anode bend rectification which occurs, the output at the desired carrier frequency will possess side-bands corresponding to the modulation of both signals, and no amount of selectivity in the succeeding circuits will discriminate against this spurious modulation. The effect can therefore only be avoided by a considerable degree of selectivity between aerial and first valve, or by operating in the region of zero bias and controlling the gain by loose input coupling, " lossing " and screen potential variation. One drastic expedient which is sometimes adopted is to fit a switch which disconnects the aerial entirely, leaving the receiver to operate upon the " pick-up " of the coils and wiring.

62. (i) The manner in which cross-modulation is introduced may be shown by assuming the amplifier valve to possess a characteristic represented by an equation of the form

$$i_a = A + B\,v_g + C\,v_g^2 + D\,v_g^3 \,\ldots \qquad (1)$$

Let $v_g = W \sin \omega_1 t\,(1 + K \sin \alpha t) + U \sin \omega_2 t\,(1 + M \sin \beta t)$,

where W is the amplitude of the " wanted ", U the amplitude of the " unwanted" signal, while ω_1 and ω_2 correspond to radio frequencies which are modulated at frequencies $\frac{\alpha}{2\pi}$ and $\frac{\beta}{2\pi}$ to depths of K and M respectively. The important term in introducing cross-modulation is $D\,v_g^3$.

$$v_g^3 = \{W \sin \omega_1 t\,(1 + K \sin \alpha t) + U \sin \omega_2 t\,(1 + M \sin \beta t)\}^3 \qquad (2)$$

$$= (a + b)^3$$

$$= a^3 + 3a^2b + 3ab^2 + b^3.$$

Expanding the term corresponding to $3ab^2$ we obtain a component of v_g^3, say v_g', where

$$v_g' = 3W \sin \omega_1 t\,(1 + K \sin \alpha t)\,\{U^2 \sin^2 \omega_2 t\,(1 + M \sin \beta t)^2\}. \qquad (3)$$

Since $U^2 \sin^2 \omega_2 t = \frac{U^2}{2}(1 - \cos 2\omega_2 t)$ and $(1 + M \sin \beta t)^2 = 1 + 2M \sin \beta t + M^2 \sin^2 \beta t$, there will be a component of v_g', say v_g'', where

$$v_g'' = 3W \sin \omega_1 t \times \frac{U^2}{2} \times 2M \sin \beta t$$

$$= 3WU^2M \sin \omega_1 t \sin \beta t$$

$$= \frac{3WU^2M}{2}\left\{\cos(\omega_1 - \beta)t - \cos(\omega_1 + \beta)t\right\}, \qquad (4)$$

which represents side-band components associated with the frequency $\frac{\omega_1}{2\pi}$ of the desired signal, but corresponding to the modulation carried by the undesired signal.

(ii) The degree to which cross-modulation is introduced is therefore proportional to the square of the amplitude of the unwanted signal at the grid of the valve and is also proportional to the coefficient of V_g^3, i.e. to D, in equation (1). It follows that the degree of cross-modulation is approximately proportional to the rate of change of curvature of the valve characteristic. This accounts for the increase of cross-modulation which occurs when the operating point is brought on to a curved portion of the characteristic by the application of large negative bias. Cross-modulation is almost entirely eliminated by the use of variable-mu valves in the radio-frequency stages.

Modulation rise

63. This is another phenomenon which occurs when operating upon a markedly curved portion of the characteristic. Under these conditions, the amplification depends to some extent upon the grid swing, being greater for large inputs than for small. It follows that when a modulated voltage is received, the peaks of the modulation envelope receive greater amplification than the troughs, and in the output the ratio of peak to trough is greater than in the input. This effect varies approximately with the square of the applied carrier amplitude, and is also proportional to the curvature of the $g_m - V_g$ curve. It is therefore of much greater importance in ordinary screen-grid valves and R.F. pentodes than in valves of the type described in the following paragraph.

The variable mutual conductance valve

64. This valve is generally referred to as the variable-mu valve. Its object is to give a considerable control of amplification by grid bias variation, without the attendant disadvantages. Both tetrodes and pentodes may be given the desired characteristics, which are obtained by

special design of the control grid. Whereas in ordinary valves the latter is of uniform spacing throughout its length, in the variable-mu valve it is made in two or more sections of varying degrees of spacing. Fig. 39 shows the electrodes of a variable-mu screen-grid valve in cross-section, and it can be seen that the middle portion of the control grid is of comparatively open mesh. It will therefore exercise a smaller degree of control than the upper and lower portions, which are of fairly close mesh. Even if the grid is given a considerably negative potential, therefore, an appreciable anode current will flow, and the $I_a - V_g$ curve tails off slowly with increasing negative bias, instead of possessing a comparatively sharp cut-off. In fig. 40 the $I_a - V_g$ curves of a variable-mu screen-grid valve (full line) and an ordinary screen-grid valve (dotted line) may be compared. The latter can only accommodate, without curvature distortion, a grid swing of about 1·5 volt; to do this the grid bias has a fairly critical value of about $-0·75$ volt. When given a negative bias of about 15 volts, however, the variable-mu valve will accommodate a grid swing of about 10 volts, the characteristic being practically straight over this range. As the grid bias decreases, the length of the straight portion available for distortionless amplification also decreases. For an input swing of 1·5 volts, the bias may be only about $-0·75$ volt, as in the screen-grid valve.

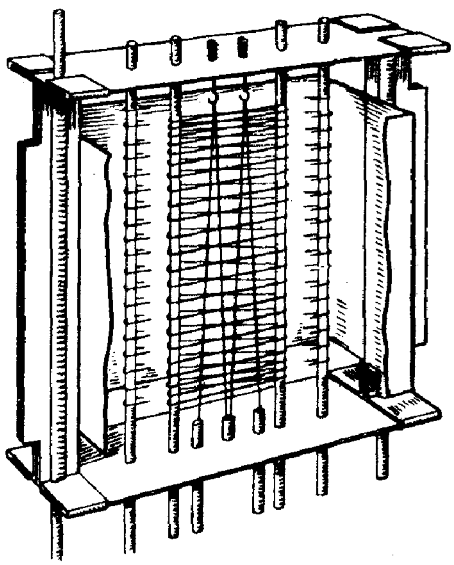


FIG. 39, CHAP. XII.—Electrodes of variable-mu screen grid valve.

65. The small inset diagram in fig. 40 shows the variation of mutual conductance with grid bias and may be compared with the $g_m - V_g$ curve of a screen-grid valve (fig. 38). The mutual conductance is seen to be, approximately, inversely proportional to the bias voltage. This is the feature which makes the valve so useful for gain control, for if the bias voltage is varied in such a manner that it is always proportional to the input voltage, the output voltage will remain constant. In modern R/T receivers, therefore, the magnitude of the bias voltage in the radio-frequency stages is often made to depend upon the amplitude of the received carrier. Such a receiver is said to be fitted with automatic volume control (A.V.C.).

Automatic volume control

66. In receivers fitted with automatic volume control, a portion of the carrier frequency input to the detector (or second detector in a super-heterodyne receiver) is rectified, and the resulting current passed through a resistance which is so arranged that an increase in rectified current, i.e. an increase in carrier amplitude, applies a corresponding negative bias to the gain-controlled valve or valves. Considerable radio-frequency amplification must be available, in order that the rectified current in the resistance will be sufficient to provide the maximum bias voltage called for. For this reason, A.V.C. is rarely found in other than super-heterodyne receivers.

For simplicity, however, the principle is illustrated in fig. 41 as it would be applied in a simple receiver comprising one radio-frequency amplifying stage and a linear grid rectifier. The first valve (V.S.G.) is a variable-mu S.G. valve and the second (T) a triode. The latter is the signal rectifier ; its input voltage is also applied to the A.V.C. amplifier valve (S.G.) the output of which is applied to the diode (D). This operates as a half-wave rectifier and establishes a P.D. across the reservoir condenser C_1 and resistance R_1, the upper plate of the condenser being at negative potential with respect to the filament line (L.T. negative).

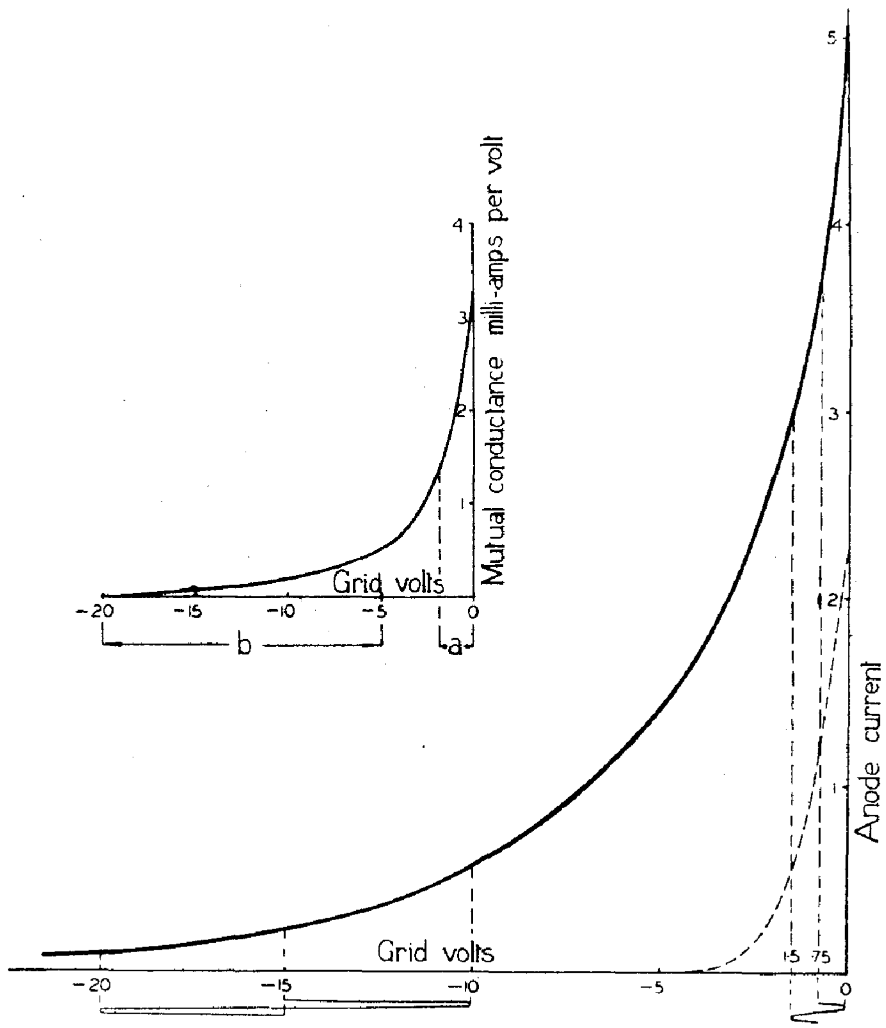


FIG. 40, CHAP. XII.—Characteristic curve of variable-mu screen-grid valve.

This P.D. is applied to the grid of the first valve via the resistance R_2, which is therefore biased to an extent depending upon the amplitude of the input to the diode. The desired effect would not be completely attained by the system so far described, because the A.V.C. would operate to some extent upon all signals, instead of only upon those exceeding a certain amplitude. This can be overcome by applying a suitable bias to the anode of the diode D, as indicated by the inclusion of the battery E in the diagram. In practice it is not convenient to insert a battery at this point, and the necessary delay voltage, as it is called, is obtained by means of a tapping on a resistance connected between the negative L.T. and negative H.T. terminals. (*See* paragraph 76).

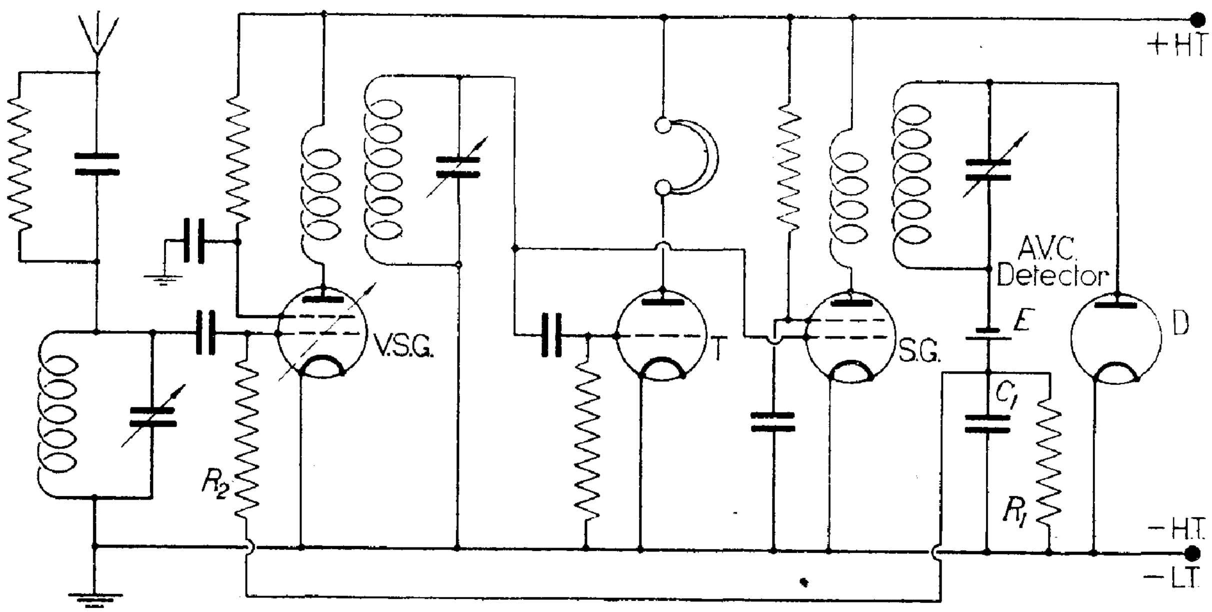


FIG. 41, CHAP. XII.—R/T receiver with A.V.C.

Audio-frequency amplification

67. After detection, it is generally desirable to utilize one or more stages of amplification, the final valve operating as a power amplifier and supplying the reproducing instrument—telephone receivers or loud speaker as the case may be. For voltage amplification, resistance-capacitance coupling is preferred for reasons stated in Chapter XI. An electrical output of the order of 150 milliwatts from the final power amplifier is sufficient to give good speech in two pairs of telephone receivers connected in parallel. This may be obtained from a small power valve, i.e. one capable of dissipating from 500 to 1,000 milliwatts, with an input swing of about 10 volts. A small loud speaker may also be operated in this way, but a high sound level cannot be expected. It is sometimes necessary to obtain a greater power output than can be given by an ordinary power valve. When this is so, several possibilities present themselves. They are the employment of

(i) A super-power output triode, i.e. a triode capable of dissipating one or more watts.
(ii) Two or more ordinary power valves in parallel.
(iii) A pentode.
(iv) Power triodes or pentodes in push-pull.

The super-power triode

68. This valve differs from the ordinary output triode in that it is designed to accept a greater grid swing. Its anode A.C. resistance is usually low, from 1,000 to 2,000 ohms, and its mutual conductance normal, e.g. about 2·5 milliamperes per volt. The $I_a - V_a$ characteristics of a typical valve of this kind are given in fig. 42. This valve has an anode A.C. resistance of about 1,600 ohms and its mutual conductance is 3 milliamperes per volt, the amplification factor being 5. In the diagram the load line is that of a dynamic resistance of 3,475 ohms, which is approximately twice the anode A.C. resistance. The permissible dissipation is 3 watts and the operating potentials are $E_a = 145$ volts, $E_g = -17{\cdot}5$ volts. Without entering the regions of anode current curvature and grid current flow, the permissible grid swing is 35 volts, and the maximum undistorted output, as shown by the shaded area, is 485 milliwatts. The efficiency is therefore only 16 per cent. For comparison the area corresponding to the output obtained with a grid swing of 11 volts is also shown in vertical shading; this is equal to 50

milliwatts. A small power valve with a grid swing of this order will give an output of about 100 to 150 milliwatts, thus, unless the super-power valve is supplied with an ample input voltage, the power output obtainable may be less than that given by a small power valve. The figure of merit (i.e. $\frac{\mu^2}{r_a}$) for the valve under discussion, is only ·015 against ·038 for the valve, V.R.22.

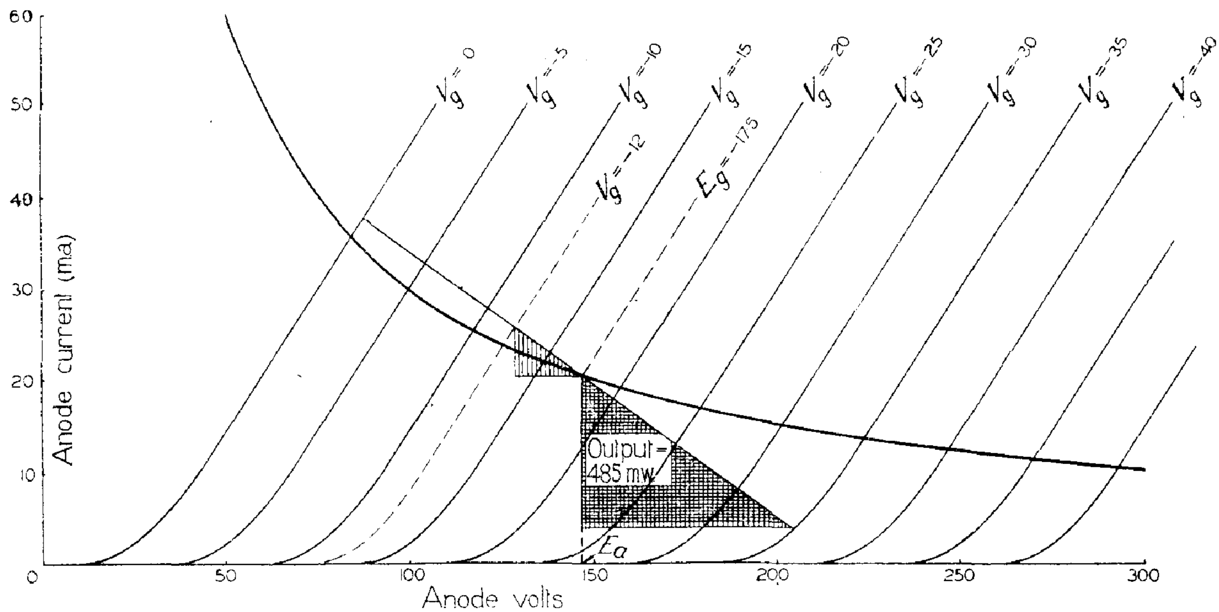


FIG. 42, CHAP. XII.—I_a — V_a curves of super-power triode.

Two power valves in parallel

69. (i) With this arrangement the amplification factor is unchanged, but the effective anode A.C. resistance is only one-half that of a single valve; it must be understood that only valves of the same type may be connected in parallel. The optimum load for maximum undistorted output is equal to r_a and the maximum grid swing that of a single valve. The anode current consumption will be doubled, as will also the power taken from H.T. supply, hence this method is unsuitable for use where only dry cell or inert batteries are available. The power output is twice that obtainable from one valve of the same type, if the optimum loading conditions are achieved in each case.

(ii) Provided that the required grid swing is available, therefore, it is preferable to use a super-power valve rather than two power valves in parallel. Both arrangements suffer from the following disadvantages :—

(*a*) A large steady anode current flows in the output choke or primary winding of output transformer, and unless specially designed the core may be magnetically saturated. Even if this is not so, the high flux density will cause heavy hysteresis loss ; this loss varies at different instants during each cycle, and thus gives rise to amplitude distortion, while the reduction of inductance due to the fall of incremental permeability with increase of flux density causes additional frequency distortion.

(b) Ordinary H.T. batteries are incapable of supplying this anode current. Either super-capacity batteries, accumulator batteries, or mains supply may be used.

(*c*) A large alternating component of anode current flows through the H.T. battery (or other supply device). Earlier stages of the receiver must therefore be very thoroughly decoupled, otherwise low-frequency oscillations may be established. It is sometimes possible to prevent this by reversing the connections to one winding of an intervalve or output transformer, but this is not the best practice, for the energy transfer which previously maintained the oscillation will then impose considerable damping and consequent reduction in overall amplification.

Use of pentode valve

70. The pentode is a five electrode valve, and its characteristics have already been discussed in Chapter VIII. It is chiefly used when greater power output is required than is obtainable from a power valve for the same input grid swing, without serious increase of either H.T. voltage or anode current. Considerable care is necessary in the choice of correct load impedance. Fig. 43 shows the I_a — V_a curves of a pentode, with load lines AB, CD, EF, representing anode circuit resistance loads of approximately 40,000 ohms, 13,700 and 4,000 ohms respectively. The maximum permissible input voltage with this particular valve has been taken as 5 volts (peak), and the working H.T. voltage as 150 volts, the anode current in the absence of an applied input voltage being 10·5 milliamperes. Taking the 40,000 ohm load, an input of 5 volts (peak)

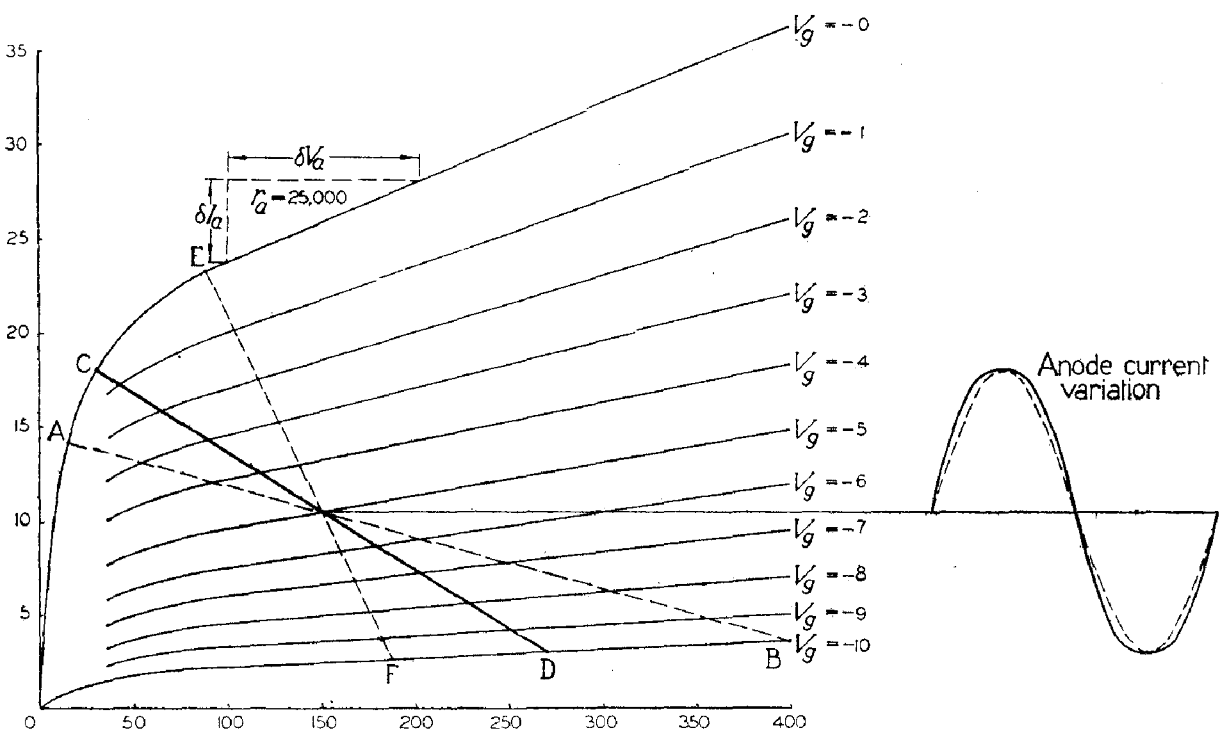


FIG. 43, CHAP. XII.—I_a — V_a curves of output pentode.

will cause the following variations of anode current and anode-filament P.D., namely (i) at the positive peak of input voltage the anode current will rise to 13·7 milliamperes and the anode-filament potential will fall to 12 volts, (ii) at the negative peak of input voltage the anode current will fall to 5 milliamperes and the anode-filament P.D. will rise to about 400 volts. As the variations of current and voltage on the positive and negative half-cycles are quite dissimilar, distortion will obviously occur. With an anode circuit load of this magnitude, another undesirable effect will also arise ; when the anode-filament P.D. falls as low as 12 volts, the flow of electrons between filament and anode will be retarded and a negative space charge will form in the vicinity of the anti-secondary or suppressor grid ; the control grid then ceases to effect any control upon the anode current. The combination of these two effects will inevitably give rise to severe distortion.

71. (i) If the load resistance is too low, the variations of anode current and anode-filament P.D. during the positive and negative half-cycles, e.g. on the load line EF, will again be unsymmetrical and distortion will occur. With a suitably chosen resistance, however, the excursions of both anode current and anode-filament P.D. will be nearly symmetrical about the mean operating point. Thus with the 13,700 ohm load, the anode current increases and decreases by 8 milliamperes as the grid voltage undergoes a complete cycle, the variation of anode-filament P.D. being also symmetrical, and the output will be comparatively free from distortion. It will also be observed that the minimum anode-filament P.D. is about 35 volts, which is sufficiently high to prevent the negative space charge effect at the anti-secondary grid.

(ii) The anode A.C. resistance of a pentode varies greatly with grid potential; if specified, it is usually taken as the ratio $\frac{dV_a}{dI_a}$ measured at the operating H.T. voltage on the curve corresponding to $V_g = 0$. In the diagram the value of r_a at this point is approximately 25,000 ohms, i.e. twice the value of load resistance giving maximum undistorted output. In general, the optimum load of the pentode may be taken as one-half the anode A.C. resistance as defined above, or alternatively as being equal to the D.C. resistance of the valve at the operating point. In fig. 43 this is $\frac{150 \text{ volts}}{10{\cdot}5 \text{ milliamperes}} = 14{,}300$ ohms. In this respect there is a marked difference between optimum conditions for pentode and triode valves, and one cannot be substituted for the other without complete re-design of the output stage.

(iii) On the right of fig. 43 is shown, in solid line, the anode current variation caused by a sinusoidal variation of grid voltage when operating with the 13,700 ohm load, together with a sine curve (dotted line) for comparison. Since the wave-form differs from the sine curve in the same manner in both half-cycles, odd harmonics must be present; as a rule the third harmonic is most conspicuous. For the same percentage distortion, this is more objectionable than the second harmonic, which is exactly one octave higher than the fundamental. The small scale of the drawing does not admit of very accurate computation, but assuming that only the third harmonic is present, the distortion is about 7 per cent.

Valves in push-pull

72. (i) The use of valves in push-pull connection has already been considered in Chapter IX. When used for audio-frequency amplification, the circuit is as shown in fig. 44. The output impedance of the previous stage is the primary of a special form of iron-core transformer having a centre-tapped secondary winding, and one-half of the secondary voltage is applied to grid and

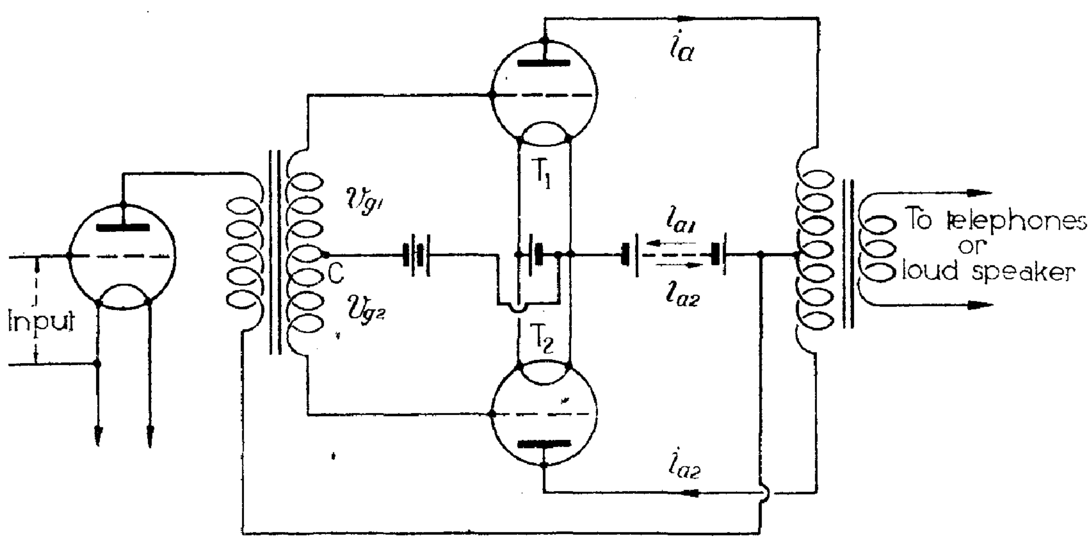


FIG. 44, CHAP. XII.—Push-pull output.

filament of each valve. In the anode circuit of each valve is connected one half of the primary winding of the output transformer, the H.T. supply being fed to both valves via the centre tap. Either Class A or Class B operation may be used. Referring to fig. 44, suppose an alternating voltage to be induced in the secondary winding of the input transformer. Whenever the grid of the valve T_1 is at a positive potential with respect to the centre point C of the winding, the grid of the valve T_2 will be at an equal negative potential, and vice versa. In the absence of a signal voltage, a steady electron current flows from filament to anode of each valve, its value depending upon the H.T. supply voltage and grid bias. Provided the two valves have identical characteristics, these steady anode currents will be equal, and will flow in opposite directions round the core of the output transformer, hence no resultant magnetic flux will be established in the latter. This in itself will remove one possible cause of distortion.

(ii) When a signal voltage is applied to the primary of the input transformer, equal voltages v_{g1}, v_{g2} will be applied to the grids of the two valves, in antiphase ; the resulting anode currents are i_{a1} and i_{a2}, and these currents flow in the two halves of the output transformer primary. As an increase in the value of i_{a1} is accompanied by a simultaneous decrease in the value of i_{a2} and vice versa the induced E.M.F. in the two halves of the winding will be in phase and of equal amplitude, hence the two voltages may be considered to assist each other in producing an alternating flux in the core and a consequent induced secondary E.M.F. ; on the other hand the alternating currents i_{a1} and i_{a2} are in opposition so far as the common connection between negative of filament and centre tap of output winding is concerned, and provided these currents are equal no alternating current will flow through the H.T. source.

Class A amplification

73. Assuming that the two valves have identical characteristics, the total anode circuit load, often referred to as the " anode to anode " load, should be rather less than $4r_a$, and the grid bias rather more than would be used with a single valve, because the anode current may be allowed to swing to a somewhat lower value without introducing appreciable distortion. The power output of a perfectly balanced Class A amplifier is twice the output obtainable from a single valve (assuming optimum load and equal grid swing in each case). As the push-pull arrangement will accept a rather greater grid swing, it would appear possible to obtain nearly three times the output given by a single valve, for the same degree of distortion, but this is rarely so in practice owing to the difficulty of obtaining two valves with identical characteristics. The advantages of the Class A push-pull amplifier over the previously discussed alternatives are as follows :—

(i) No appreciable variation of battery current during each cycle, and consequently little tendency to transfer energy to earlier stages by battery coupling.

(ii) The absence of a steady magnetizing current in the primary winding of the output transformer, permitting the use of a cheaper and lighter transformer.

(iii) A slight increase in permissible grid swing, compared with that of a single valve or two valves in parallel. The super-power valve however has the advantage in this respect.

Quiescent push-pull

74. This term is frequently used to denote the type of circuit in which the grids are biased to cut-off point, so that in the absence of an alternating input voltage the anode current is negligible. This is correctly termed Class B amplification, but this appellation is generally applied to a special arrangement which is described later. Either triodes or pentodes may be used in quiescent push-pull. In either case the permissible grid swing is approximately double that of a single valve, and theoretically the output should be quadrupled. Since however each half of the output transformer is energized only during alternate half-cycles, the output power is only slightly greater than that obtainable under Class A conditions. Although the second harmonic variations of anode current are in opposition in the load, they are in phase in that

portion of the anode circuit which is common to both valves, i.e. the H.T. battery, and complete decoupling of earlier stages is therefore essential. The principal disadvantage of quiescent push-pull working is the difficulty of securing a perfectly matched pair of valves.

The Class B valve.

75. The form of quiescent push-pull arrangement generally referred to as "Class B" amplification utilizes a special valve which comprises two carefully matched triodes in a single envelope, the filament being common to both electrode systems. Only a small negative bias, say 1·5 volt, is used, some valves being designed to operate with zero bias, and grid current flows during a considerable portion of each cycle. The input impedance is correspondingly low, and it is necessary to provide a power input rather than merely a wattless input voltage as is generally the case. The preceding stage is operated as a power amplifier and is called the driver of the Class B valve, to which it is coupled by an iron-core transformer with centre-tapped secondary. The transformer must be designed to match the input impedance R_i of the Class B valve to the A.C. resistance r_a of the driver valve, its ratio being $\sqrt{\frac{R_i}{r_a}}$, while the resistance of its secondary winding must be low compared to R_i. The output transformer must also be carefully designed in order that its internal resistance and leakage reactance may be small, otherwise considerable distortion arises owing to the variation of effective load at different frequencies. The remarks regarding variation of battery current in quiescent push-pull amplifiers are equally applicable to Class B. Quiescent push-pull and Class B amplification are chiefly used where only dry-cell batteries are available for H.T. supply. Any device having considerable internal impedance, e.g. a so-called H.T. eliminator, can only be used if special steps are taken to ensure that the terminal P.D. does not vary with the load current to any appreciable extent.

The paraphase amplifier

76. In the preceding paragraphs it has been assumed that the input circuit of the push-pull amplifier is perfectly symmetrical. It is difficult to ensure this at the higher audio-frequencies, owing to the difference between the capacitance of each end of the transformer winding with respect to earth. The varying iron losses in this transformer also cause amplitude distortion. For both reasons it would appear advantageous to precede the push-pull output stage by a resistance-loaded voltage amplifier. This is achieved in the paraphase amplifier by means of a phase reversing valve, the connections being given in fig. 45. Here T_1 is the amplifying valve following the rectifier; the resistance R_1, grid condenser C_1, and resistance R_2 forming an output network. A portion of the voltage across R_2 is applied to the grid and filament of the phase reversing valve T_2, which has the output network R_3 C_2 R_4. When the two networks are correctly balanced, equal voltages are developed across the resistances R_2 and R_4, in antiphase, and these are applied to the push-pull (Class A) triodes T_3, T_4. The correct balance is obtained by adjusting the tapping point on the resistance R_2. To do this a sinusoidal input is supplied to the valve T_1 and the operator listens in the telephone receivers which are connected in the common anode lead of the output stage. When a correct balance is obtained there will be no current variation in this circuit and consequently no sound output from the receivers.

77. The circuit diagram of a typical super-heterodyne R/T receiver embodying several of the features discussed in previous paragraphs, is given in fig. 46. An aperiodic aerial coupling supplies the input circuit L_1 C_1 of the frequency changer which is a triode-pentode valve, the pentode section having variable-mu characteristics. The triode section of the frequency changer acts as the oscillator valve in conjunction with the circuit L_4 C_4 and reaction coil L_5. The band-pass filter L_2 C_2 L_3 C_3 is tuned to the intermediate frequency and is not adjustable. The intermediate frequency amplifier valve is a variable-mu radio-frequency pentode. Its output circuit supplies the double-diode detector valve. The anode A_1 rectifies the signal voltage, the output voltage being developed in the circuit R_1 C_5, R_2. The latter resistance is a kind of potentiometer, and is used as a control of the average output sound level, this level being then

maintained automatically. The automatic volume control operates as follows: rectification takes place at the anode A_2 of the second detector, and a voltage proportional to the carrier amplitude is developed across the resistance R_3. This voltage is applied, via the resistances R_4, R_5 to the control grids of both frequency changer and intermediate frequency amplifier valves. The

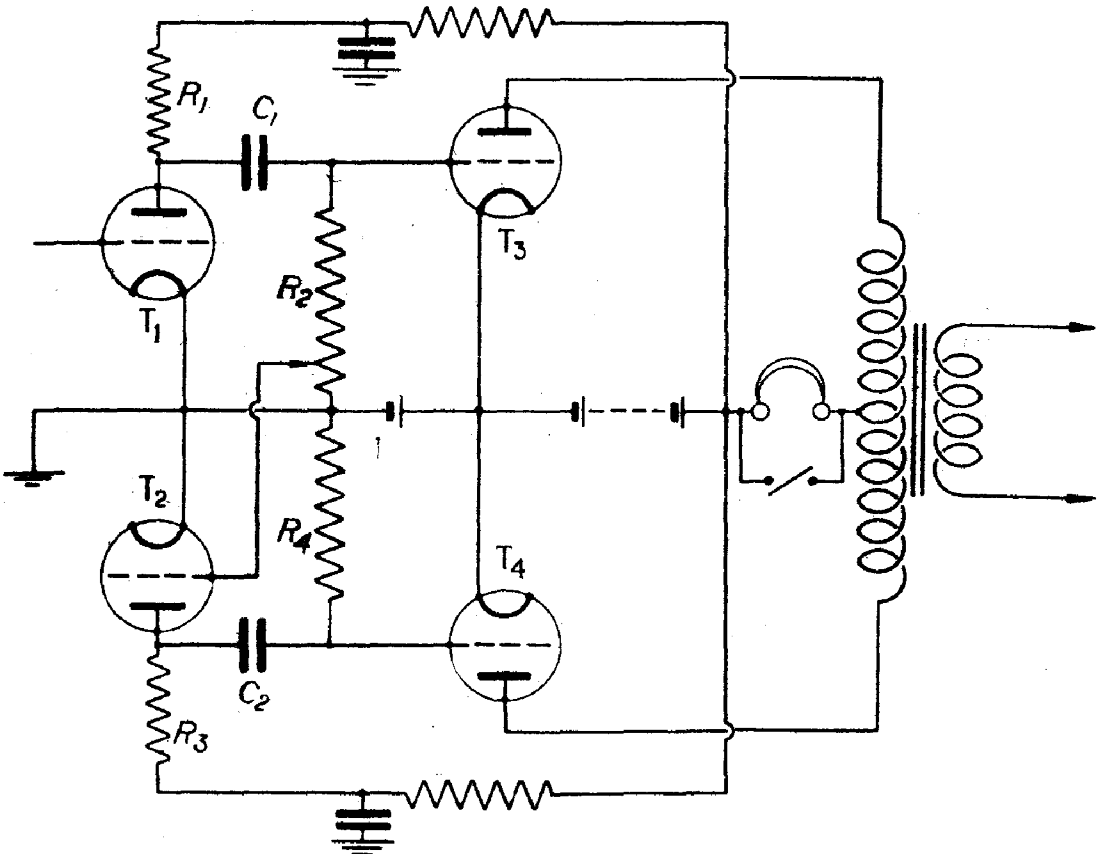


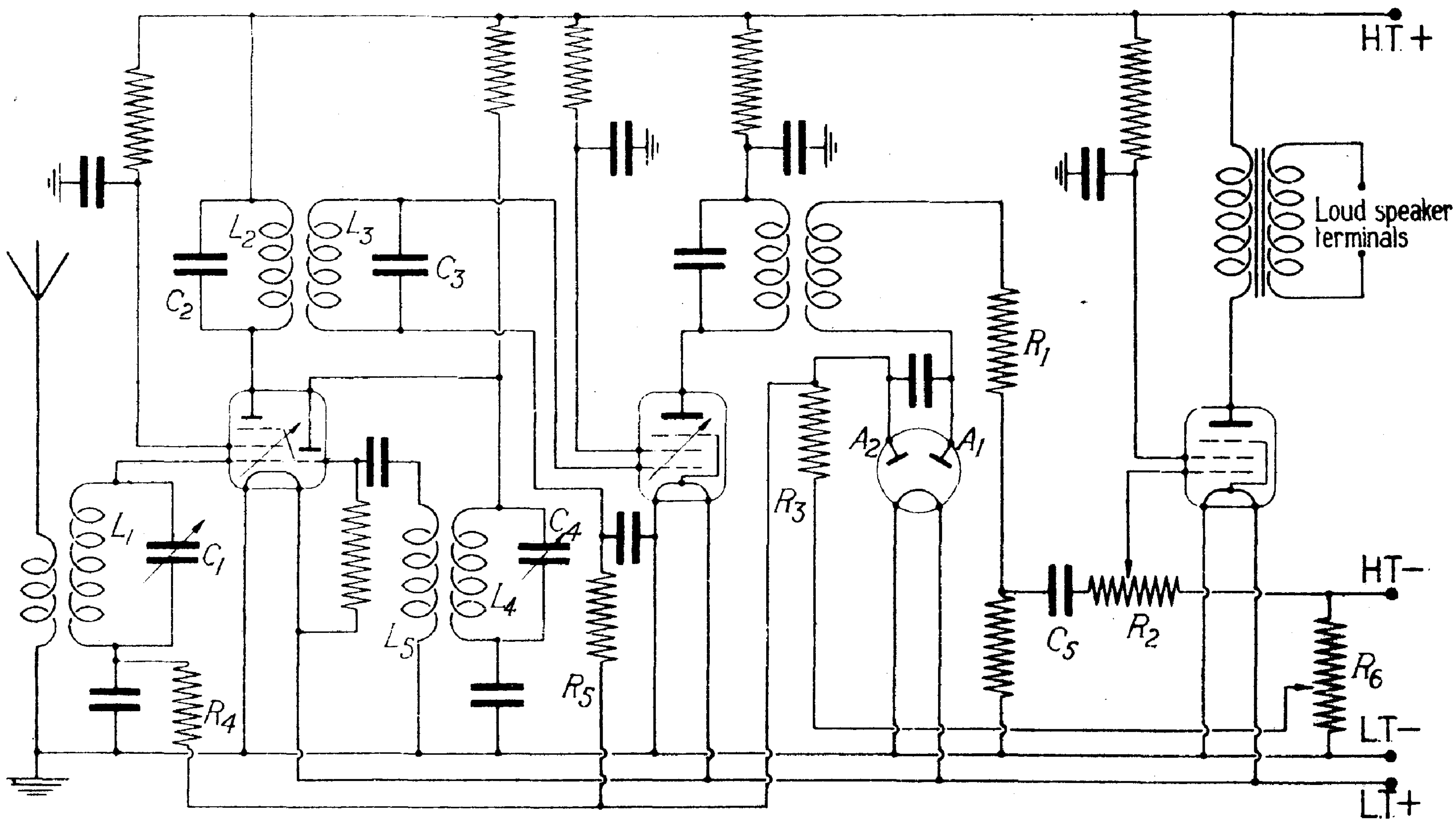
FIG. 45, CHAP. XII.—Paraphase amplifier.

anode A_2 of the second detector is maintained at a mean negative potential with respect to the filament by means of a tapping on the resistance R_6 thus giving a definite delay to the operation of the A.V.C. This resistance also serves to give the required bias to the pentode output valve.

PHASE AND FREQUENCY MODULATION

Phase modulation

78. Phase modulation is performed by maintaining the wave at a constant amplitude, while introducing a cyclical phase shift with reference to the phase of the carrier under non-modulated conditions, the phase shift being proportional to the amplitude of the modulating current or voltage. Such a wave is represented by the heavy line of fig. 47, while the relative phase of the unmodulated wave, which is assumed to be of sinusoidal wave-form, is also shown in the diagram. In the particular conditions illustrated the modulated wave advances in phase during the first quarter of a cycle of the modulating frequency, and then commences to fall back, momentarily assuming its original phase at the end of the first half-cycle. During the next half-cycle the modulated wave lags behind its normal phase, the lag reaching a maximum value at the negative peak of the audio-frequency cycle, after which the phase angle decreases.



SUPER-HETERODYNE RECEIVER FOR R/T

FIG. 46
CHAP XII

This Page Intentionally Blank

At the end of one complete cycle of modulation the wave has the same phase as it would possess in the absence of modulation. If the phase shift introduced by the modulation at any instant is $\varphi(t)$ radians, a phase-modulated current can be represented by the equation

$$i = \mathcal{I} \sin \{ \omega_r t + \varphi(t) \}$$

$\frac{\omega_r}{2\pi}$ being the carrier or unmodulated frequency. The angle $\varphi(t)$ is not constant but varies sinusoidally at the audio-frequency, $\frac{\omega_a}{2\pi}$. If we define the modulation ratio M in such a manner

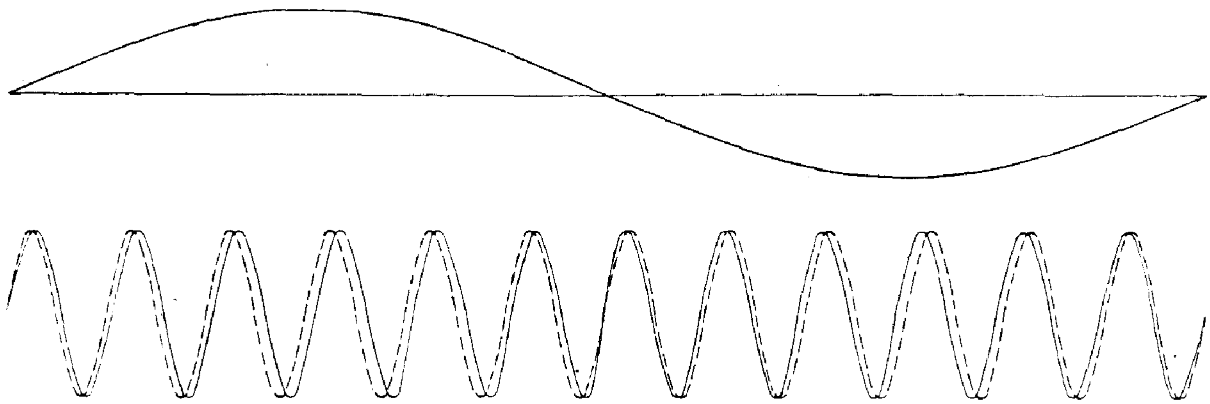


FIG. 47, CHAP. XII.—Phase-modulated wave.

that a modulation ratio of unity will cause a peak phase shift of one complete radio-frequency cycle or 2π radians, $\varphi(t) = 2\pi M \sin \omega_a t$ and therefore

$$i = \mathcal{I} \sin \{ \omega_r t + 2\pi M \sin \omega_a t \}$$

Phase modulation is rarely if ever deliberately adopted for communication purposes, but occurs to some extent in amplitude-modulated systems, as explained in paragraph 40. It is particularly liable to occur when a modulated amplifier is employed, since the magnitude and phase angle of the load impedance generally varies to some extent with the frequency. It is chiefly of importance because of its close resemblance to frequency modulation.

Frequency modulation

79. This type of modulation is achieved by maintaining the constant amplitude of the radiated wave, while varying the frequency in accordance with the amplitude of modulation. Such a wave is shown in fig. 48 which should be compared with the phase-modulated wave of fig. 47. It will be observed that the two waves are very similar in character. The difference between phase and frequency modulation lies partly in the amount of phase shift during any half-cycle of the modulation frequency. If the phase shift in a phase-modulated signal is 2π radians, and there are 1,000 radio-frequency cycles in each half-period of modulation, there will be only 499 radio-frequency cycles in the first quarter period and 501 cycles in the second quarter period of modulation, 501 cycles during the third quarter period and 499 during the fourth, so that the average radio frequency during each half period is equal to the unmodulated or carrier frequency. In a frequency-modulated wave, however, the peak phase shift may be more than 2π radians, and there may be many more radio-frequency cycles during one half period of modulation than in the half period of opposite sign. Hence the average radio frequency during any one half period of modulation is not equal to the frequency of the unmodulated oscillation. It can be shown that the equation representing a frequency-modulated current is

$$i = \mathcal{I} \sin \left\{ \omega_r t + \frac{f_r M}{f_a} \sin \omega_a t \right\}.$$

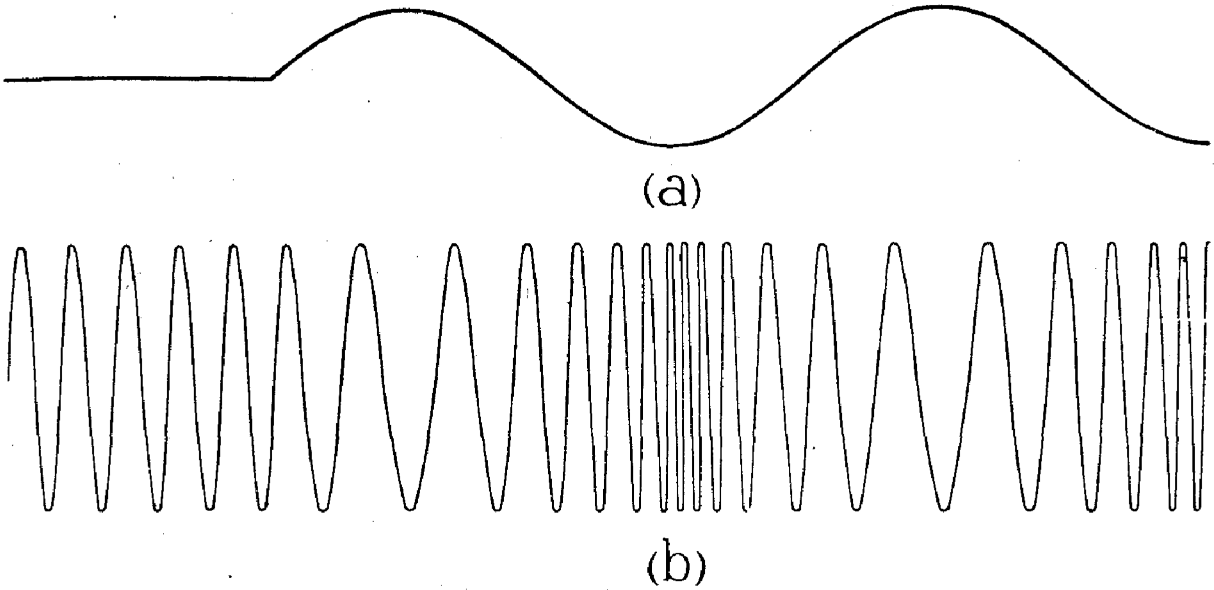


FIG. 48, CHAP. XII.—Frequency-modulated wave.

80. The frequency-modulated wave may therefore be regarded as a particular form of phase-modulated wave, in which the peak phase shift is $\frac{f_r M}{f_a}$ instead of $2\pi M$, i.e. for a given carrier frequency f_r and modulation ratio M the peak phase shift is inversely proportional to

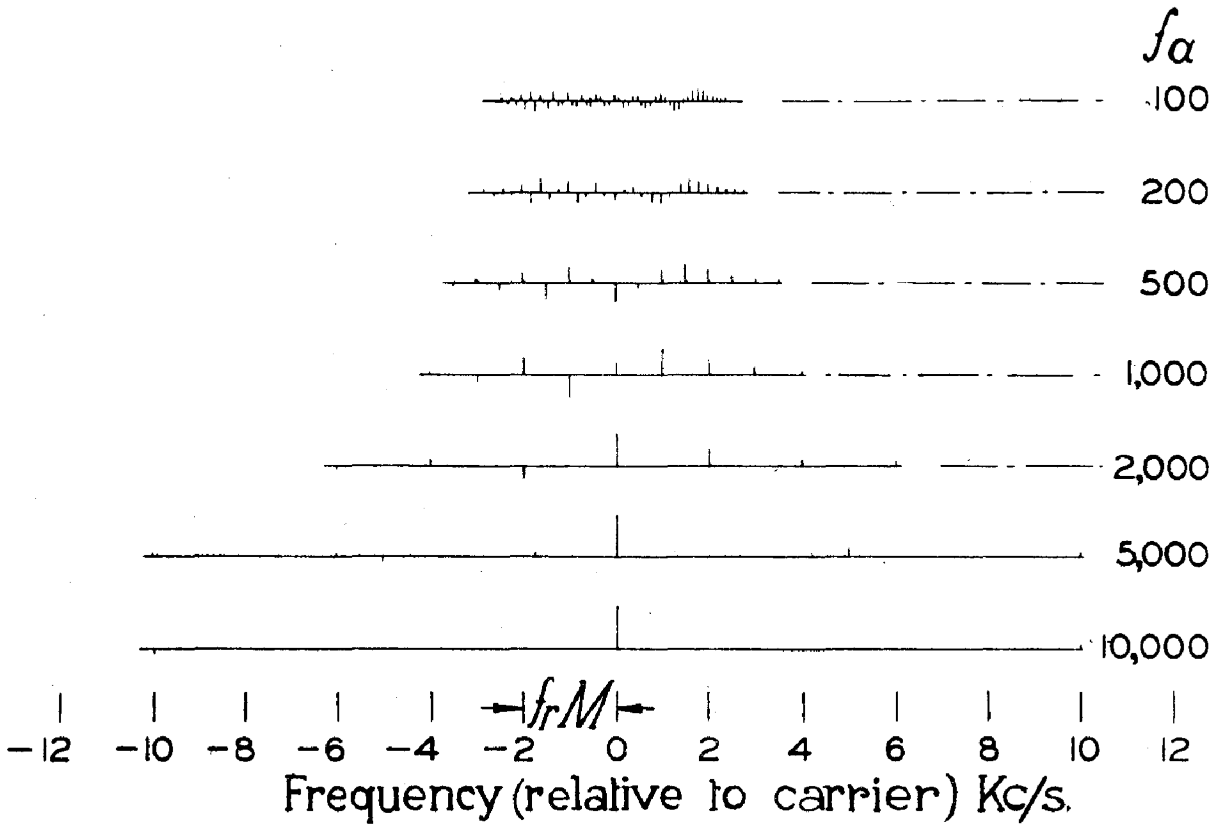


FIG. 49, CHAP. XII.—Amplitude and phase of components of frequency-modulated wave.

the modulation frequency f_a, instead of being a constant for all modulating frequencies as in phase modulation. The expression $f_r M$ will be referred to as the frequency deviation. As in the case of amplitude modulation, the wave may be resolved into the sum of a number of components. The frequencies of these components are $f_r, f_r \pm f_a, f_r \pm 2f_a, f_r \pm 3f_a, f_r \pm 4f_a$, etc., representing a carrier with side-bands. Modulation by a single audio frequency thus produces an infinite number of side-frequencies instead of only one pair. The relative magnitudes and phases of the carrier and side-band components, for a frequency deviation of 2 kc/s and for various modulation frequencies, are shown in fig. 49. When the peak phase shift is less than one radian, the amplitude of the side frequencies $f_r \pm f_a$ is approximately proportional to the frequency deviation $f_r M$, and the carrier component of frequency f_r has an amplitude practically equal to the amplitude of the unmodulated wave. If however the peak phase shift is greater than one radian, the carrier component decreases in amplitude and may be very small or even zero, while the components of frequency $f_r \pm 2f_a, f_r \pm 3f_a$, etc. may, be very prominent, and there may be side-bands extending up to the extreme limits of the frequency deviation. For practical purposes, the total band width occupied by a frequency-modulated wave may be taken as rather more than twice the frequency deviation or rather more than twice the modulation frequency, whichever is the greater. It is seen therefore that frequency modulation is no solution to the problem of frequency allocation, the total band width required being somewhat greater than that occupied by an amplitude-modulated transmission of comparable quality.

81. To some extent, frequency modulation must occur in a nominally amplitude-modulated system, particularly in circuits in which the carrier frequency varies with the instantaneous power taken by the oscillatory circuit. The deliberate adoption of this form of modulation for high frequency transmission appears to offer certain advantages, but sufficient experience has not yet been gained to make it clear whether these are offset by its disadvantages. A frequency-modulated system of high quality demands a much more elaborate circuit than an amplitude-modulated system. One of the earliest attempts to realize a frequency-modulated wave consisted of a master-oscillator transmitter in which a condenser microphone of capacitance C_m was connected in parallel with the tuning capacitance C of the master-oscillator, but such a transmitter will produce a purely frequency-modulated wave only (i) if the ratio $\frac{C_m}{C}$ is vanishingly small, so that the frequency deviation and equivalent depth of modulation approach zero, and (ii) if the frequency of the unmodulated oscillation is extremely constant.

82. In a high fidelity frequency-modulated system, then, the following conditions must be satisfied.

(i) The mean frequency, f_r, must be stable.
(ii) There must be no amplitude modulation.
(iii) The frequency deviation must be independent of the modulation frequency and directly proportional to the amplitude of the modulating current or voltage.

Condition (ii) is comparatively easy to satisfy at high frequencies where the ratio of frequency deviation to mean audio frequency can be kept quite small, say one part in ten thousand, for the resonance curve of the master-oscillator circuit will be practically flat over this limited deviation. Condition (iii) can be met by a careful consideration of the method of modulation control. Condition (i) is not so easily satisfied; for instance, if an attempt is made directly to vary the frequency of the master-oscillator in the manner outlined in the previous paragraph, it follows that the oscillator must be of low inherent stability.

83. (i) During the last few years, a method has been developed which appears completely to solve this problem. It depends upon the close resemblance between phase and frequency modulation. It has already been stated that in a frequency-modulated signal the peak phase shift is $2\pi M$ and in a frequency-modulated signal is $\frac{f_r M}{f_a}$ being therefore inversely proportional to the modulating frequency. In the system to be described, the modulating voltage is caused to shift the phase of a current derived from a source of constant frequency by an amount which

is directly proportional to the amplitude of the modulation and inversely proportional to its frequency. For reasons which will be apparent later, the modulated wave is then subjected to considerable frequency multiplication before it is applied to the final power amplifier. The initial process is to produce the required phase shift, the circuit arrangements to this end being shown in fig. 50. The master-oscillator may have a frequency of from 50 to 100 kc/s, and an E.M.F. from this source is supplied to the balanced modulator T_a T_b, and also, in synphase, to a radio-frequency amplifier valve T_1. The balanced halves L_a C_a L_b C_b of the output circuit of the modulator are acceptor circuits for the master-oscillator frequency (parallel feed being employed by means of the resistances R_a R_b) and the anode currents in the coils L_a L_b are in phase with the corresponding grid-filament input voltages. The output from the modulator feeds a modulation amplifier valve T_2 through an inductive coupling, the effect of the latter being that the output voltage of T_2 is 90° out of phase with the output of the master-oscillator.

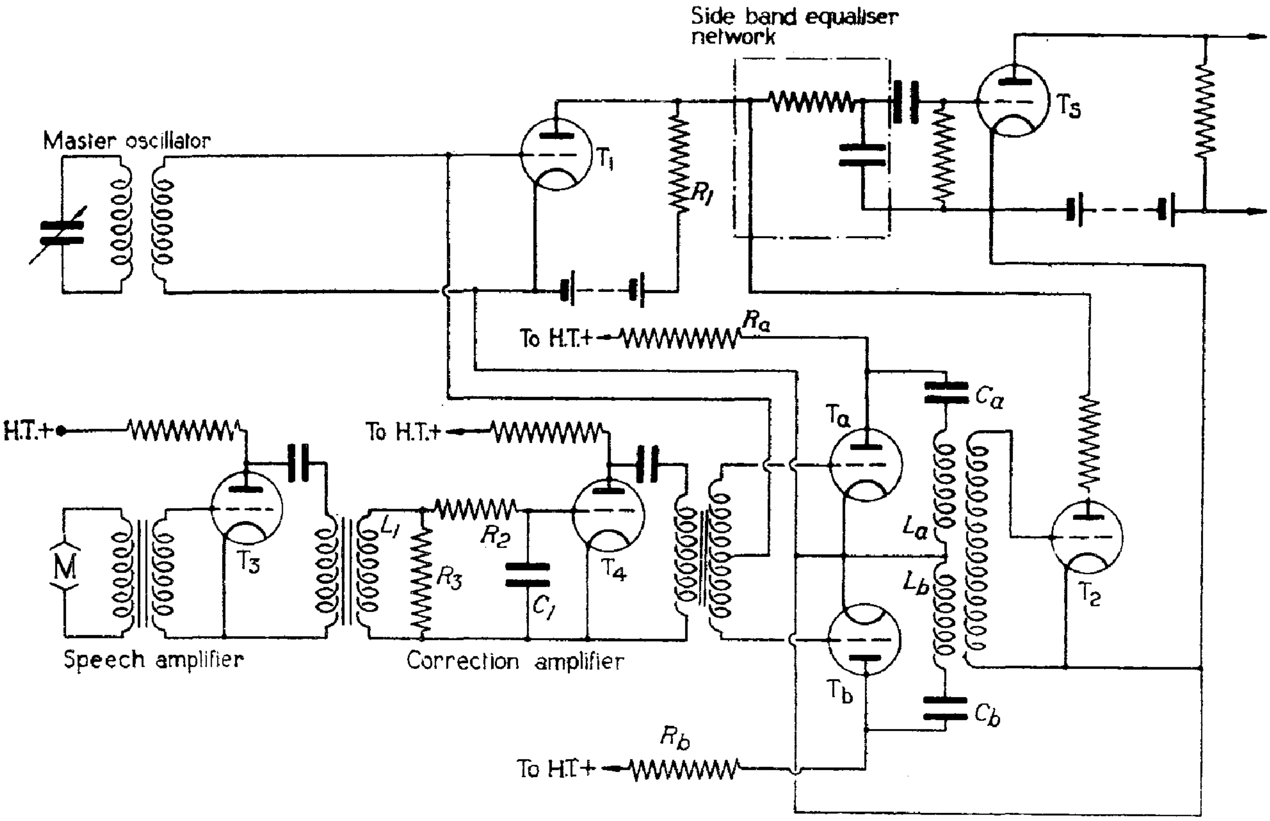
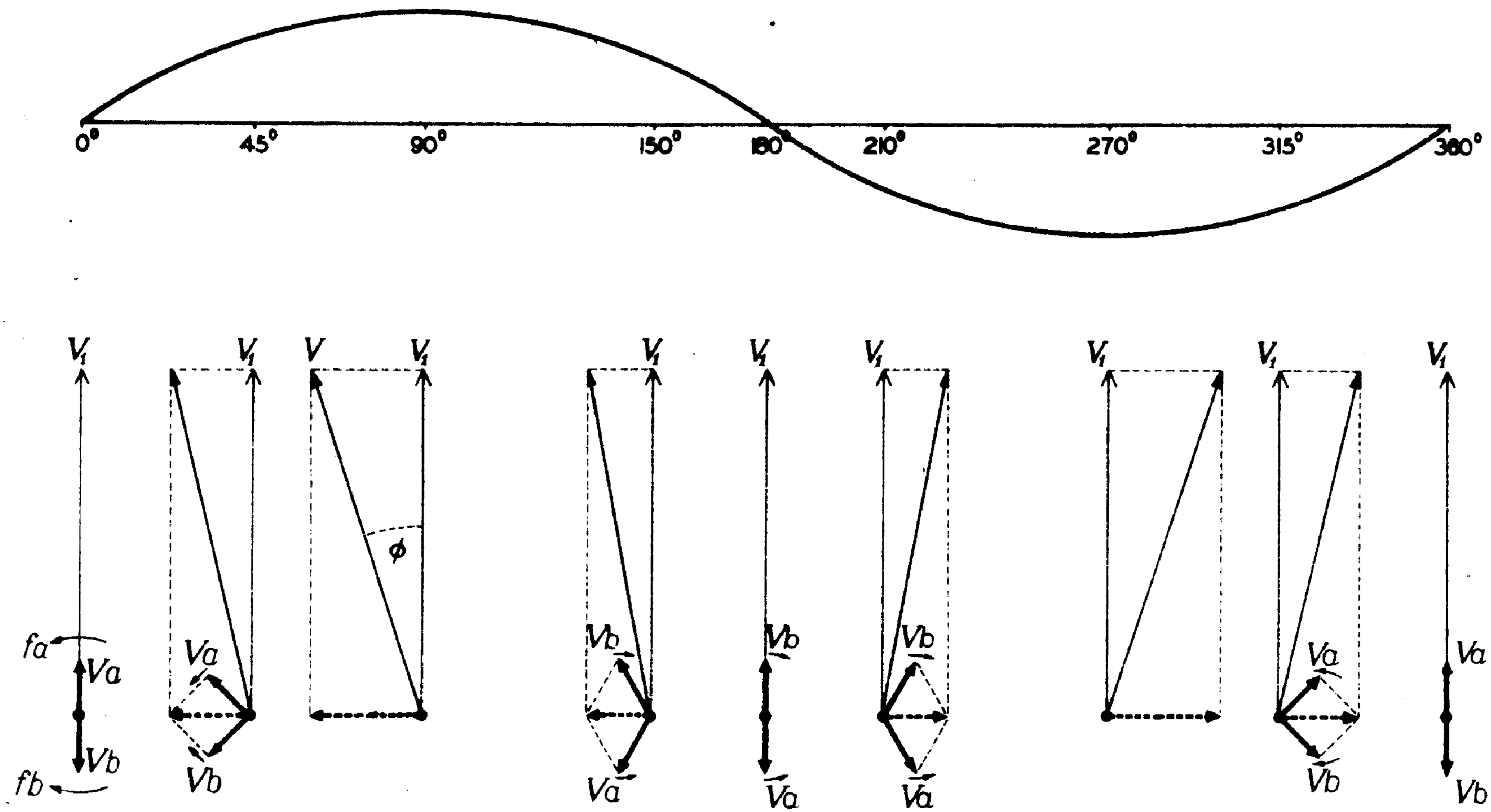


FIG. 50, CHAP. XII.—Modulating stages, frequency-modulated transmitter.

The unmodulated amplifier valve T_1 and modulated amplifier valve T_2 work into a common resistance load R_1 and the voltages developed by the two amplifiers, between the ends of this resistance, are in quadrature. The vector diagrams of fig. 51 illustrate the changes in phase (and incidentally in amplitude) of the P.D. across the resistance R_1. The vector V_1 represents the voltage due to the valve T_1 while the vectors V_a and V_b represent the output voltages of the valve T_2, due to the valves T_a T_b of the modulator. The vector V_1 revolves at ω_r, V_a at $\omega_r + \omega_a$ and V_b at $\omega_r - \omega_a$ radians per second. With respect to the vector V_1, V_a rotates at $+\omega_a$ radians per second and V_b at $-\omega_a$ radians per second, so that if V_1 is considered to be stationary, V_a and V_b may be considered to rotate f_a times per second in the counter-clockwise and clockwise directions respectively. The peak voltage across the resistance R_1 at several different instants during one audio-frequency cycle is therefore as shown in the diagram. It will be observed that the phase of the resultant voltage varies between φ and zero during this period, i.e. φ is the maximum phase shift, occurring at the peak of the audio-frequency cycle.



VECTOR DIAGRAM SHOWING PRODUCTION OF PHASE MODULATION

FIG. 51
CHAP. XII

This Page Intentionally Blank

(ii) Provided that φ is not allowed to exceed about 30°, it will be very nearly proportional to the magnitude of the sum of the modulating components. Referring to the diagram, it will be seen that $\tan\varphi = \frac{V_a + V_b}{V_1}$. As φ is made smaller and smaller, so $\tan\varphi \to \sin\varphi \to \varphi$, i.e. for small phase shifts $\varphi \doteqdot \frac{V_a + V_b}{V_1}$. The magnitude of the resultant voltage V varies between the values V_1 and $\sqrt{V_1^2 + (V_a + V_b)^2}$, so that there is a slight degree of amplitude modulation. This again is small if $V_a + V_b$ is small compared to V_1 and in any case is easily removed by a limiting stage.

84. Hitherto we have considered the phase shift to be the same for any value of f_a whereas if the final result is to be truly a frequency-modulated wave, the phase shift must be inversely proportional to f_a. This is achieved by means of the audio-frequency correction network associated with the audio-frequency amplifier valve, T_4, preceeding the balanced modulator (fig. 50). The first valve T_3 serves merely to couple the microphone to the attenuation network $L_1\, C_1\, R_2\, R_3$. The reactance of C_1 at the lowest modulation frequency is small compared to the resistance R_2. Let a voltage $v = \mathcal{V} \sin \omega_a t$ be applied to the series circuit $C_1\, R_2$. The P.D. across the condenser will be

$$v_c = \frac{v}{\sqrt{R_2^2 + \left(\frac{1}{\omega_a C_1}\right)^2}} \times \frac{1}{\omega_a C_1}$$

As $\frac{1}{\omega_a C_1}$ is always small compared to R_2, v_c is approximately equal to $\frac{v}{\omega_a C_1 R_2}$, i.e. $v_c \propto \frac{1}{\omega_a}$, which is what is required. As the maximum phase shift is less than 30°, and this is only permissible on the lowest audio-frequencies, it is necessary to appreciate the order of the phase shift which will be obtained when the highest audio-frequencies are transmitted. If the band to be covered is 100 to 5,000 cycles per second, and the phase shift for 100 cycles per second is 25°, at 5,000 cycles per second it will be $25° \times \frac{100}{5{,}000} = 0{\cdot}5°$, and for a wider band it would be still less. This amount of phase shift would produce side bands of negligible amplitude; the apparent difficulty is however removed by frequency multiplication.

85. The conditions for the distortionless amplification of a frequency-modulated wave are

(i) the gain must be constant over the whole band width occupied by the signal.

(ii) the phase change introduced by the amplifier must be directly proportional to the frequency.

These conditions ensure that the relative phases and amplitudes of the side bands are unchanged by amplification. It will easily be seen that frequency multiplication increases the frequency deviation in the same ratio as the mean frequency. It may be taken that if the side-bands in the final output are to be of sufficient amplitude to give the equivalent of 100 per cent. amplitude modulation, a phase shift of 45° will be required at the highest audio-frequencies. In the example taken above where the phase shift of the modulator stage is only $0{\cdot}5°$, it would be necessary to employ a frequency multiplication of $\frac{45°}{\cdot 5} = 90$. Actually, for very high quality transmission it may be necessary to multiply from 500 to 1,000 times. It should be particularly noted that with this system of modulation the master-oscillator may be designed to possess the highest possible frequency constancy, since no attempt is made directly to vary its frequency. The frequency deviation must be symmetrical about the carrier frequency, and the simple balanced modulator described above will not fulfil this requirement. Referring to fig. 51 it is seen that if the side-band currents in T_a and T_b are equal, say i_a, the side-band voltages applied to the modulation amplifier will be proportional to $(\omega_r + \omega_a)i_a$ and $(\omega_r - \omega_a)i_a$, so that the vectors V_a, V_b will not be of equal amplitude. This inequality is easily removed by the insertion of a correcting network at a suitable point in the amplifying chain.

86. Suppose that this system is applied to a transmitter which is normally driven by a master-oscillator at 2 Mc/s, with a frequency multiplication of 24, i.e. 3 doublers and a tripler, giving an operating frequency of 48 Mc/s. To adapt it to frequency multiplication, the new master-oscillator may be designed to operate at 61·25 kc/s, and the initial phase shift would be introduced at this frequency, a multiplication of 32 being necessary before coupling to the original transmitter; thus the overall multiplication would be $32 \times 24 = 768$. It is also of interest to note that changing the frequency of a frequency-modulated oscillation by the heterodyne method does not change the frequency deviation, e.g. if a frequency $f_r + f_a$ is caused to beat with a frequency f_o, the number of beats per second is $(f_r + f_a) \sim f_o$, say $(f_r - f_o) + f_a$. This enables an ingenious system to be used for a change of operational frequency.

87. If the above transmitter is required to change its operational frequency in the usual manner a considerable amount of readjustment would be required. If however the modulated master-oscillation is subjected to considerably greater frequency multiplication and then heterodyned by an oscillation of adjustable frequency, it is possible to dispense with a considerable amount of the readjustment. For example suppose the original oscillation (61·25 kc/s) to be multiplied 256 times, giving 15,680 kc/s. If this is heterodyned by an oscillator having a frequency of 13,680 kc/s, the output after rectification will be at 2,000 kc/s, and will still carry modulation with the original frequency deviation. The 2,000 kc/s oscillation may now be multiplied by 24 giving the final output at 48 Mc/s. If the final frequency is to be changed to 40·32 Mc/s, there is no necessity to interfere with the master oscillation or pre-heterodyne multiplier stages, for if the heterodyne oscillator is adjusted to 14,000 kc/s, the modulated beat oscillation will be 1,680 kc/s, and on adjusting the three doubling and one tripling stage to deal with this frequency the final output will be at 40·32 Mc/s. With this method of adjustment then, those stages which require the most careful manipulation can be set up once and for all, subsequent frequency adjustment being performed in the heterodvne oscillator and the four succeeding stages only.

Detection of frequency-modulated wave

88. The application of a frequency-modulated signal to any circuit network having a response dependent upon frequency will result in a wave which is modulated both in frequency and amplitude. Such a wave may be rectified in the same manner as a purely amplitude-modulated signal. The choice of a suitable circuit network presents a compromise between the conflicting claims of sensitivity and selectivity. If, for example, a frequency-modulated current passes through an inductance L of negligible resistance, the voltage v_L developed across its terminals is linearly amplitude-modulated. Thus if

$$i = \mathcal{I} \sin\left(\omega_r t + \frac{\omega_r M}{\omega_a} \sin \omega_a t \right)$$

$$v_L = - L \frac{d i}{d t}$$

$$= - \omega_r L \mathcal{I} \left(1 + M \cos \omega_a t \right) \cos \left(\omega_r t + \frac{\omega_r M}{\omega_a} \sin \omega_a t \right)$$

i.e. the voltage wave is frequency-modulated and is also amplitude-modulated to a depth M. Since however in all practical cases M is small compared to unity this method is too insensitive to be of value as a means of translating pure frequency modulation into amplitude modulation.

89. Various methods of greater sensitivity have been proposed, all of which depend upon the shape of the resonance curve of a circuit possessing both inductance and capacitance. Suppose a frequency-modulated voltage of amplitude $\mathcal{E}$ to be applied to a series-resonant circuit as in fig. 52a. Let its mean frequency be $\frac{\omega_r}{2\pi} = f_r$ and the peak frequency deviation be $f_r M$. Then if the resonant frequency of the circuit is f_o, fig. 51b, the voltage across the inductance will have

the value $\mathscr{V}_o$, during periods of no modulation, while during the time sinusoidal frequency-modulating is occurring with a peak deviation f_rM the voltage will vary between $\mathscr{V}_o + \mathscr{V}_1$ and $\mathscr{V}_o - \mathscr{V}_1$ as in the diagram, i.e. the grid-filament voltage will be amplitude-modulated. Such circuits are most effective when connected in a push-pull arrangement, in order that any slight amplitude modulation of the original signal shall be cancelled out, and the translation from frequency-modulation to amplitude-modulation may then be very nearly linear; it will not be entirely so since the slope of the resonance curve is not quite constant over the operating range. The distortion due to this is not serious unless the frequency swing is sufficient to carry the operating point over the peak of the resonance curve in which event distortion becomes so serious that speech is quite unintelligible. If, however, care is taken to limit the frequency deviation this method of detection appears to offer possibilities for aircraft communication. It should be noted that if the mean frequency is changed before detection from a high to a low value, as in a super-heterodyne receiver, the ratio of frequency deviation to mean frequency is increased, and a greater rectified current is obtained than if the detection is directly performed at the initial frequency.

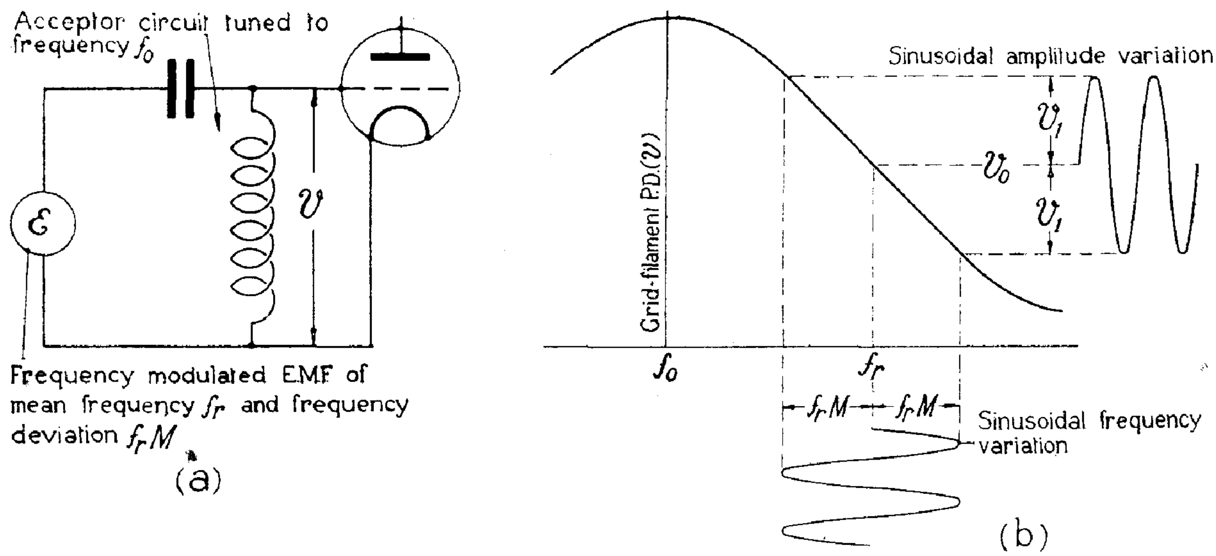


FIG. 52, CHAP. XII.—Translation of frequency modulation into amplitude modulation.

The type of receiver generally indicated is therefore a double super-heterodyne receiver in which the incoming very high frequency signal is first changed from the order of 100 Mc/s to say 5 Mc/s and afterwards to about 500 kc/s, thus providing considerable discrimination against second channel interference. Certain limiting stages must be introduced before the detector stage in order to eliminate any adventitious amplitude modulation which may be introduced either during propagation or in the tuned circuits of the receiver.

Signal-noise ratio and interference

90. It is claimed for frequency modulation (when carried out on very high frequencies with a large frequency deviation and a correspondingly wide frequency band) that the signal noise ratio is very much better than with amplitude modulation. In the latter case, the smaller the band of frequencies admitted by the receiver, the better is the signal-noise ratio, whereas in frequency modulation the greater the frequency deviation the wider is the band embraced by the side-frequencies and the greater is the depth of amplitude modulation after translation in the detector stage. Consequently, the signal-noise ratio increases with the frequency band embraced by the signal. For example, comparing an amplitude-modulated system with a frequency band of 20 kc/s with a frequency-modulated system having 100 kc/s frequency

deviation for the same audio-frequency range, and assuming equal power to be drawn from the mains by both transmitters, the signal-noise voltage ratio of the frequency-modulated system is about 34 times that obtained by amplitude modulation. These are of course theoretical values, but it is claimed that improvements of this order are practically realizable. The actual improvement is however approximately inversely proportional to the initial noise level, i.e. the signal-noise ratio decreases as the noise level increases.

91. The possibilities of interference with frequency-modulated reception have not been completely investigated. The following results are probable, but are still to be verified experimentally.

(*a*) Amplitude-modulated signals should not interfere with frequency-modulated signals unless the two carriers are capable of combining to produce an audio-frequency beat note, i.e. the two carriers are nearly equal in frequency or separated by nearly twice the intermediate frequency.

(*b*) Second channel interference can occur between frequency-modulated signals exactly as between amplitude-modulated signals.

(*c*) It is claimed that two frequency-modulated transmissions can be separated, even if their side-bands overlap, provided that there is no audio-frequency beat between the carriers.

(*d*) No information is yet available with respect to cross-modulation.

Effect of propagation path

92. If an amplitude-modulated signal arrives at a receiver by two or more paths of different length, it undergoes what is called selective fading (Chapter XIV). In the same circumstances, a frequency-modulated signal is subjected to attenuation of some of the side-band frequencies to a greater extent than others, thus giving rise to distortion. Its use is therefore limited, either to short distances where only the direct ray is of importance, or to transmission at very high frequencies where there is no ray reflected from the ionosphere. The reflection at the surface of the earth must be taken into account, but at present there is little if any information available regarding the influence of these phenomena.

Summary

93. The present state of frequency modulation may therefore be summarized as follows :—

(i) Under certain conditions of frequency and propagation path a frequency-modulated system provides quite a practical method of communication.

(ii) Such frequencies are limited to the very high frequency band.

(iii) Propagation must take place over a single path.

(iv) The apparatus is rather more complicated than for an amplitude modulation system of comparable quality, but not unduly so.

(v) The signal-noise ratio appears to be very much better than in the corresponding amplitude-modulated system.

(vi) At the present stage of high frequency technique it should be possible to allocate as many frequency-modulated as amplitude-modulated systems within a given frequency band. This may not be true in the future, as the frequency stability of very high frequency amplitude-modulated systems may be considerably improved enabling allocation to be made much more closely than at present.


CPL V WOOTON
23