SECTION 2

ADVANCED PRACTICAL RADIO ENGINEERING

TECHNICAL ASSIGNMENT

F.M. RECEIVERS

238 E

F.M. RECEIVERS

FOREWORD

When Professor Armstrong first introduced his f.m. system, many skeptics looked askance at it, in spite of the remarkable demonstration he offered in proof of its noise-free capabilities. He himself stated that several years earlier he had proved mathematically that the system had no advantages over amplitude modulation, but in spite of such self-inflicted discouragement, proceeded to develop the system until he had satisfied himself that it had remarkable possibilities.

Today f.m. is here to stay. There may be some justice to the claim of some engineers that if as much effort had been put into improving a.m., comparable results would be obtained, but the fact remains that f.m. is expanding rapidly as a broadcast service, as well as for point-to-point and other forms of communication, and the engineer who does not care to learn the theory and application of this type of modulation had just as well reconcile himself to a minor post in industry.

This assignment deals primarily with the general theory of frequency modulation, its ability to suppress interference of various kinds, and the general design considerations pertaining to f.m. receivers. However, f.m. modulating methods are also discussed, although a more complete exposition appears in the assignment on modulation.

It is recommended that this assignment be studied carefully, and the topics re-read, if necessary, until the student feels satisfied he has mastered its contents. The result will be a well-grounded knowledge of the principles of this subject, and a better appreciation of its capabilities. This is a reward worth striving for; a sufficient recompense for the effort put into the study of this assignment.

E. H. Rietzke,
President.

- TABLE OF CONTENTS -

F.M. RECEIVERS

F.M. RECEIVERS

INTRODUCTION

Before going into the theory of modulation, and particularly frequency modulation, it will be of value to make a general survey of this new service to see what it comprises, and what its advantages are.

The transmission of intelligence has been accomplished from earliest times by means of sound and light waves. Thus, even the chirp of a cricket, or the light produced by a firefly, is presumably a mating call.

Man has extended the methods of transmission to include not only sound and light, but electricity as well. In telegraphy, for example, a special code of dots and dashes is employed to break up or *modulate* a direct current in order to transmit intelligence to a remote point. The d.c. may be regarded as a zero-frequency carrier.

In telephony, the d.c. carrier is varied or modulated by the speech components themselves, although speech in itself may be regarded as a code for certain concepts and mental pictures. In another form of modulation of the d.c. carrier, the latter is varied in accordance with the light and shade of a picture or scene; this represents the wire transmission of facsimile and television.

When it was attempted to transmit directly through space (wireless) it was found necessary to substitute a high-frequency wave or carrier for the d.c. The methods of modulating such a high-frequency carrier are greater in number than for the d.c. carrier; amplitude, frequency, phase, or pulse-time modulation can be employed.

The most obvious form of modulation is amplitude modulation (a.m.), in which the amplitude of the r.f. carrier is varied at the modulation rate. It was the first employed, and is in wide use today. Although it has been treated in previous assignments, a brief review of a more mathematical nature will be given farther on in this assignment. Pulse-time modulation is discussed in the assignment on pulse techniques, and will not be treated here.

Frequency modulation (f.m.) is the process of varying the carrier frequency at the modulation rate. For example, suppose the carrier frequency is 100 mc, and the modulation frequency is 1,000 c.p.s. Then the carrier frequency will be varied *above and below* its 100-mc value 1,000 times a second.

The student may immediately ask, "How far above and below?" The answer is, "This depends upon the *amplitude* (loudness) of the 1,000-cycle note." If the note is of the maximum loudness that can be handled by the system, then it has been standardized that the 100 mc shall vary a full 75 kc or .075 mc above and below the unmodulated value of 100 mc. Thus, the frequency would vary back and forth from 100 mc - .075 = 99.925 mc to 100 + .075 = 100.075 mc, 1,000 times a second. It is as if one were to rock the tuning condenser of an oscillator back and forth, 1,000 times a second, through a range causing the oscillator to oscillate between the values of 99.925 mc and 100.075 mc. For a 1,000-cycle tone of half the amplitude, the frequency variation will be half, or 37.5 kc above and below the carrier frequency. Thus, the loudness of the tone determines the extent of the frequency variation, and the frequency of the tone determines the number of to-and-fro variations per second.

Some surprise may be occasioned by the use of such a mode of modulation, since it appears to be more of an unconventional or freak method, compared to a.m., which seems a natural form of modulation. This, however, is precisely the reason why f.m. is of value. Signals propagated from the transmitter to any receiver have to contend with certain inherent enemies to reception. These enemies are interference and noise; interference from other stations on the same channel and on adjacent channels, thermal noise, and man-made static, such as ignition and motor-sparking noises.

The interfering signals in general represent amplitude variations; their frequency variations are very small. For example, a strong pulse will shock-excite the receiver tuned circuits and cause them to oscillate at their natural frequency, a fixed value. Hence, their effect upon a *frequency-sensitive* detector, such as is used in an f.m. receiver, will be small; much smaller than their effect upon an *amplitude-sensitive* detector, such as is used in an a.m. receiver.

Another factor that redounds to the advantage of f.m. is the fact that the volume range it can handle in the form of frequency variation or deviation is extremely large if the carrier is at a sufficiently high frequency, *and this modulation is independent of the carrier amplitude*. The normal f.m. band is from 88 to 108 mc, so that a deviation as great as 75 kc is but a small percentage of the carrier frequency anywhere in this range.

In amplitude modulation, on the other hand, the amplitude, for 100 per cent modulation is limited to the carrier amplitude; i.e., the carrier can vary from twice its normal (unmodulated) value to zero. The significance of this is that noise and interference, predominantly variable in amplitude, cannot be swamped out by a.m. on a low-level carrier, whereas the small frequency variation of the noise can readily be swamped out by a carrier undergoing a large frequency deviation, even though the amplitude of the wave is low. In short, the more one can make the desired modulation differ in nature from the undesired interference modulation, the better are the chances of having a high signal-to-noise ratio. F.M. has just this kind of an advantage over a.m.

Another advantage has to do with the transmitter. The power output depends upon the amplitude of the output current of the unit, *and not on its frequency*. Hence if the frequency varies during modulation rather than the amplitude, the output power of the transmitter can remain constant at all times, and the unit can operate at peak efficiency and output. Since vacuum tubes are expensive sources of power, it is clear that f.m. operation results in definite economies in this respect over a.m. operation.

The f.m. receiver is in general

exactly like the a.m. receiver, except in two important details: instead of an ordinary second detector, it employs a special frequency detector, and it usually employs a so-called limiter stage or stages ahead of the detector, that remove any amplitude variation in the wave that may be produced within the receiver or in space (owing to the superposition of noise pulses, etc.). Several forms of f.m. detectors are available, and those most commonly used are described in the text.

Hence, briefly, the microphone and its associated amplifying equipment causes the frequency of the transmitter to vary through a range and at a rate depending respectively upon the amplitude and frequency of the impinging sound waves. The resulting frequency variable antenna currents radiate this form of modulated energy into space, where it is picked up by the receiving antenna. It is then amplified by an r.f. amplifier, changed to an intermediate frequency by the mixer tube and local oscillator and is then amplified by the i.f. system. The i.f. carrier varies in frequency similar to the r.f. carrier picked up on the antenna. Any amplitude variations are then clipped off or limited by one or more limiter stages; these do not affect the frequency variable characteristic of the wave.

It is then impressed on circuits whose response varies with frequency. As a result, the *constant-amplitude* variable frequency wave is converted by these circuits into a *variable-amplitude* (and variable frequency) wave. A suitable detector, like a balanced diode circuit, then converts the variable amplitude i.f. wave into an audio wave; this is amplified by an audio amplifier; and applied to the loudspeaker.

THEORY OF MODULATION

GENERAL CONSIDERATIONS.—In studying the various forms of modulation, it is necessary to start with the general equation for a sine wave. To be specific, suppose a sine wave of current be considered. Its equation is

$$i = I \sin(2\pi f t + \phi)$$

$$= I \sin(\omega t + \phi) \qquad (1)$$

where i is the instantaneous value of the wave; I is its maximum or peak amplitude; f is the frequency; $\omega = 2\pi f$; t is the time as measured from some arbitrary starting point or moment; and ϕ is a suitable angle called a phase angle.

In seeking to modulate such a wave with the intelligence to be transmitted, one observes that three quantities can be varied:

1. If the peak amplitude I in itself varies with time in accordance with the intelligence or modulation to be transmitted, the carrier is said to be *amplitude modulated.*

2. If the frequency f is caused to vary with time in accordance with the intelligence to be transmitted, the carrier is said to be *frequency modulated.*

3. If the phase angle ϕ is made to vary with time, the carrier is said to be *phase modulated.*

Actually, as will be shown later, any variation in f or ϕ produces both f.m. and p.m.; the two of necessity occur simultaneously. Whether the system is considered f.m. or p.m. depends upon whether the frequency variation or the phase angle is proportional to the modulation amplitude. As will be shown later, owing to the use of preemphasis and deemphasis networks, it is a moot point whether the actual system in use today is f.m. or p.m.; for convenience, and because fundamentally its operation is based on f.m., the system is said to be one of frequency modulation.

COMPARISON OF A.M. AND F.M..—In Fig. 1(A) is shown an a.m. wave, and in (B) an f.m. wave. Note from Fig. 1(A) that the amplitude, as suggested by the dotted lines, varies in a sinusoidal manner, but at a lower frequency than the carrier itself. The dotted-line envelope represents the intelligence or modulation; in the figure this is simply a low-frequency sine wave.

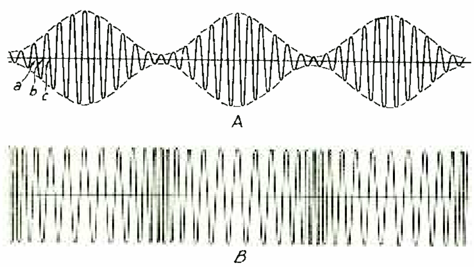


Fig. 1.—A.M. and F.M. waves.

The counterpart in f.m., as shown in Fig. 1(B), is a wave of constant amplitude, but variable period (time of one cycle). The wave has somewhat the appearance of an accordion bellows.

The a.m. wave of Fig. 1(A) crosses the axis in equal time intervals; i.e., ab = bc, etc. This would indicate that the frequency is constant. However, the fact that the amplitude is continually changing implies that the wave is not a sine wave, for a sine wave has a constant amplitude. An analysis made in the next section shows that this wave may be resolved in three sine waves: one of the same frequency as when it is unmodulated (carrier frequency); one of a frequency higher than the carrier by the modulating frequency (upper side band), and one of a frequency lower by the modulating frequency (lower side band).

In the case of the f.m. wave, a similar resolution is possible, but now the number of side bands is theoretically infinite. The side bands are spaced to either side of the carrier at frequency intervals equal to the modulating frequency. Their amplitudes rapidly decrease as one proceeds in the spectrum to either side of the carrier, hence only a limited number are required for adequate representation of a frequency-modulated wave. This means that but a moderate band width is required of the transmitter and receiver circuits.

As will be shown later, if the frequency deviation of the carrier is made sufficiently large, higher signal-to-noise ratios may be obtained in f.m. systems than in a.m. systems, although it is contended by some that the comparison is not made on an entirely fair basis. Nevertheless, f.m. in practice has important advantages over present-day a.m. as regards signal-to-noise ratio, although the system will require several years of use to establish its true worth.

One further point to note is that the ordinary a.m. diode detector, for example, responds to amplitude variations in the carrier wave, but not to frequency variations. When it is desired to detect an f.m. wave, the latter must first be transmitted through a frequency-sensitive network in order that its frequency variations produce corresponding amplitude variations, which can then be detected. The arrangement, in one form, is known as a discriminator circuit, but all f.m. detectors require translation by one means or another from frequency to amplitude variation.

ANALYSIS OF MODULATION

AMPLITUDE MODULATION.—In order more fully to appreciate the foregoing remarks, it will be necessary to analyze the various types of modulation in greater detail. Consider amplitude modulation first. The fundamental equation of an a.c. wave is

$$i = I \sin \omega_c t \qquad (1a)$$

in which ω_c refers to the carrier angular velocity.* The phase angle ϕ of Eq. (1) has been omitted, since no loss in generality will be incurred by such deletion.

Now assume I itself varies in a sinusoidal manner at some low modulating frequency ω_m. Since in general I will vary to a certain fraction m_a of its average value I_{avg}, express I as follows:

$$I = I_{avg} + m_a I_{avg} \sin \omega_m t$$

$$= I_{avg} [1 + m_a \sin \omega_m t] \qquad (2)$$

The maximum value for m_a is unity; this represents 100 per cent modulation. Then, as $\sin \omega_m t$ varies between the limits ±1, I will vary between the limits $I_{avg}[1 + 1] = 2 I_{avg}$ to $I_{avg}[1 - 1] = 0$. Thus, $2 I_{avg}$ and 0 represent the limits of the envelope of the wave. For values of m_a less than unity, I will vary between smaller limits, but in all cases its average value will remain I_{avg}.

Now substitute Eq. (2) in Eq. (1_a) and multiply out:

$$i = I_{avg}[1 + m_a \sin \omega_m t] \sin \omega_c t$$

$$= I_{avg} \sin \omega_c t$$

$$+ m_a I_{avg} \sin \omega_m t \sin \omega_c t \qquad (3)$$

The first factor on the right represents a sine wave of frequency ω_c and *fixed* amplitude I_{avg}; this is the carrier component that is present whether the wave is modulated or not.

The second factor on the right, when expanded by trigonometry, will yield the upper and lower side bands. From trigonometry it is known that

$$\sin A \sin B = \frac{1}{2} \cos (A - B)$$

$$- \frac{1}{2} \cos (A + B) \qquad (4)$$

Let $A = \omega_c t$ and $B = \omega_m t$. Then Eq.

*From now on the word *frequency* will be employed to designate either f or $\omega = 2\pi f$, even though ω is, strictly speaking, not the frequency, but the *angular velocity*.

(3) becomes

$$i = I_{avg} \sin \omega_c t + \frac{m_a I_{avg}}{2} \cos(\omega_c - \omega_m)t - \frac{m_a I_{avg}}{2} \cos(\omega_c + \omega_m)t \qquad (5)$$

The second term on the right is the lower side band; its frequency is $(\omega_c - \omega_m)$. The third term is the upper side band; its frequency is $(\omega_c + \omega_m)$. (A cosine term has the same shape as a sine term, but is 90° ahead of it in phase.) Note that a.m. has produced in addition to the original carrier, two side bands of amplitudes $m_a I_{avg}/2$. For 100 per cent modulation, $m_a = 1$, and these amplitudes become $I_{avg}/2$ or half of the carrier amplitude. Each has therefore *one-quarter* of the carrier power; the two together add 50 per cent to the original carrier power.

In Fig. 2 are shown the carrier and side bands in the frequency spectrum, and their wave shapes.

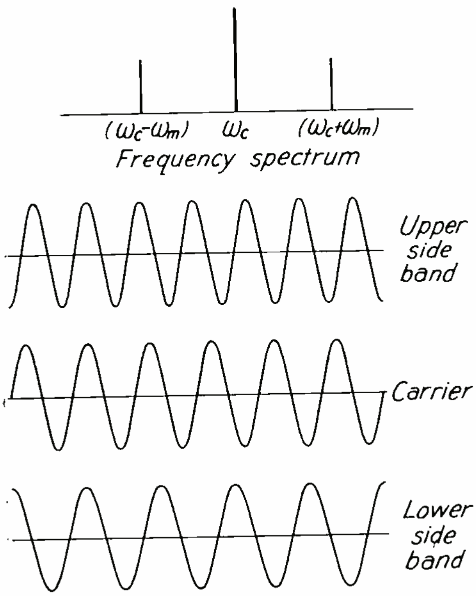


Fig. 2.—Side bands and carrier of an amplitude-modulated wave.

Observe that the cosine side-band waves start at negative and positive maximum at t = 0, whereas the carrier starts at zero, since $-\cos(\omega_c + \omega_m)t = -\cos(0) = -1$; $\cos(\omega_c - \omega_m)t = \cos(0) = 1$, and $\sin \omega_c t = \sin(0) = 0$ at $t = 0$.

This alignment at t = 0 is important and, as will be shown farther on, produces amplitude modulation. If the side bands involved sine instead of cosine functions, and also were of opposite sign, a form of frequency modulation (Armstrong system) would result.

FREQUENCY MODULATION.—In frequency modulation, ω_c varies with time, and I is constant. Suppose ω_c varies sinusoidally with time at a modulating frequency ω_m. Then, similar to Eq. (2), we can write

$$\omega \equiv \omega_c (1 + k_f \cos \omega_m t) \qquad (6)$$

where k_f is similar to m_a, and is known as the *frequency modulation constant.* Since the cosine, like the sine, varies between the limits ±1, ω varies between the limits $\omega_c(1 + k_f)$ and $\omega_c(1 - k_f)$.

For example, if ω_c has the value $(2\pi \times 100 \times 10^6)$, or the carrier frequency f_c is 100 mc, and $k_f = 0.001$, then the instantaneous frequency f varies from 100(1 + .001) = 100.1 mc to 100(1 - .001) = 99.9 mc. The frequency shift or deviation is 0.1 mc or 100 kc to either side of the 100-mc carrier.

Concepts of Phase and Frequency.—The equation for a sine wave is fundamentally in terms of a peak amplitude I and a *phase angle* (ωt). This angle varies directly with time, if ω is constant. If ω is itself variable with time, then the phase angle varies in correspondingly more complicated manner with time. In order to analyze the wave, the phase angle must be known in terms of time. If instead, the frequency ω is given as a function of time, it is first necessary to find the corresponding functional relationship between the phase angle and time before the analysis can be made. It is therefore necessary to examine more critically the method of generating an a.c. wave.

Sine-wave voltages were originally generated by rotating conductors in a magnetic field. As illustrated in Fig. 3, a conducting

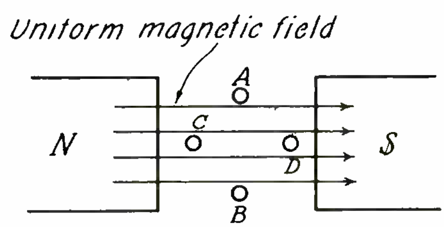


Fig. 3.—Generation of an a.c. wave by the rotation of a loop in a uniform magnetic field.

loop AB (or CD) is revolved with constant angular velocity in a uniform magnetic field, and thereupon generates a sine-wave voltage.

Thus, when the loop is in position AB, the loop sides move parallel to the flux and the instantaneous generated voltage is zero; when it is in position CD, the rate of cutting of the flux is a maximum and the instantaneous generated voltage is at a maximum or peak value. During one revolution, one sine-wave cycle is developed, and at the same time the loop has rotated through an angle of $360° = 2\pi$ radians.

Suppose the loop executes 60 revolutions per second. Then the total number of cycles developed is 60, and the total number of radians swept out is $2\pi \times 60 = 377$ radians. The *rate* at which the radians are swept out is 377 radians *per second*, and the number of cycles (each consisting of 2π radians) is 60 per second. The latter figure is denoted as the *frequency;* if the rotation of the loop is uniform, then the frequency is constant, and the total phase angle up to any moment is simply 2π radians times the frequency times the time. Thus, if the frequency is constant and known, as well as the time, the phase angle is thereupon also known, and the equation for the voltage wave is simply

$$e = E \sin 2\pi ft = E \sin \omega t \qquad (7)$$

which is a well-known formula. The important point to note is that frequency is *rate of change* of phase angle with respect to time, divided by 2π; i.e., $f = \omega/2\pi$.

It will be of interest to pursue this matter in greater detail, and to use the above numerical values. One revolution will take 1/60 sec., and will produce 1 cycle and 2π radians. One-half revolution will take 1/120 sec. and will sweep out π radians; one-quarter revolution will require 1/240 sec. and sweep out $\pi/2$ radians; and similarly for smaller and smaller intervals of time.

Suppose, however, that the loop swept out π radians instead of $\pi/2$ radians in the first quarter of the time required for one revolution; $\pi/4$ radians in the second quarter period; $\pi/2$ radians in the third quarter period; and $\pi/4$ radians in the fourth quarter.

The total number of radians swept out in the period would be $\pi + \pi/4 + \pi/2 + \pi/4 = 2\pi$ (one revolution), but clearly the rate at which they are swept out is very variable, so that the *instantaneous* frequency is variable over the cycle. Thus, during the first quarter period, the instantaneous frequency is π rad./$(2\pi \times 1/240$ sec.) = 120 c.p.s.; during the second quarter it is $(\pi/4)/(2\pi \times 1/240)$ = 30 c.p.s.; for the third quarter it is $(\pi/2)/(2\pi \times 1/240)$ = 60 c.p.s.; and during the fourth quarter it is 30 c.p.s., the same as for the second quarter.

The important thing to note is that frequency is proportional to rate of change of phase angle,* and this is exactly analogous to the relationship between velocity and distance traversed; i.e., velocity is the rate of change of distance with respect to time. If, therefore, a certain frequency variation is desired, a corresponding amount of phase-angle variation must be produced, and frequency modulation must be accompanied by phase modulation, and phase modulation by frequency modulation.

The phase and frequency can, however, vary over many cycles instead of during one cycle. Thus, for a constant frequency of 60 cps, there would be 15 cycles in 1/4 sec, or $2\pi\ 15 = 30\pi$ radians swept out. Now suppose that the rotating loop slowed down so that 28π instead of 30π are swept out. Then the frequency will have decreased from 60 cps to a lower value and then back to 60 cps in 1/4 sec.

Next suppose that in the next 1/4 sec 32π radians are swept out, or 2π radians over the normal value. The frequency must have risen above 60 cps and then back to 60 cps in this 1/4 sec. Then, in the next 1/4 sec assume that 28π radians are swept out once more, and in the fourth 1/4 sec 32π radians are swept out, and so on.

The total number of radians in one second is $(28 + 32 + 28 + 32)\pi = 120\pi$ or $120\pi/2\pi = 60$ cycles, which is the same number of cycles as if the speed had been constant. Yet the actual rotation has been variable; the frequency varied from below 60 cps to above that value. In the first quarter second, the AVERAGE frequency is $(28\pi/2\pi)/(1/4) = 56$ cps. This means that the frequency dropped UNIFORMLY from 60 to 52 cps in 1/8 sec. and then back to 60 cps in the next 1/8 sec.; the average frequency was $(60 + 52)/2 = 56$ cps as stated previously.

In the next 1/4 sec. the average frequency is $(32\pi/2\pi)/(1/4) = 64$ cps. This means that the frequency rose uniformly from 60 cps to 68 cps in 1/8 sec. and then back to 60 cps in the next 1/8 sec. The same reasoning holds for the remaining two quarters of a second.

Thus the frequency drops 8 cps and then returns to 60 cps in 1/4 sec.; it then rises 8 cps and then returns to 60 cps in the next 1/4 sec., and so on. There is consequently a frequency variation or DEVIATION of 8 cps either way, or a total deviation of 16 cps from a minimum to a maximum frequency. Furthermore, there are TWO such COMPLETE deviations per second.

*Students of the calculus will immediately recognize that $f = (1/2\pi)\ d\phi/dt$, and conversely, $\phi = 2\pi \int f dt$.

The number of radians swept out in 1/4 sec. was $(30 - 28)\pi = 2\pi$ radians less, or $(32 - 30)\pi = 2\pi$ radians more than that swept out at a UNIFORM speed. Now suppose that the 2π radians were lost or gained in 1/8 sec. intervals instead of 1/4 sec. intervals. Thus, in the first 1/8 sec. interval, we would have 13π radians swept out instead of the normal 15π radians. The average frequency would be $(13\pi/2\pi)/(1/8) = 52$ cps, and the minimum would be 44 cps.

In the second 1/8 sec. interval, we would have 17π radians swept out, the average frequency would be 68 cps, and the maximum frequency would be 76 cps. In short, the frequency deviation would now be twice as much as before, or ± 16 cps from 60 cps, and it would occur twice as often, or 4 complete cycles per second.

This in turn would mean that 1). the modulating wave is TWICE as great in amplitude, and 2). twice as great in frequency. If the modulating wave were of the same amplitude as before, but twice as great in frequency, then the frequency deviations would be ± 8 cps as originally, but would occur 4 instead of 2 times per second. This is all illustrated in Fig. 3A; the uniform variation above and below 60 cps means a sawtooth modulating wave.

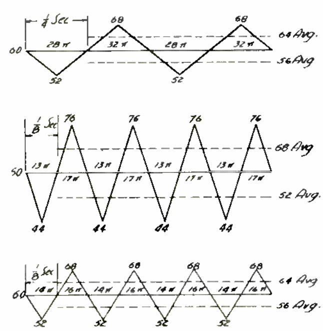


Fig. 3A —Sawtooth frequency modulation at two deviation frequencies and rates.

From all this we can derive the fundamental principles of f.m.:

1. If the deviation in frequency is to represent the AMPLITUDE of the modulating wave, then the higher the modulating frequency, the smaller must be the deviation in PHASE ANGLE, or the phase angle deviation varies *inversely* with the modulating frequency. Thus, for a frequency deviation of ±8 cps, the phase angle deviation is $\pm 2\pi$ radians if the modulating frequency is 2 cps, and $\pm\pi$ if the modulating frequency is 4 cps.

2. For a given modulating frequency, the deviation in phase angle is *directly* proportional to the corresponding frequency deviation. Thus, for a modulating frequency of 4 cps, if the frequency deviation is ± 8 cps, the phase deviation is $\pm\pi$ radians; if the frequency deviation is ± 16 cps, the phase deviation is $\pm 2\pi$ radians.

When a quantity is directly proportional to one factor, and inversely proportional to another, it is proportional to their quotient. Therefore, if the equation for the frequency variation is that given by Eq. (6) or (6a), namely,

$$\omega = \omega_c(1 + k_f \cos \omega_m t) \qquad (6)$$

$$\omega = \omega_c + k_f\omega_c \cos \omega_m t \qquad (6a)$$

where $k_f\omega_c$ is the peak frequency deviation from the carrier value ω_c, then the corresponding equation for the phase-angle variation is

$$\phi = \omega_c t + \frac{k_f\omega_c}{\omega_m} \sin \omega_m t \qquad (8)$$

Note in Eq. (8), the ratio $k_f\omega_c/\omega_m$, which indicates that the peak *phase-angle* deviation is directly proportional to the *peak frequency* deviation $k_f\omega_c$, and inversely proportional to the modulating frequency ω_m. The function $\cos \omega_m t$ of Eqs. (6) and (6a) has been changed to $\sin \omega_m t$ in Eq. (8). This is because the rate of change of a sine function is a cosine function (which is a sine function 90° leading), on the basis that a sine function changes most rapidly when passing through its zero values, which means that the *rate of change* of a sine function has its peak values where the sine function has its zero values.

This is illustrated in Fig. 4. For convenience, the quantity $k_f\omega_c/\omega_m$ has been denoted by the single symbol m_f, known as the *frequency-modulation factor*, in contradistinction to k_f, which is known as the *frequency-modulation constant*. Note from Fig. 4 the

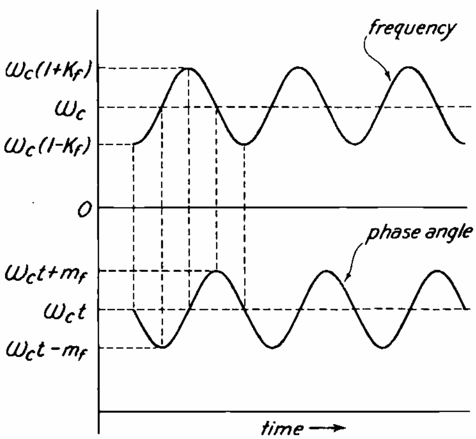


Fig. 4.—Relation between sinusoidal frequency variation and corresponding phase variation.

maximum deviations of the frequency from the carrier value ω_c, and the maximum deviations in the phase angle from its nominal value $\omega_c t$. In addition, note the alignment between the two waves: the upper one leads the lower one by 90°.

Now that the phase angle ϕ has been determined corresponding to the frequency variation expressed by Eqs. (6) and (6a), the final equation for the frequency- (or phase-) modulated wave can be written. It is

$$i = I \sin[\omega_c t + m_f \sin \omega_m t] \quad (9)$$

This equation can then be expanded by trigonometrical and other mathematical methods to exhibit the infinite number of side bands it contains. Before doing so, however, it will be of value to discuss the practical implications of the relationship between phase and frequency derived above.

Phase and Frequency Modulation.—It has just been shown that frequency modulation is accompanied by phase modulation, and similarly, phase modulation is accompanied by frequency modulation. The question naturally arises, "When do we call the system f.m., and when do we call it p.m.?"

The answer is to be found in reference to the modulating wave. Briefly, if the *frequency* deviation is proportional to the strength or amplitude of the modulation wave, then the system is f.m. Under these conditions, if the amplitude of the modulation wave is the *same* at all modulating frequencies, then the frequency deviation will be the *same* at all modulating frequencies, while the number of deviations per second will be equal to the modulating frequency. The accompanying *phase-angle variation* in this case will not be the same at all modulating frequencies, but instead will vary *inversely* as the modulating frequency.

If the *phase-angle* variation or deviation is proportional to the modulation amplitude, then the system is p.m. In this case the *frequency deviation* will increase *directly in proportion* to the modulation frequency: a thousand-cycle note will produce ten times the frequency deviation that a one-hundred cycle note *of the same amplitude* will produce.

In general f.m. is preferred to p.m. in that the former fixes the maximum deviation regardless of the modulating frequency, and this in turn facilitates the design of the band-pass networks that have to transmit the modulated wave. However, the actual transmitter may involve a design in which inherently the phase-angle variation rather than the frequency variation is proportional to the modulation amplitude.

This is the case of the Armstrong and the G.E. phasitron tube systems. Here the audio modulating voltage inherently produces a phase-angle variation from the nominal value, or phase shift, that is proportional to the amplitude of the audio voltage. (The modulating systems will be described in greater detail in a later assignment.) Such an inherent relationship will automatically produce phase rather than frequency modulation; the frequency deviation will increase directly with the audio frequency.

To prevent this effect and to obtain frequency modulation, the phase-angle deviation must decrease as the audio frequency increases in order to keep the frequency deviation constant. This is accomplished in the Armstrong system by attenuating the higher audio frequencies so that the amplitude of the audio voltage input to the transmitter will vary *inversely* as the audio frequency, whereupon the phase-angle deviation will vary likewise, and the frequency deviation will be independent of the audio frequency. This is shown in Fig. 5. The curve is an equilateral hyperbola; this is the graph for functions of the form

$$y = k/x \qquad (10)$$

where k is a constant of proportionality.

In the G.E. system a magnetic field applied to the phasitron tube

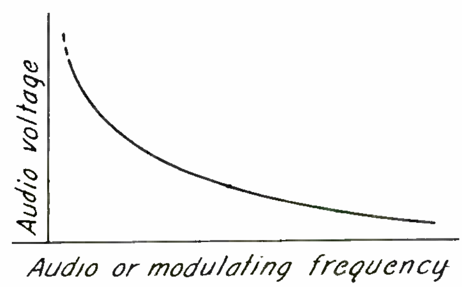


Fig. 5.—Audio-frequency response curve required to give f.m. in the Armstrong system.

produces the desired phase shift in the carrier wave. The magnetic field is produced by passing the audio current through a coil surrounding the tube. The coil is predominantly inductive, hence the audio current is given by

$$I = E/j\omega_m L \qquad (11)$$

where I is the audio current,
E is the (sine-wave) audio frequency voltage,
L is the inductance of the coil,

and ω_m is the audio or modulating frequency.

For a fixed amplitude of E, I varies *inversely* as ω_m, as is evident from Eq. (11). Hence, the higher the audio frequency, the smaller is I. Since the flux produced by the coil is directly proportional to I, it is clear that flux will also vary inversely as ω_m. Finally, since the carrier phase shift is directly proportional to the flux, it, too, will vary inversely with ω_m. Therefore, by means of a simple coil, frequency modulation can be obtained from a system that is inherently phase modulation.

In another system, an oscillator tank circuit has a tube paralleling it. The grid and plate of the tube are so connected to one another that it acts as a capacitive or inductive reactance, as desired. It is therefore known as a reactance tube. When connected in parallel with the tank circuit, it can be arranged to act as a shunt inductance (Fig. 6), and thus raise the frequency of the oscillator. The extent to which it shunts and hence lowers the tank inductance depends upon its G_m. This in turn varies with the audio voltage injected into its grid circuit via R_g since the tube is of the supercontrol type.

At the same time the grid is coupled to the plate via C_g to the R-C network shown. The latter network makes the tube function as a shunt inductance, suggested by the dotted lines in Fig. 6. If R and C are reversed, the tube functions as a variable capacitor.

In either case—variable capacitor or variable inductance—the reactance tube causes the oscillator frequency, rather than the phase angle, to vary at the audio modulating rate. Hence, this type of circuit inherently produces f.m. rather than p.m., and does not require the audio characteristic of Fig. 5.

Preemphasis and Deemphasis.—Farther on in this assignment there will be analyzed the noise characteristics of an f.m. system. It will there be shown that the higher the noise frequency component, the greater its output. At the same time, it is a characteristic of ordinary speech and music that very little energy is present in the higher audio-frequency components. Hence, in the output of the f.m. receiver, the signal-to-noise ratio for the higher audio

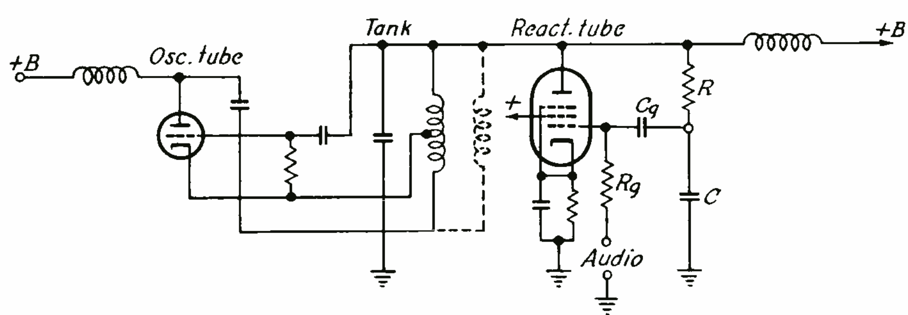
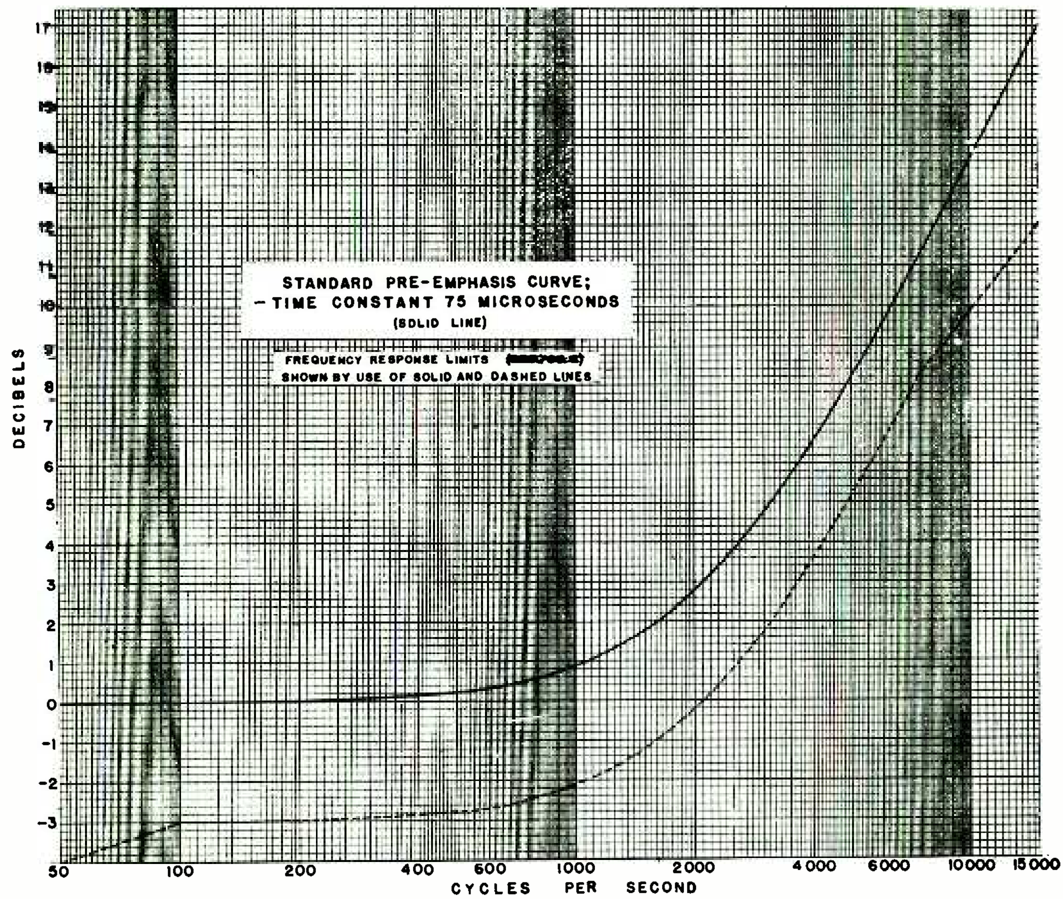


Fig. 6.—Reactance tube used as a shunt inductance to the tank circuit.

frequencies will not be as favorable as for the lower-frequency components.

To counteract this effect, it has been agreed and standardized to accentuate the high-frequency response of the audio amplifier in the transmitter (preemphasis), and then to attenuate the high-frequency response in the receiver (deemphasis), so that the overall response is flat. This is an additional variation in the audio response curve that is superimposed on the high-frequency attenuation shown in Fig. 5, as required for the Armstrong system.

The preemphasis curve is shown in Fig. 7. It will be observed that the output rises about 17 db from 50 to 15,000 c.p.s. The form of the curve is specified to



(Courtesy F.C.C.)

Fig. 7.—Preemphasis curve showing limits for transmitter design.

be the same as that for the variation in *impedance* with frequency of a series R-L circuit whose time constant is 75 μsec. The series impedance in complex form is

$$Z = R + j\omega L$$

$$= R(1 + j\omega L/R) \qquad (12)$$

Since L/R has the dimensions of time, it can be denoted by the time constant T_L, so that Eq. (12) becomes

$$Z = R(1 + j\omega T_L) \qquad (13)$$

Since T_L is chosen as 75 μsec, Eq. (13) becomes

$$Z = R(1 + j\omega\, 75 \times 10^{-6})$$

or in absolute value

$$|Z| = R\sqrt{(1 + \omega^2(75 \times 10^{-6})^2} \qquad (14)$$

On a db basis (remembering that $|Z|$ represents the ratio of output to input *voltage* rather than *power*)

$$20 \log E_i/E_o = 20 \log |Z| = 20 \log R + 20 \log \sqrt{1 + \omega^2(75 \times 10^{-6})^2}$$

$$= 20 \log R + 10 \log[1 + \omega^2(75 \times 10^{-6})^2] \qquad (15)$$

In the receiver, a circuit as shown in Fig. 8 is employed to

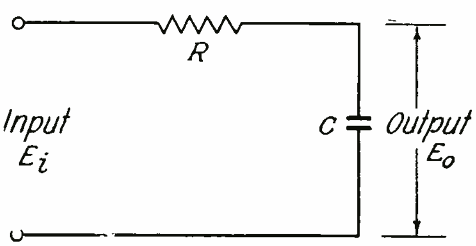


Fig. 8.—Standard deemphasis network for an f.m. amplifier.

attenuate or deemphasize the output voltage. The ratio of E_o/E_i is equal to the ratio of the reactance of C to the total impedance, or

$$E_o/E_i = \frac{1/j\omega C}{R + (1/j\omega C)} = \frac{1}{1 + j\omega CR} \qquad (16)$$

Let CR = T_c, the time constant for this R-C circuit. Then Eq. (16) can be written as

$$E_o/E_i = \frac{1}{1 + j\omega T_c} \qquad (17)$$

and in absolute value

$$|E_o/E_i| = 1/\sqrt{1 + \omega^2 T_c^2} \qquad (18)$$

This is the reciprocal of Eq. (14), except for the numerical value of R in the latter equation. This means that the *overall* transmission including both transmitter and receiver is the *product* of the individual transmission factors, or

$$|E_o/E_i|\,(\text{overall}) = \frac{R\sqrt{1 + \omega^2 T_L^2}}{\sqrt{1 + \omega^2 T_c^2}} \qquad (19)$$

If both T_L and T_c are made equal to 75 μsec., the two expressions under the square root signs cancel, and Eq. (19) reduces to

$$|E_o/E_i|\,(\text{overall}) = R \qquad (20)$$

a constant *independent of frequency*. In short, the overall frequency response will be flat.

The significance of the above analysis is that if the voltage ratio at the particular network used in the transmitter varies with frequency in accordance with Eq. (13), then a very simple R-C circuit as shown in Fig. 8 can be employed in the receiver to restore the response to a uniform value. Since there are thousands of receivers for every transmitter, it will be appreciated that any simplification in the receiver circuit is very much justified from an overall economic viewpoint.

Thus, the network theoretically required to give a voltage ratio in the transmitter of the form of Eq. (13) is shown in Fig. 9. It is clear from an inspection of this figure that

$$\frac{E_o}{E_i} = \frac{R + j\omega L}{R + j\omega L - j\omega L} = \frac{R + j\omega L}{R}$$

$$= 1 + j\omega L/R = 1 + j\omega T_L \qquad (21)$$

which is the reciprocal of Eq. (17) if $T_L = T_c$, in which case an overall flat frequency response will be obtained. However, it will be observed from Fig. 9 that a *negative* inductance (-L) is required.

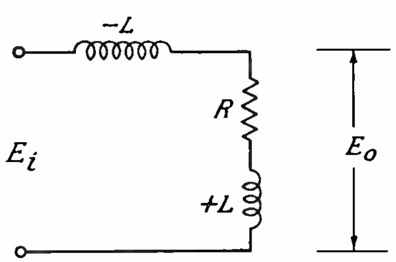


Fig. 9.—Theoretical preemphasis network required in an f.m. transmitter.

Such a circuit element is not physically realizable, and the actual preemphasis circuits (to be discussed in a subsequent assignment) are actually more complicated and yet only approximate in their transmission characteristic.

It is for that reason that limit curves are given in Fig. 7, between which the transmission curve of the actual preemphasis circuit must lie. However, the added complexity in the transmitter is—as previously mentioned—justified on an economic basis by virtue of the simple inexpensive R-C circuit that can be employed in the receiver.

The use of preemphasis of the higher audio frequency actually makes the frequency deviation at these higher frequencies increase over that of the lower frequencies. This means that the carrier phase shift rather than the carrier frequency shift or deviation is more nearly the same at all audio frequencies; i.e., the commercial f.m. system is really more nearly a p.m. system!

Actually, the practical system is somewhere in between an f.m. and a p.m. system, since the preemphasis is not in direct proportion to frequency. However, the preemphasis is built up on a system that is basically f.m. rather than p.m., and is for the purpose of obtaining a better signal-to-noise ratio at the higher audio frequencies. Hence, it is usual to call the actual system an f.m. rather than a p.m. system.

In passing, it is of interest to note that in the case of the Armstrong and G. E. methods of modulation, the preemphasis is superimposed on the initial attenuation curve of Fig. 5; i.e., the actual preemphasis curve is obtained by multiplying corresponding ordinates of Figs. 5 and 7. The net result is the same as for any other type of modulator, such as the reactance tube, and the output from any f.m. receiver will be the same regardless of how the f.m. transmitter is modulated.

F.M. SIDE BANDS.—A complete analysis of the side-band frequencies generated in f.m. involves the use of higher mathematics. However, if some of the latter material is assumed, it is possible to show the number and distribution of these side bands, with the further aid of some simple trigonometrical relations. This has been done in Appendix I.

The equation for a frequency-modulated wave was found to be:

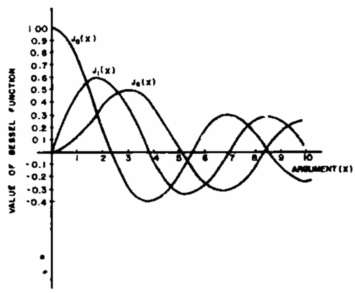
$$i = I \sin[\omega_c t + m_f \sin \omega_m t] \qquad (9)$$

This can be expanded into a series of terms as follows:

$$\begin{aligned} i = I\{ & J_0(m_f)\sin \omega_c t && \text{carrier frequency} \\ & + J_1(m_f)[\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] && \text{first order side bands} \\ & + J_2(m_f)[\sin(\omega_c + 2\omega_m)t + \sin(\omega_c - 2\omega_m)t] && \text{second order side bands} \\ & + J_3(m_f)[\sin(\omega_c + 3\omega_m)t - \sin(\omega_c - 3\omega_m)t] && \text{third order side bands} \\ & + \ldots \text{ etc.}\} && (22) \end{aligned}$$

In this expression, the various J's, such as J_0, J_1, J_2, etc., represent the amplitudes of the various side bands, when multiplied by I. They are known as Bessel functions of the first kind. Thus, J_0 is a Bessel function of the first kind and *zero* order; J_1 is a Bessel function of the first kind and first order, and so on.

The *argument* of each of these functions is m_f, just as θ is the argument of sin θ. As m_f varies, $J_0(m_f)$, $J_1(m_f)$, $J_2(m_f)$ all vary in their individual, particular manners. These variations are shown in Fig. 10*. Note that each



(*Courtesy of Communications, by Nathas Marchand.*)

Fig. 10.—Plot of Bessel functions of the first kind and first three orders.

of these functions passes through zero as m_f increases.

It will be recalled that

$$m_f = \frac{\text{frequency deviation}}{\text{modulation frequency}}$$

Suppose a certain audio note has a pitch corresponding to 500 c.p.s., and has a loudness that produces a frequency deviation of the carrier of 1,200 c.p.s.

*A rather complete table of these is given in Jahnke and Ende's "Tables of Functions," Dover Publications.

Then

$$m_f = \frac{1200}{500} = 2.4$$

From Fig. 10 it will be found that $J_0(2.4) = 0$. This means that the amplitude of the carrier frequency for this value of m_f is zero!

This, however, is not so surprising as it may seem at first. Since the amplitude of a frequency-modulated wave is constant, the power output is constant for all degrees of modulation. For some modulation factor m_f, such as 2.4, all the energy (initially in the carrier for no modulation) has gone into the side bands, and no energy is present in the carrier.

For other values of m_f, the first order side bands, whose frequencies are ω_m above and below the carrier frequency ω_c (just as in amplitude modulation), may be equal to zero. For still other values of m_f, the second order, third order, or even higher order side bands may be equal to zero.

The second order side bands involve $\omega_c \pm 2\,\omega_m$; the third order side bands involve $\omega_c \pm 3\,\omega_m$, and so on. Thus, while in a sinusoidally amplitude-modulated wave, only one pair of side bands is produced, and the carrier amplitude remains unchanged for all degrees of modulation; in f.m. an infinite number of side bands are produced, spaced at frequency intervals of ω_m from each other and symmetrically disposed about the carrier. This is shown in Fig. 11.

Observe that the amplitudes of the side bands in general decrease as their order increases. Of course, for some value of m_f, a particular pair may drop out completely, and higher orders still be present, but the general trend is as mentioned above. This is readily apparent from an inspection

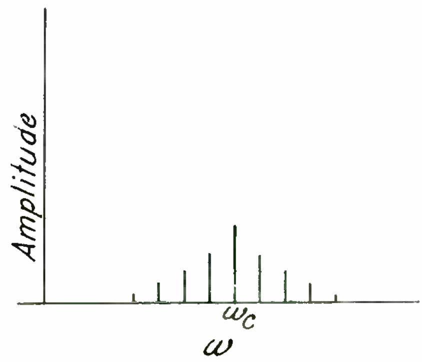


Fig. 11.—Frequency distribution and relative amplitudes of some of the side bands of a sinusoidally frequency-modulated wave.

of Fig. 10, the J_2 order is in general lower than the J_1, and the J_1 is lower than the J_0.

Because the amplitudes become practically negligible for the high order side bands, their omission does not cause serious distortion of the f.m. wave. This means that in spite of the fact that theoretically an infinite number of side bands are required to be transmitted by any circuit, in actual practice transmission of only a limited number of the lower order pairs is absolutely necessary. For example, in the practical f.m. broadcast system, a maximum deviation of 75 kc to either side of the carrier is permitted; this corresponds to the loudest audio tone that can be handled. The amplifier band width required is but 200 kc, which is not so very much greater than the maximum frequency deviation.

It is of interest to compare the loudest tones than can be handled (corresponding to a deviation of 75 kc) for frequencies of 50 c.p.s. and 15,000 c.p.s., the two limits of the audio spectrum as normally employed in f.m. For 50 c.p.s., the value of m_f is 75,000/50 = 1,500, a rather large number. All the corresponding Bessel coefficients are small (as can be noted from the trend of the curves in Fig. 10 as m_f increases to large values). Thus, even the J_0, J_1, and J_2 coefficients are small.

These small amplitude side bands, however, are spaced but 50 cycles apart in the spectrum, hence a large number can be accommodated in the 200-kc band width. On the other hand, for the 15,000-cycle note, m_f = 75,000/15,000 = 5, a much smaller number. In this case $J_0(5)$, $J_1(5)$, and even $J_2(5)$ are appreciable, although the successive higher order coefficients become smaller and smaller.

This means that the lower order side bands are important, and the higher order much less so. However, since the spacing in the spectrum is here 15,000-cycles for the various orders, only a few can be accommodated in the 200-kc band width, so that it is fortunate that the higher order side bands drop more rapidly in amplitude for this high-frequency note. The relative spacing and amplitudes are shown (not to scale) in Fig. 11.

VECTOR REPRESENTATION.—The previous analysis can be made more vivid and clear by the use of rotating vectors to represent the various forms of modulation. In addition, the use of vectors facilitates the analysis of noise and other interference phenomena.

Consider an unmodulated carrier whose equation is

$$i = I \sin \omega_c t \qquad (23)$$

As has been stated previously in this course, such a sinusoidal function of time can be represented by a counterclockwise rotating vector which makes $\omega_c/2\pi$ revolutions per sec., and thereby sweeps out ω_c radians per sec. However, ordinarily in a vector diagram, all vectors are assumed to be of the same frequency, and so maintain their relative phase angles at all times. A single vector diagram therefore represents the relations between them—namely, magnitude and phase for all moments of time.

Amplitude Modulation.—Suppose the carrier is amplitude modulated with a sine wave. As was shown previously, it develops two side bands, each of magnitude 1/2 mI, where I is the carrier amplitude, and m is the per cent, or rather fraction, of modulation, and has a maximum value of unity, so that the maximum amplitudes of the two side bands, for 100 per cent modulation, are each I/2.

If ω_m is the modulating frequency, then the upper side band has a frequency of $(\omega_c + \omega_m)$; and the lower side band has a frequency of $(\omega_c - \omega_m)$. It would therefore appear that a single vector diagram cannot be drawn for these three vectors, since they are of different frequencies.

This is true in the strict sense of the definition of vectors; however, a modified interpretation permits such a diagram to be drawn. Suppose the observer were to assume that he rotates in synchronism with the carrier vector. Then the carrier vector would appear to be stationary.

The upper side-band vector (u.s.b.), however, would appear to rotate *counterclockwise* at a rate of $\omega_m/2\pi$ revolutions per sec., and the lower side band (l.s.b.) would appear to rotate *clockwise* at the same rate of $\omega_m/2\pi$ revolutions per sec. At some instant of time the vectors would appear as in Fig. 12(A); at another instant of time as in 12(B); at a third instant of time as in

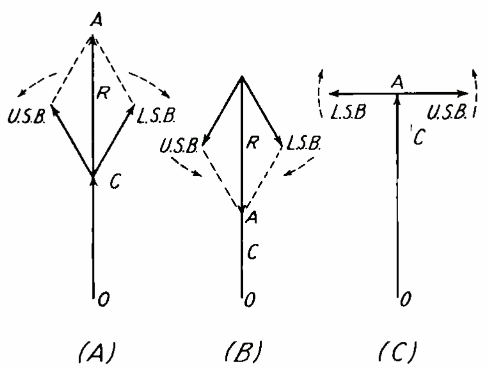


Fig. 12.—Representation of an amplitude-modulated wave by a carrier vector C, an upper side-band vector u.s.b., and a lower side-band vector, l.s.b.

12(C), and so on. At all times u.s.b. and l.s.b. make equal angles with respect to C. The resultant R of u.s.b. and l.s.b. is therefore *at all times collinear with C* (in line with C). Hence, R either adds or subtracts in line with C to give the *overall resultant* OA which is therefore also collinear with C.

The significance of this is that the overall resultant OA appears to the rotating observer to be stationary in space the same as C, even though u.s.b. and l.s.b. both appear to rotate in opposite directions. The fact that OA appears stationary means that it is at all times at carrier frequency; in amplitude modulation the frequency appears constant.

The amplitude of OA, however, varies at the modulating frequency, which is that of the two side bands relative to C. Thus, in Fig. 12(A), OA exceeds OC; in 12(B), it is less than C; and in 12(C) it is momentarily equal to C. At some other instant of time, when u.s.b. and l.s.b. are both facing upward and in line with C, their resultant R is twice either in amplitude, and the overall resultant OA is a maximum (peak value of the modulation envelope). At an instant 180° away in the modulating cycle, the two side-band vectors will be facing downward and in line with C; OA and hence the modulation

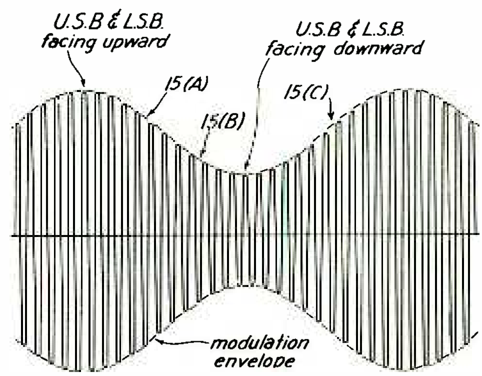


Fig. 13.—Modulated wave showing points on its envelope corresponding to Fig. 12(A), (B), and (C).

envelope will then be at a minimum. These positions are all indicated in Fig. 13.

Thus, the vector diagram of Fig. 12, or rather a series of such diagrams, can be used to portray an amplitude-modulated wave. One diagram, however, is usually sufficient; from inspection of it one can tell how the overall resultant is going to vary.

Frequency Modulation.—Next consider a frequency-modulated carrier. In this case—as shown in Eq. (22)—an infinite number of pairs of side bands are produced. Each pair of frequencies such as $(\omega_c + 3\,\omega_m)$ and $(\omega_c - 3\,\omega_m)$, for example, can be combined into a resultant similar to R in Fig. 12. Then all of these resultants can be combined with the carrier vector to give the final overall resultant similar to OA in Fig. 12.

However, there is one important difference between this vector combination and that for amplitude modulation. In amplitude modulation (refer to Eq. 5), the carrier involved $\sin\,\omega_c t$ and the side bands $\mp\cos\,(\omega_c \pm \omega_m)t$. The side bands thus give rise to a resultant R (Fig. 12) that is *collinear* with C.

In f.m., the first-order side bands (refer to Eq. 22), involve $\pm\sin(\omega_c \pm \omega_m)$ instead of the $\mp$ cosine functions. Hence, the lower side band is reversed in direction, and appears in the third quadrant at a moment when upper side band appears in the second quadrant, as is illustrated in Fig. 14. *Their resultant is at right angles to the carrier vector.* This is extremely important, for as shown in Fig. 14, the combination of the carrier and the two first order side bands gives rise to vector OA, which is only slightly longer than the carrier vector C, *but is no longer collinear with C.*

This means that as the side-band vectors rotate, and maintain equal angles with a line *perpendicular* to C, they cause the overall resultant OA to swing to and fro with respect to C. The opposite extreme position is OA'. Thus, the overall resultant

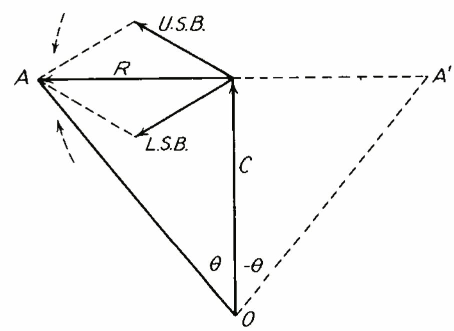


Fig. 14.—Combination of carrier vector with two first-order side bands to give approximate frequency modulation.

oscillates about the carrier vector C at a rate determined by the modulating frequency, and the extreme limits of its oscillation are $\pm\theta$.

A general rule is that if the carrier is a sine function, then if the side bands are cosine functions of unlike sign, or sine functions of like sign, their resultant is collinear with the carrier vector, so that a.m. is obtained; if they are cosine functions of like sign, or sine functions of unlike sign, their resultant is at right angles to the carrier vector, and f.m. is obtained.

Angle θ is the phase shift of the overall resultant with respect to the carrier, and represents the phase modulation that the 90° shift in the side bands has produced. The rate of change of θ with respect to time represents the frequency modulation that inherently accompanies phase modulation, and vice-versa.

However, true f.m. does not

involve any change in amplitude of the wave, whereas Fig. 14 clearly shows that the overall resultant varies from a maximum value OA = OA′ to a minimum value of C. This is because only the first order side bands have been added to the carrier vector. If all the infinite number of side bands are added, an overall vector will be obtained whose length is constant, but whose phase angle varies.

For example, suppose the second-order side bands are added to the carrier and first-order side bands. Note, from Eq. (22), that the second-order side bands involve $\sin(\omega_c \pm 2\omega_m)t$. The two positive sine function gives rise to side-band vectors that combine to give a resultant *collinear* with the carrier; in short, the resultant R_2, Fig. 15, of the second-order side bands is perpendicular to the resultant R_1 of the first-order side bands, and parallel to C.

As is clear from the figure, the overall resultant OA is more nearly equal to the amplitude of the unmodulated carrier, but at a phase angle θ with respect to it.

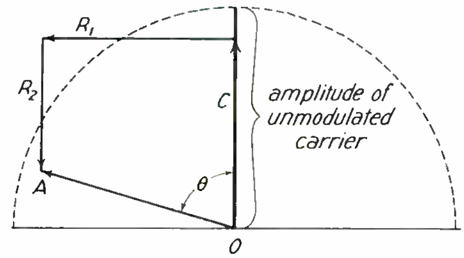


Fig. 15.—Combination of carrier, first-order, and second-order side bands.

For true f.m. or p.m., the tip of the overall resultant should at all times lie on the dotted-line circle; this means its amplitude will at all times be constant, but its phase angle and frequency will shift at the modulation rate.

As higher order side bands are added, this constancy of amplitude

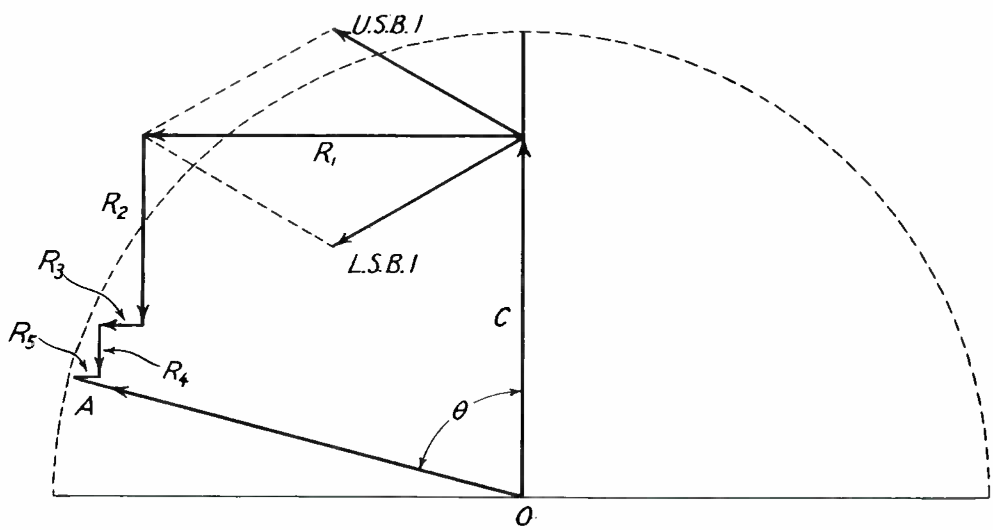


Fig. 16.—Combination of five pairs of side-band vectors with the carrier.

is more closely obtained. Mention was made previously that the amplitudes of the higher-order side bands progressively decrease. Hence, as indicated in Fig. 16, the addition of just a few pairs of side bands (five pairs in the figure) results in nearly perfect f.m. or p.m. (For clarity, only the first-order side bands u.s.b.1 and l.s.b.1 and their resultant R_1 has been drawn; for the others, only the resultants are shown.)

Fig. 16 is very illuminating with respect to the behavior of the components given in Eq. (22). Thus, it shows that the amplitudes of the higher-order side bands should be less than those of the lower-order side bands in order to piece out the overall resultant so that it just touches the circle.

It also shows why a finite band width is sufficient in practice to give satisfactory f.m. If higher-order side bands are attenuated or cut off, the overall resultant OA may not quite touch the circle at all times, but its phase (and frequency) deviation from the carrier will not be materially affected by such attenuation.

Armstrong Modulation System.—Fig. 14 indicates a practical method of accomplishing frequency modulation (the Armstrong system). Although modulation will be discussed in another assignment, it will be of interest to discuss the principles of the Armstrong method here. A carrier wave of 200 kc is first amplitude-modulated in a special push-pull amplifier whose grids are in parallel for r.f. but in push-pull for a.f., and whose plates are connected in push-pull with the load impedance.

As a result, the output contains the side bands, *but not the carrier component*. Assume for simplicity that the modulation is a single sine wave. Then just two side bands are produced. These are then fed into a 90° phase-shift network, and combined with the original carrier, as is shown in Fig. 17, together with further operations on the resulting wave.

The process up to this point can be represented vectorially as in Fig. 18. In (A) is shown the ordinary amplitude-modulated wave, as produced in the usual modulator. In the case of the special balanced modulator, however, the output is

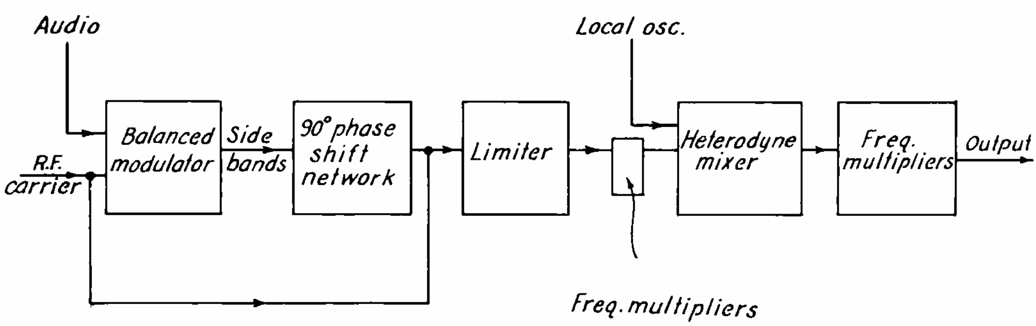


Fig. 17.—Block diagram of Armstrong system of frequency modulation.

as shown in (B); only the side bands appear in the output, and the carrier is suppressed by the action of this type of circuit. The phase-shift network produces the 90° shift as shown in (C), so that when the side bands are now combined with the original carrier C, as in (D); essentially phase and frequency modulation are produced. (This diagram is identical with Fig. 14.)

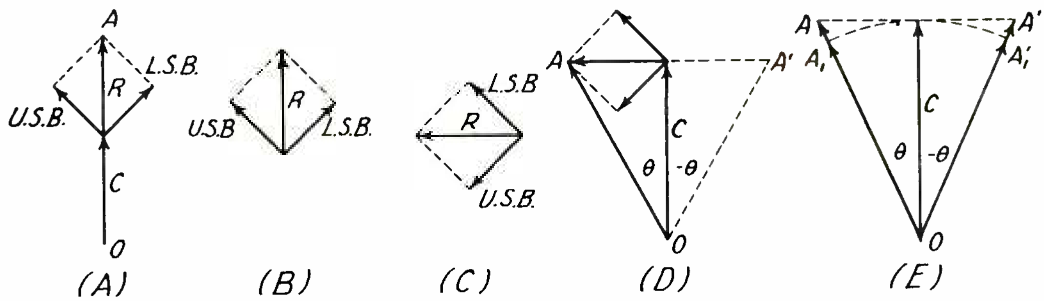


Fig. 18.—Successive steps in producing f.m. by means of the Armstrong method.

Note that the amplitude of the overall resultant varies as well as the phase-shift angle θ as the side-band vectors rotate. The maximum values are OA = OA′; the minimum value is C. There are two variations per rotation of the side-band vectors, and hence two per modulation cycle. Thus the amplitude modulation that is inherently produced by this method is of double the modulating frequency, or a second (and also higher even) harmonics of the incoming audio.

Since amplitude modulation is not desired, and since it is a distortion product anyway, it is removed by a limiter circuit. This circuit clips the wave down to a constant amplitude regardless of its phase or frequency, and operates by virtue of the cutoff property of a vacuum tube. The resulting clipped wave contains harmonics of the *r.f.* frequency and its side bands, but these can be eliminated by passing the clipped wave through ordinary tuned circuits.

The result is a wave that is essentially frequency modulated. As shown in Fig. 18(E), the amplitude, which previously varied from OA to C to OA′, now remains constant at the value $OA_1 = C = OA_1'$, while the phase shifts through a range of $\pm\theta$ from the carrier frequency.

For small amplitudes of the side bands relative to the carrier, the angle θ will be small. As shown in Fig. 19, the resultant

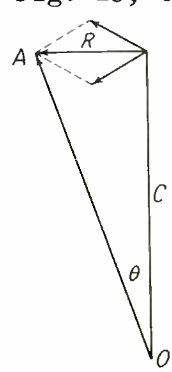


Fig. 19.—Geometry involved in the Armstrong f.m. system.

R of the two side bands has a maximum value equal to twice the amplitude of either side band. The angle θ is clearly such that

$$\tan \theta = R/C \qquad (24)$$

If R is small compared to C, then tan θ is small, and in this case the angle in radians is very nearly equal to its tangent, or

$$\theta \cong R/C \qquad (25)$$

Since the carrier amplitude C is maintained constant, Eq. (25) shows that θ is directly proportional to R alone, the resultant of the two side bands. It was previously shown that R represents the amount by which the envelope in a.m. exceeds the unmodulated carrier (see Fig. 13); that is, R is directly proportional to the input modulation voltage.

Hence θ, the phase shift of the resultant wave with respect to the carrier wave, is directly proportional to the input modulation voltage, and therefore this system produces phase modulation, as was indicated earlier.

Suppose the modulation (audio) frequency is 100 c.p.s., and θ maximum is 25°. Then the number of degrees per sec. variation from the carrier value will be 100 × 25° = 2,500° per sec. Next suppose the modulation frequency is doubled: 200 c.p.s. If θ remained the same, the number of degrees per sec. variation would be 200 × 25 = 5,000° per second. Since frequency is proportional to degrees per sec., it is clear that if the phase shift is the same for all audio frequencies, the frequency deviation will be proportional to the audio frequency: a necessary consequence of phase modulation.

To obtain f.m.; i.e., have the frequency deviation remain the same at all audio frequencies, it is necessary to have the amplitude of the audio input to the modulator vary inversely with the frequency, as was indicated in Fig. 5, in which case the phase shift θ will also vary inversely with the frequency, and the frequency deviation will then be independent of the audio frequency. The result will then be f.m. instead of p.m. This conclusion was reached earlier in this assignment; the vector representation helps to make it clearer.

In order for Eq. (25) to hold to any reasonable degree of approximation, θ must be no more than about 25.5° at the lowest audio frequency—say 20 c.p.s.* for a maximum distortion of 5% (departure of linearity between θ and the modulating voltage). At higher audio frequencies θ will be smaller—as just explained—in order to obtain f.m. instead of p.m., and the distortion will become progressively smaller.

The restriction of θ to a small value because of distortion considerations precludes obtaining a large frequency deviation. On the other hand, a large deviation is necessary if the advantage of suppression of noise and static inherent in this system is to be obtained. It is therefore necessary to expand the initially small deviation to the 75 kc desired.

Merely multiplying the fre-

*See Jaffe, D.L., "Armstrong's Frequency Modulator," *Proc. I.R.E.*, April 1938, for a discussion of the distortion produced in the Armstrong system.

quency by means of vacuum tube stages will not produce the desired result in that the carrier frequency will be too high by the time the desired 75-kc deviation is obtained. Hence only a certain amount of multiplication up to 12,800 kc is obtained (64 times). Then this output is mixed with a 11,900-kc (local) oscillator, and a difference beat frequency of 12,800 - 11,900 = 900 kc obtained.

This 900-kc carrier, however, contains the same phase shift and frequency deviation as the 12,800-kc wave, and these are 64 times as great as initially produced. Thus, on a carrier but 900/200 = 4.5 times the initial carrier, there is obtained frequency modulation 64 times as great. Subsequent multiplication can bring the carrier up to the desired value, and its frequency deviation up to 75 kc.

REDUCTION IN INTERFERENCE

It was mentioned previously that f.m. can produce a higher signal-to-noise ratio at the output of the system than occurs at the input, and that static and other interference can be rendered practically negligible. This is true if the frequency deviation is sufficiently great, and a limiter, or a discriminator that is insensitive to amplitude variations, is employed.

The interference to be discussed here consists of:

1. Random thermal noise.
2. Common-channel interference.
3. Adjacent-channel interference.
4. Impulse noise.

THERMAL NOISE.—The basis of all noise reduction can be understood by a study of the effect of thermal noise on an f.m. wave. Thermal noise is a random voltage generated in any resistor owing to the haphazard excursions of the free electrons in the resistor at room temperatures. At some moment of time there may be more electrons traveling in one direction through the resistor than in the opposite direction; at another moment of time an excess of electrons may be travelling in the opposite direction. Such motions represent current flow: as shown in

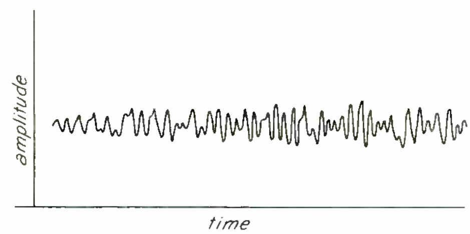


Fig. 20.—Picture of thermal noise, showing its random nature with time.

Fig. 20 this flow, when plotted against time, is completely random in nature.

The quantitative nature of noise is treated more completely in the ultra-high frequency section of this course. Suffice it to say here that when thermal noise is analyzed on a frequency basis, it is found to consist of sinusoidal components completely filling the spectrum, from the lowest to the highest frequency employed. These components are all about equal in amplitude; a more accurate statement is that the energy contained in any one band width is the same as in any other equal band width located in another part of the spectrum, and is proportional to the resistance and the absolute temperature.

The phase of the components is completely random in nature. Hence, if one wishes to study the noise in a given band width centered on some specific carrier, he must take each component individually into consideration, and not try to combine two into a side-band pair with respect to the carrier. This will be the basis of the analysis given in this assignment.

In passing, it will be of interest to note that in the standard broadcast band, atmospheric static (which is different in characteristics from thermal noise) is so strong that the broadcast wave must be of considerable strength successfully to override it. In doing so, it can successfully override the weaker thermal noise generated in the receiver, so that thermal noise is not an important factor here.

At the higher f.m. carrier frequencies—around 100 mc—atmospheric noise is negligible, although man-made static, such as from the ignition systems of automobiles and airplanes, may be quite appreciable. In this range of the spectrum, however, antennas—such as half-wave dipoles—are quite small. The energy picked up by these is from an area in space on the order of a quarter-wave length around the dipole, and hence quite small. As a result, the signal picked up is not so very much greater than the thermal noise generated in the set, and hence the latter is of considerable importance in design considerations.

In analyzing the performance of an f.m. wave in the presence of thermal noise, consider for simplicity an unmodulated carrier. This passes through the receiver circuits accompanied by the various noise components that can pass through the pass band of the circuits. In Fig. 21, C represents the carrier vector, and N represents one sinusoidal component of the noise voltage.

Suppose the noise component N is of a higher frequency than C.

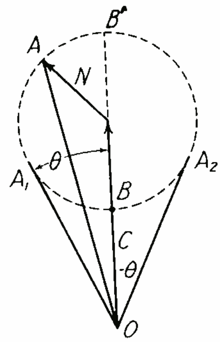


Fig. 21.—Combination of an unmodulated carrier and a noise component.

It will therefore rotate counterclockwise on the diagram. As shown, it describes a circle about the tip of C. The overall resultant is OA; this vector clearly exhibits a phase oscillation to either side of C, as well as amplitude modulation.

Thus the maximum phase swing is $\pm\theta$, corresponding to extremes positions OA_1 and OA_2. The maximum amplitude excursions are from a minimum of OB to a maximum of OB′. Suppose a limiter or an amplitude-insensitive detector is employed. Then the amplitude variations will either be eliminated (by the limiter) or will be ineffective in producing any audio output.

The phase modulation $\pm\theta$, however, is not eliminated by a limiter, *and will produce an audible noise output.* The frequency of this noise component at the output of the detector is the number of $\pm\theta$ swings per second. This, as is clear from a study of Fig. 21, is equal to the number of revolutions of N about C; that is, to the excess in frequency of N over that of C. For example, if C has a frequency of 100 mc, and N has a frequency of 100.001 mc, then there will be $(100.001 - 100)10^6 =$ 1,000 c.p.s. alternations in θ, and the detector output will be a 1,000-cycle note.

Suppose next that N has a frequency of 100.002 mc. In this case the output will be a 2,000-cycle note. Recall that the various components of thermal noise are of equal amplitude. Therefore, the maximum phase shift or deviation will still be the same, *but the rate of change of θ, or the frequency deviation, will be twice as great.* This means that at the output of the detector or discriminator, the 2,000-cycle noise component will have double the amplitude of the 1,000-cycle component.

Hence, the higher the frequency of N relative to C, the higher is the frequency of the *output* noise component, *and the greater is its amplitude.* This means that the noise output in general sounds higher pitched for f.m. (more like a hiss) than it does for a.m. For a.m., the noise output amplitude is directly proportional to the input amplitude of N and independent of its frequency. For example, in a.m., the signal-to-noise ratio at the output is the same as at the input, namely, N/C. In the case of f.m., the ratio is not so simply expressed, but is not at all difficult to derive.

In Fig. 22 is shown the noise vector, N, the carrier vector, C, and their resultant R, which varies with time as N rotates around the

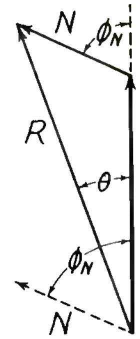


Fig. 22.—Vector diagram used to derive signal/noise ratio for f.m.

tip of C at a frequency equal to the difference between that of C and N. At some particular moment the angle between N and C is ϕ_N, and $\phi_N = 2\pi(f_N - f_C)t$ where f_N = frequency of noise component, and f_c = carrier frequency.

There is a theorem in trigonometry that states that the length of the resultant of two vectors is equal to the square root of the sum of their squares minus twice their product times the cosine of their included angle. As applied to Fig. 22.

$$R = \sqrt{C^2 + N^2 \pm 2\,NC\cos\phi_N}$$
$$= \sqrt{C^2 + N^2 \pm 2\,NC\cos 2\pi(f_N - f_c)t} \quad (26)$$

This formula gives the variation in amplitude of R with time as N rotates counterclockwise with respect to C. This variation in the amplitude can be eliminated by the use of a limiter, and hence need produce no noise at the output of the f.m. detector.

However, R makes an angle θ with respect to C that also varies

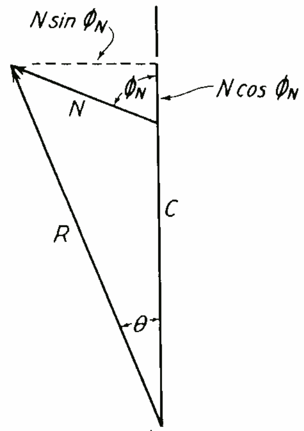


Fig. 23.—Vector diagram used to evaluate angle θ.

with time, and represents frequency modulation. This angle can be derived by referring to Fig. 23. It is clear that

$$\tan\theta = \frac{N \sin\phi_N}{C + N\cos\phi_N}$$

$$= \frac{N \sin 2\pi(f_N - f_C)t}{C + N\cos 2\pi(f_N - f_C)t} \quad (27)$$

If the noise amplitude N is small relative to C, N cos ϕ_N is negligible compared to C, and θ is so small that tan $\theta \cong \theta$, so that Eq. 27 reduces to

$$\theta \cong \frac{N}{C}\sin 2\pi(f_N - f_C)t \quad (28)$$

Angle θ is the phase shift of R relative to the unmodulated carrier C, and represents the phase modulation produced by the addition of the noise component N to C. This angle is in general very small. Its maximum value occurs when $\sin 2\pi(f_N - f_C)t = \pm 1$, when it is equal simply to N/C. Suppose N is 0.1 C. Then

$$\theta \cong 0.1 \text{ radian} = 5.73°$$

The phase shift that occurs in the carrier C when it is fully modulated by an audio signal may be several revolutions of C first in one direction and then in the opposite direction relative to its unmodulated position, and 5.73° is a very small angle in comparison.

However, in an f.m. system, it is not the phase shift or deviation that produces output, but rather the accompanying frequency deviation, which is the *rate of change* of phase shift with respect to time. For θ as given by Eq. (28), the corresponding frequency deviation is*

$$f_D = \frac{N}{C}(f_N - f_C)\cos 2\pi(f_N - f_C)t \quad (29)$$

Note the factor $(f_N - f_C)$ in the formula for the frequency deviation. It is absent in Eq. (28) for the phase deviation. This is to be expected: the frequency deviation f_D is proportional not only to the phase deviation producing it, but also to the number of times per second or frequency $(f_N - f_C)$ that θ varies about the unmodulated position of the carrier C. Thus Eq. (29) substantiates the physical reasoning presented earlier to the effect that the higher f_N is relative to f_C, the greater is the amplitude of the noise output, since the output is proportional to f_D.

*In differential calculus notation, $f = (1/2\pi)\, d\theta/dt$.

The highest noise frequency f_N that can be passed in an actual system is one-half the i.f. band width above (or below) the carrier frequency f_c. Since the band width is 200 kc, f_N can exceed f_c by at most 100 kc. However, in this case, the *noise output* has a frequency of 100 kc, and is inaudible. The highest value $f_N - f_c$ can have and yet be audible is about 15 kc (upper end of audio band).

This places a maximum value of 15,000 for $(f_N - f_c)$. Suppose, as before, N/C = 0.1. Then, from Eq. (29),

$$f_D = 0.1(15{,}000) = 1{,}500 \text{ c.p.s.}$$

The maximum deviation for a *desired signal* is 75,000 c.p.s. Therefore the modulation factor for the noise is

$$m_f(\text{noise}) = 1{,}500/75{,}000 = 1/50$$

or the noise will be 1/50 = 2 per cent of the maximum signal; i.e.,

$$\text{signal-to-noise ratio} = 50 : 1$$

If amplitude modulation had been employed, 100 per cent signal modulation would be a variation in carrier amplitude between 2C and 0 peak-to-peak. In other words, the modulation peak amplitude would equal C, the carrier amplitude. Hence if N/C = 0.1, the signal-to-noise ratio would be 1/0.1 = 10 : 1.

From this it is clear that f.m. in this particular case has a five-fold advantage over a.m. The advantage is even greater than this, however, The comparison was made for a noise component whose audio output frequency was 15,000 c.p.s. For lower frequency components (f_N closer to f_c), the output will be correspondingly less.

In Fig. 24 is shown the proportional manner in which the noise components increase as their frequency *above* or *below* f_c becomes greater and greater. In the preceding analysis f_N was taken as greater than f_c; exactly the same results are obtained if f_N is less

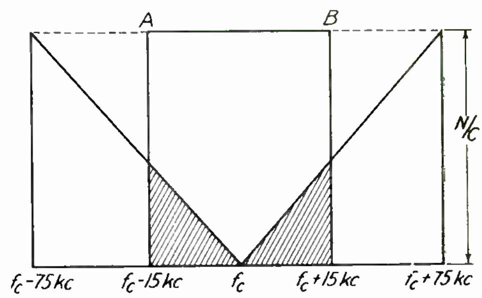


Fig. 24.—Noise-to-signal ratio for f.m. and a.m. systems.

than f_c. Hence, two branches to the curve are obtained, as shown.

Note that the noise-to-signal ratio does not rise to its a.m. value N/C until $f_N = f_c \pm 75$ kc. Since the noise is audible only up to values $f_N = f_c \pm 15$ kc, it is clear that the maximum amplitude is but 15/75 = 1/5 of N/C, and that is for a 15-kc noise component. For lower noise components (f_N closer to f_c), the noise amplitude is much less.

In the case of a.m., the noise-to-signal ratio is constant at the value N/C from f_c - 15 kc to f_c + 15 kc, and is therefore represented by the rectangle whose top is AB. It is clear from the figure that the noise-to-signal ratio for a.m. is much greater than it is for f.m.

The comparison in Fig. 24 is

based on signal amplitudes. If the comparison is made on an energy basis, then the ordinates must all be squared. Rectangle AB remains a rectangle when the ordinates are squared, but the two shaded triangles form a parabola. The *total* noise-to-signal in either case can be found by adding arithmetically the squared ordinates, since energy is not a vector quantity, and the random phase of the individual noise components, which prevents them from being added directly, does not prevent their energy contents (as represented by the squared ordinates), from being so added. This summation is best done by means of the integral calculus, and yields the following result*

$$\frac{S_f/N_f}{S_a/N_a} \text{ (r.m.s. values)} = \sqrt{3}\,\frac{F_D}{F_a} \qquad (30)$$

where S_f/N_f is the signal-to-noise ratio at the output of the f.m. receiver; S_a/N_a is the signal-to-noise ratio at the output of the a.m. receiver, F_D is the peak frequency deviation, and F_a is the maximum audio frequency to be amplified.

According to Eq. (30), if F_D = 75 kc, and F_a = 15 kc, the "improvement" is $\sqrt{3}(75/15)$ = 8.66 times, or the signal-to-noise output for f.m. is 8.66 times that for a.m. Although the analysis was based on the action of noise components on an unmodulated carrier, it holds for a modulated carrier as well, so long as the noise components N are small in comparison with the carrier amplitude C.

It will be of value to look into the above improvement from a physical viewpoint. It was shown that if N is small compared to C, the phase shift θ is only a few degrees. The corresponding noise frequency modulation increases with increase in the noise frequency f_N over the carrier frequency f_c, but a limit is reached when $(f_N - f_c)$ equals 15 kc. Fig. 24 shows how the noise increases in amplitude to either side of f_c, but ends abruptly at $f_c \pm 15$ kc.

The peak frequency deviation produced by a noise component 15 kc higher or lower than f_c is, by Eq. (29),

$$f_D = (N/C)\,(15)\,\text{kc}$$

and this is the maximum value for f_D.

Thus, in the case of noise, maximum detector output, which is proportional to f_D, does not exceed 15 kc (for which N = C and N/C = 1), and the audio noise frequency is then 15 kc. On the other hand, even a 50-cycle modulation *signal* can impart the full 75-kc deviation to the carrier C if it is loud enough. This is because the modulating equipment can produce sufficient phase shift to provide the 75-kc frequency deviation, even if the phase shift occurs but 50 times a second.

The noise component, N, on the other hand, produces its maximum phase shift angle by adding its amplitude at right angles to C at the proper moments in its rotation about C, and if N is small compared to C, the phase shift angle is small, so that the frequency deviation, even at the maximum value of $(f_N - f_c)$ = 15 kc,

*See Crosby, Murray G., "Frequency Modulation Noise Characteristics," *Proc. I.R.E.*, April 1937.

is less than 15 kc.

It is therefore clear that as the peak deviation F_D for the signal is increased for a desired audio band width F_a, the improvement in signal/noise for f.m. over that for a.m. will increase correspondingly, in addition to an inherent advantage given by the factor $\sqrt{3}$ in Eq. (30). The 200-kc i.f. band width necessary for the 75-kc deviation admits a wide band of noise components, but only 15 kc of these to either side of the carrier produce any audible output. On the other hand, the full resources of this band width are available and used to pass even the lowest signal-frequency side bands; they utilize this band because they produce sufficient phase shift and hence frequency deviation to require such a band width. In short, no matter how large the i.f. band width is made to accommodate a large frequency deviation system, the noise can utilize only 15 kc of it; the signal frequencies, all of it.

The action of the f.m. system employing a limiter or discriminator that is insensitive to amplitude variation, is to suppress the noise (or any other similar interference) to a point where it is generally inaudible, whereas an a.m. system has no such suppressing effect. This is based on the assumption, however, that the noise amplitude N is small compared to the carrier amplitude C. If, on the other hand, N is large compared to C, the opposite effect takes place, and the noise tends to suppress the desired signal. A figure often employed is that the carrier power should be about 6 db greater than that of the noise, or the amplitude ratio should be about 2 : 1, in order that suppression of the noise begin to take place. Below this threshhold ratio, the noise begins to come through to a degree comparable to that in a.m. systems, and when the noise amplitude exceeds that of the carrier, the carrier is suppressed, and the results are inferior to an a.m. system.

Returning to the condition of N small compared to C, note that the use of preemphasis and deemphasis circuits produces an even greater signal-to-noise ratio. The reason is that the preemphasis of the wanted signal compensates the high-frequency deemphasis of the signal, but the noise, lacking such compensation, is reduced by the deemphasis circuit to a very low value.

Above about 1,500 c.p.s., the receiver audio response drops practically linearly with frequency. This just about exactly compensates for the tendency of the higher-frequency noise components to give a greater deviation and hence f.m. detector output. As a result, the actual noise output after the deemphasis circuit does not rise beyond 1,500 c.p.s., but remains practically constant.

As a result, the noise diagram of Fig. 24 is modified and appears as shown in Fig. 25. A

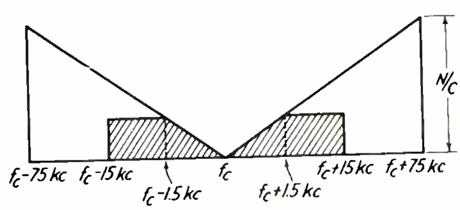


Fig. 25.—Noise-to-signal ratio when deemphasis is used.

simple and approximate ratio for the superiority or improvement of f.m. over a.m. is D/1,500. For a peak deviation of D = 75,000 c.p.s., the ratio becomes 75,000/1,500 = 50 : 1. This should be compared with the previous figure of 8.66 : 1 when deemphasis is not employed.

COMMON-CHANNEL INTERFERENCE.— Common-channel interference is the interference caused to the reception by an f.m. receiver of a stronger or nearer station, by a weaker or more distant station on the same channel. Essentially, the problem is that of a desired carrier and a weaker interfering sinusoidal component. This was analyzed previously in the case of thermal noise: any sinusoidal component of thermal noise is equivalent to the interfering carrier. Eq. (29) therefore holds here, if N be interpreted as the amplitude of the interfering carrier, and f_N as its frequency, which differs from the desired carrier frequency f_c owing to inherent drift between two transmitters unless synchronized.

The modulation factor for this interference is

$$m_f\text{(interference)} = \frac{(f_N - f_c)}{F_D} \times \frac{N}{C} \quad (31)$$

where, as before, F_D is the maximum frequency deviation for the desired modulating signal. If deemphasis in the receiver is taken into account, N is attenuated in the same proportion as $(f_N - f_c)$ increases above a value of about 1,500 c.p.s., so that if the interfering carrier drifts in frequency from f_c by more than 1,500 c.p.s., m_f does not increase.

Actually the interfering carrier N swings back and forth in frequency relative to the desired carrier C, so that an average frequency difference of $(f_N - f_c)$ = 1,500 c.p.s. is a reasonably good approximation. Hence the interference, when deemphasis is employed, is given by

$$m_f\text{(interference)} = \frac{1500\ N}{F_D\ C} \quad (32)$$

or, for F_D = 75 kc, this becomes

$$m_f\text{(interference)} = \frac{1500}{75000} \times \frac{N}{C}$$

$$= \frac{N}{50\ C} \quad (33)$$

The modulation factor m_f gives the fraction of the maximum deviation F_D that the interfering carrier produces, and hence its relative loudness. The signal-to-noise ratio for f.m. is the reciprocal of this, or

$$S_f/N_f = \frac{1}{m_f} = \frac{50\ C}{N} \quad (34)$$

In amplitude modulation, the interfering carrier N would produce a variation of N units in the envelope of C. Signal amplitude modulation of C would produce a variation of C units in the envelope (from 0 to 2 C peak-to-peak). Hence, the a.m. signal-to-noise ratio would be

$$S_a/N_a = C/N \quad (35)$$

The ratio of the two is therefore

$$\frac{S_f/N_f}{S_a/N_a} = \frac{50\ C}{N} \times \frac{N}{C} = 50$$

or the f.m. system produces 1/50 of

the interference in its receiver that is produced in the receiver of the a.m. system.

Such an improvement permits f.m. stations on the same frequency or channel to be spaced geographically much closer than a.m. stations. Also, since f.m. carrier frequencies are much higher than a.m. carrier frequencies, few sky wave reflections are produced to cause interference to stations at great distances from the one in question. Interference-free operation can therefore be expected even when the same channel is used for a large number of stations; this permits better coverage of those sections of the country that can economically justify the cost of a station or stations.

There is, of course, a small area between two stations where their signal strengths are about equal, and here the interference is as bad for both as in the case of a.m. stations. Another phenomena noted is "bursts," owing to the vagaries of propagation, the weaker station may intermittently override the stronger one and take away the program from it in the receiver. Such phenomena must be investigated more thoroughly, and more will be known about this as the service is expanded.

ADJACENT-CHANNEL INTERFERENCE.—The interference caused by a carrier in one channel upon a receiver tuned to a carrier in an adjacent channel is known as adjacent-channel interference. Although the two carriers are sufficiently far apart (200 kc) so that their interaction will produce f.m. above audibility, their adjacent side bands may be sufficiently close in frequency to produce audible, interfering tones.

This effect depends in great measure upon the selectivity of the receiver, and since adjacent channels are involved, i.f. selectivity is involved. Suppose the two signals are as shown in Fig. 26. The

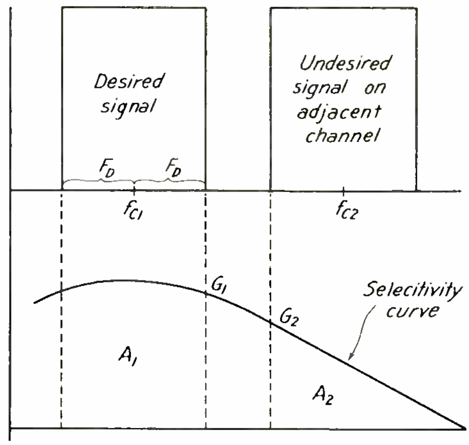


Fig. 26.—Effect of selectivity curve on two signals in adjacent channels.

selectivity curve, shown beneath the two signals, allows them to pass as shown by the two areas A_1 and A_2.

The carrier of frequency f_{c1} swings to and fro through the frequency range $\pm F_D$, and the adjacent carrier of frequency f_{c2} oscillates back and forth in similar manner.

When the desired and undesired signals move toward each other at some instant, they are amplified to the extent of G_1 and G_2, respectively. Suppose their inputs to the receiver are C_1 and C_2, respectively. Then their respective amplitudes at the limiter stage will be C_1G_1 and C_2G_2.

If C_1G_1 is greater than C_2G_2, the adjacent channel interference will be very small: at least 60 db below the signal level.* It is to be emphasized that even though G_1 is greater than G_2 if the receiver has any selectivity, the adjacent signal C_2 may be so strong that C_2G_2 can conceivably exceed C_1G_1.

In actual practice appreciable adjacent channel interference has been experienced. The F.C.C. (Federal Communications Commission) has therefore amended their rules to prohibit the location of stations on alternate frequencies (400 kc apart) in the same general area, and instead have placed them 800 kc apart. The increased separation is of value in permitting the receivers to be less selective and hence cheaper.

IMPULSE NOISE.—Impulse noises are sharp-peaked voltages occurring either at periodic or random moments, depending upon their origin. Unlike random (thermal) noise, impulse noise has a definite spectrum in that the phase of the components are not random, as in the former case. At the peak of the impulse, the sinusoidal components are all in phase, and the height of the impulse is proportional to the band width, whereas the r.m.s. value of random noise is proportional to the square root of the band width.

Thus, the peak value of the impulse noise will increase in direct proportion to the band width, and in the case of f.m., where the width is 200 kc, the peak can easily exceed the desired signal amplitude. The interference may in this case be very appreciable, as for any type of interference that exceeds the signal in amplitude.

If the pulse duration is sufficiently short, it will suppress the signal momentarily, and the effect will hardly be noticeable. This is similar to the action of the noise suppressor in some a.m. receivers, and is therefore not a feature peculiar to f.m. Ordinary impulse noises, such as due to automobile ignition systems, are under ordinary listening conditions more important than thermal noise, which normally is noticeable only when the signal level is extremely low. Hence, the former form of interference is of practical importance. It makes its presence known by a series of isolated clicks or pops, particularly during quiet moments in the program.*

The fact that the peak of the impulse noise increases as the band width is increased indicates that too great a band width is inadvisable, because then the impulse has a better chance of exceeding the signal amplitude and thus "capturing" the receiver (suppressing the signal). This in turn means that the amount of frequency deviation that can be employed is limited: although the greater the deviation, the less effective is thermal and such noises; the greater band width required makes the effect of impulse noise more objectionable.

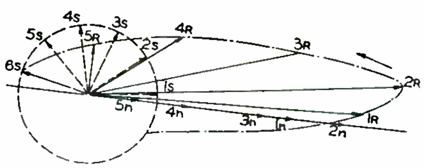
It is of interest to note graphically the action of an im-

*See, for example, S. Goldman, "F-M Noise and Interference," *Electronics*, August 1941.

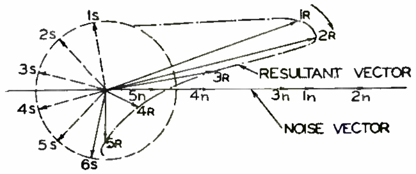
*Smith and Bradley, "The theory of Impulse Noise in Ideal Frequency-Modulation Receivers," *Proc. I.R.E.*, October, 1946.

pulse. The actual shape of the pulse at the discriminator depends practically entirely upon the amplifier characteristics, notably band width. This is because if the impulse is sufficiently short, it sets the tuned circuits into oscillation by the process of shock excitation, like a hammer blow, and the output is then relatively independent of the nature of the input pulse.

As shown in Fig. 27, the impulse may be regarded as a vector of carrier frequency, that grows to a maximum and then decreases to zero. This is characteristic of a tuned circuit or circuits shock-excited. In (A) the noise vector



(A)



(B)

(Courtesy of Proc. of I.R.E., by Smith and Bradley.)

Fig. 27.—Production of clicks or pops by impulse noise on an f.m. receiver.

in its successive changes of growth and decay is designated by 1n, 2n, 3n, 4n, and 5n. The modulated carrier vector is shown simultaneously rotating counterclockwise from its unmodulated position by the series of vectors 1S, 2S, 3S, 4S, 5S, and 6S.

The resultant of the two is given by the vectors 1R, 2R, 3R, 4R, and 5R. It will be observed that this vector follows the carrier vector in phase, but lags somewhat behind it, thereby giving rise to an extraneous click, which is not too objectionable.

Under purely chance conditions, the situation may be as in (B), Fig. 27. Here it will be observed from the R (resultant) vectors that the impulse momentarily rotates clockwise and hence opposite to the direction of rotation of the signal vector S, because the phase between 1n and 1S happens to exceed 90°. Ultimately, in either case, the phase shift reverts to that of the signal carrier, but in (B) there is a permanent loss of 2π radians; the effect is as if a sudden "step" voltage were applied to the audio amplifier and the result is a loud pop.

In the case of such sudden but thereafter constant change in phase, to which corresponds a sudden change but thereafter constant value of detector output voltage, the deemphasis time constant is the major circuit affected, the peak output is about 13 per cent of the signal amplitude corresponding to 100 per cent modulation (75-kc deviation), but this pop is the only sound heard at the time.

In the case of clicks, the change in phase is momentary and less than the deemphasis time constant. In this case its output amplitude is determined mainly by the band width of the audio system, and for 15,000 c.p.s., the amplitude may be at most 4 per cent of the signal amplitude for 100 per cent modulation. In general click interference is far less severe than pop interference, although the occurrence of either is based on pure chance.

If the receiver is mistuned, then the carrier frequency is inherently different from that of the impulse, and considerable frequency pop modulation may occur. This indicates the importance of a precise method of tuning an f.m. receiver, in addition to the importance of eliminating distortion of the program itself by mistuning.

F.M. RECEIVING CIRCUITS

R.F. AMPLIFIERS.—The design of an f.m. receiver follows closely that of an a.m. receiver so far as r.f. amplifier design is concerned. The problems are intensified by the higher carrier frequencies encountered, but image rejection considerations are the major design problem just as in the standard broadcast band.

Adequate band width in the r.f. amplifier, however, is not a major problem. In the 100-mc range, a 200-kc band width is equivalent to a 2-kc band width at 1 mc, and hence is not difficult to obtain. Indeed, with the appreciable input loading developed by the tubes at such frequencies, the problem is to obtain sufficiently high circuit Q's. This is analyzed very thoroughly in the u.h.f. section of the course and hence will not be discussed any further here.

Oscillator "pulling" is a problem in this frequency range, and a particularly vexing problem is the change in frequency that occurs after possibly 15 to 30 minutes operation, when the set has warmed up appreciably. This can in part be circumvented by allowing a certain amount of extra band width in the i.f. amplifiers to accommodate such changes in oscillator frequency, and also by the use of temperature compensating capacitors in the oscillator circuit. A.F.C. circuits are desirable, but in general are too expensive.

The problem of thermal noise is also treated very thoroughly in the u.h.f. section of the course. Suffice it to say here that in the 100-mc region an r.f. amplifier stage has sufficient gain, in spite of the appreciable input loading of the tube, to warrant its use to build up the signal strength to a point noticeably greater than the "shot" noise of the lower gain mixer tube. At a much higher frequency, depending upon the tube used, the gain is so low that the signal-to-noise ratio is not appreciably improved, so that an r.f. stage is not justified. This is the so-called "cross-over frequency;" it is considerably above the f.m. band (88 to 108 mc). Another advantage of an r.f. stage, just as in the case of the standard broadcast receiver, is the reduction in reradiation from the local oscillator.

I.F. AMPLIFIERS.—The i.f. amplifier is also similar to that for the standard broadcast receiver, and the same design formulas hold. The more-or-less standard intermediate frequency is 10.7 mc, but Philco, for example, employs 9.1 mc. Either is noticeably higher than the 450 kc or so employed in the standard broadcast receiver; this is mainly due to the higher carrier frequencies, which demand a higher i.f. for adequate image rejection, and in part due to the greater band width required, namely, 200 kc. Normally, double-tuned i.f. transformers are employed. Owing to the high intermediate frequency employed, regeneration is more apt to occur, and much more care must be employed in shielding, lead placement and length, and by-passing of screen and plate circuits.

As an example of design, consider an i.f. stage fed by a 6BA6 tube, whose G_m = 4,400 μmhos, and whose R_p = 1.5 megohms, for a plate supply voltage of 250. This stage

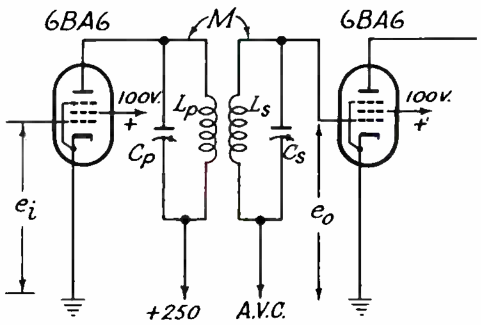


Fig. 28.—Typical i.f. stage of an f.m. receiver.

feeds the grid of a 6BA6 tube used in a second i.f. stage, so that the secondary of the i.f. transformer can be assumed to be unloaded. The circuit is shown in Fig. 28.

Assume the primary and secondary capacitors are $C_p = C_s = 33$ $\mu\mu$f. For an i.f. of 10.7 mc,

$$L_p = L_s = \frac{1}{\omega_1^2 C_p} \qquad (36)$$

or

$$L_s = L_p$$

$$= \frac{1}{(2\pi \times 10.7 \times 10^6)^2 (33 \times 10^{-12})}$$

$$= 6.69 \ \mu\text{henries}$$

(The student should refer to the previous assignment for review of the procedure that follows.)

The coupled-in resistance for the primary coil

$$R = \frac{1}{\omega_1^2 C_p^2 R_p}$$

$$= 1 \div [(2\pi \times 10.7 \times 10^6)^2 (33 \times 10^{-12})^2 \times (1.5 \times 10^6)]$$

$$= 0.1352 \text{ ohms or } 0.14 \text{ ohms}$$

This is a very small value, and indicates that the R_p loading of the driver tube is very small at such a high intermediate frequency. Suppose the primary and secondary coil Q's are each equal to 90. The secondary *circuit* Q will be the same, since the secondary is not appreciably loaded by the grid circuit of the following 6BA6 tube, but the primary circuit Q will be somewhat less than 90. The series resistance of the primary coil is

$$R_{pw} = \frac{\omega_1 L_p}{Q}$$

$$= \frac{2\pi \times 10.7 \times 10^6 \times 6.69 \times 10^{-6}}{90}$$

$$= 5 \text{ ohms}$$

The total series resistance is

$$R_{pt} = 5 + .14 = 5.14 \text{ ohms}$$

and the primary *circuit* Q is therefore

$$Q_p = \frac{\omega_i L_p}{R_{pt}}$$

$$= \frac{2\pi \times 10.7 \times 10^6 \times 6.69 \times 10^{-6}}{5.14}$$

$$= 87.6$$

Suppose a 200-kc band width is desired. Referring to Fig. 29, note that the band width between either peak and the center or carrier frequency is

$$\Delta f = \frac{200}{2\sqrt{2}} = 70.7 \text{ kc} = 0.0707 \text{ mc}$$

The fractional band width is

$$\Delta f/f = \frac{.0707}{10.7} = 0.00661$$

Let

$$\gamma_p = f_p/f_i = 1 - \Delta f/f$$

Then

$$\gamma_p = 1 - .00661 = 0.993$$

All quantities have now been evaluated that are necessary for the calculation of the required coupling necessary to give the desired band width. Since

$$k = \sqrt{\left(\frac{1}{\gamma_p^2} - 1\right)^2 + \frac{1}{2}\left(\frac{1}{Q_s^2} + \frac{1}{Q_p^2}\right)} \qquad (37)$$

we have

$$k = \sqrt{\left[\frac{1}{(.993)^2} - 1\right]^2 + \frac{1}{2}\left[\frac{1}{(90)^2} + \frac{1}{(87.6)^2}\right]}$$

$$= \sqrt{(1.014 - 1)^2 + \frac{1}{2}[\ 0001233 + .0001301]}$$

$$= .01798$$

The critical coupling is

$$k_c = \frac{1}{\sqrt{Q_p Q_s}} \qquad (38)$$

or

$$k_c = \frac{1}{\sqrt{87.6 \times 90}} = .01125$$

The ratio of the gain α_p at f_p to α_r at f_i (Fig. 29) is given by

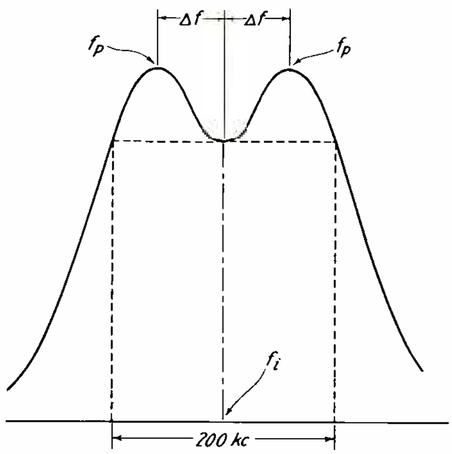


Fig. 29.—Peak half band width and total band width.

$$\frac{\alpha_p}{\alpha_r} = \frac{(k_c/k) + (k/k_c)}{\gamma_p^2\left[\sqrt{Q_p/Q_s} + \sqrt{Q_s/Q_p}\right]\left[1 - \frac{k_c^2}{4k^2}\left(\sqrt{Q_s/Q_p} - \sqrt{Q_p/Q_s}\right)^2\right]^{1/2}} \tag{39}$$

or

$$\frac{\alpha_p}{\alpha_r} = \frac{(.01125/.01798) + (.01798/.01125)}{(.993)^2\left[\sqrt{87.6/90} + \sqrt{90/87.6}\right]\left[1 - \frac{(.01125)^2}{4(.01798)^2}\left(\sqrt{90/87.6} - \sqrt{87.6/90}\right)^2\right]^{1/2}}$$

$$= 1.13$$

or the gain at the peak will be 13 per cent higher than at the center, carrier, resonant frequency. If three i.f. stages are employed, the gain variation will be

$$(\alpha_p/\alpha_r)^3 = (1.13)^3 = 1.443$$

or 44.3 per cent excess gain.

Often the band width is made appreciably less than 200 kc. For example, the overall gain may be allowed to drop 6 or even 10 db (1/2 to 1/3 in amplitude relation) at 100 kc off the center frequency. This improves the adjacent channel selectivity, and the succeeding limiter stage or stages can then eliminate the variation in amplitude as the carrier deviates its maximum amount of 75 kc from the center frequency, without too seriously affecting the frequency modulation. This can be allowed for in the above calculations by choosing Δf possibly .707 times the above value of 70.7 kc.

As a final point, the gain at the center frequency f_1 will be calculated. This is given by the formula

$$\alpha_r = \frac{G_m\sqrt{L_s/C_p}\,k}{k^2 + k_c^2} \tag{40}$$

or

$$\alpha_r = \frac{4400 \times 10^{-6}\sqrt{\frac{6.69 \times 10^{-6}}{33 \times 10^{-12}}}\,(.01798)}{(.01798)^2 + (.01125)^2}$$

$$= \frac{(4.4)(.451)(.01798)}{.000322 + .000126} = 79.6$$

This value compares favorably with those obtained in a.m. i.f. stages.

F.M. LIMITERS.—One of the basic differences between an f.m. and a.m. receiver is the use of one or more limiter stages. There are two basic reasons for the noise-free reception possibilities of an f.m. receiver: wide frequency deviation and the use of limiters. The latter, by removing the amplitude variation in the wave before detection, removes the most potent factor in the interfering wave; the remaining frequency variation in the interference is very small unless the interference is greater in amplitude than the signal, say—by a factor of two to one.

The limiter acts like a "clipper" stage in pulse techniques, and limits the excursion of the total signal to within a prescribed amplitude. Clipping may be accomplished by plate-current cutoff or by saturation effects in the plate circuit, as in the case of a pentode tube.

The latter effect is illustrated first in Fig. 30. If the load

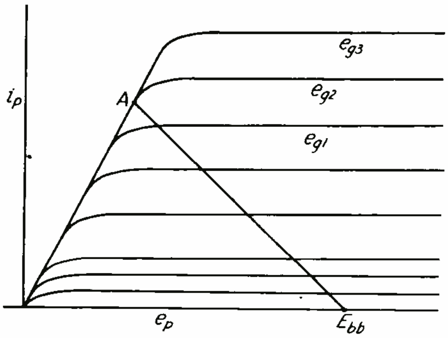


Fig. 30.—Plate family of characteristics showing limiting action owing to shape of curves.

line E_{bb} A is high enough in impedance to cut into the "knee" of the characteristics at A, then for grid voltages greater than e_{g1}, such as e_{g1}, e_{g2}, e_{g3}, etc., the plate current ceases to rise above point A, and the resultant output wave is clipped to this level.

Its appearance is as shown in Fig. 31. The dotted lines indicate

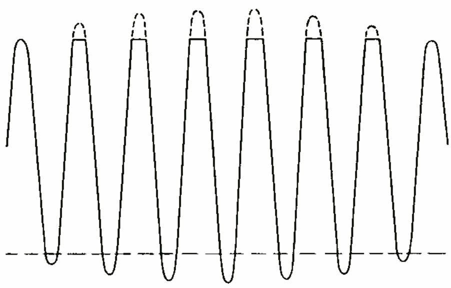


Fig. 31.—Appearance of wave after limiting.

the flattening that takes place in the tops of the waves. Clipping can also simultaneously occur in the negative direction owing to the drawing together of the plate curves near E_{bb} in Fig. 30. This is indicated in Fig. 31 by the horizontal dotted line cutting through the bottom of the peaks.

Limiting in the grid circuit by plate-current cutoff is perhaps more generally favored. This is illustrated in Fig. 32. Note the 100,000-ohm resistor and 47-$\mu\mu$f capacitor between the i.f. secondary and the cathode (ground). An alternative arrangement is shown

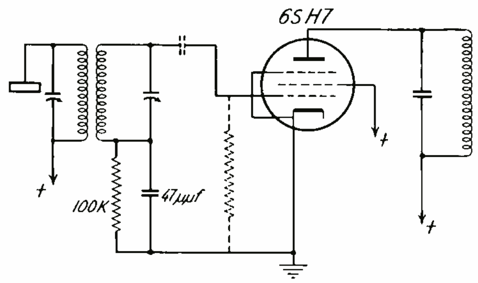


Fig. 32.—Limiter stage employing clipping in grid circuit.

by dotted lines; the action is the same in either case.

In the absence of an input voltage, the grid is simply at zero bias. When a signal is impressed via the i.f. transformer, each positive half-cycle causes grid current to flow: electrons flow from cathode to grid (the two acting as a diode) and thence in part into the 47-$\mu\mu$f capacitor, and in part through the 100,000-ohm resistor.

Since the latter is a relatively high resistance, the small capacitor readily charges up to the peak value of the input wave. On the negative half-cycle, the grid-to-cathode path is non-conducting, and the charge trapped on the 47-$\mu\mu$f capacitor can escape (the capacitor discharge) only through the 100,000-ohm resistor. In the time of one i.f. cycle, the capacitor can discharge but very little, since the time constant $47 \times 10^{-12} \times 100{,}000 = 4.7$ μsec. is much longer than the approximately 0.1 μsec. period of a 10.7 mc wave. Thus the capacitor is still nearly at its peak charge value, and on the next positive half-cycle draws just enough current to replenish the charge it lost through the 100,000-ohm resistor on the previous negative half-cycle.

Suppose, however, the wave slowly decreases in amplitude from its initial value, as when it is amplitude-modulated. For such slow variations, the capacitor is able to discharge its excess charge so as to follow the slowly decreasing amplitude. Thus, each successive peak is able to drive the grid slightly positive to replenish some of the charge lost during the preceding negative half-cycle, and the peaks all come up to a fixed amplitude.

This is illustrated in Fig. 33. The input wave would normally vary

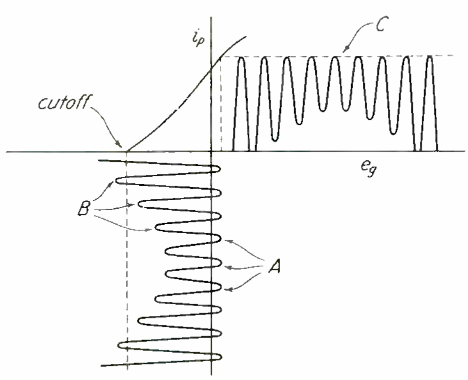


Fig. 33.—Appearance of grid input wave and plate current output wave in grid limiter.

equally in amplitude on both sides of its axis. Owing to the manner in which the positive peaks are pulled to the zero line or slightly to the right of it, as at A, all the variation in amplitude occurs on the left-hand side, B.

The plate current appears as at C. Note that some of the variation in amplitude appears on the negative half cycles. This has been purposely exhibited in the

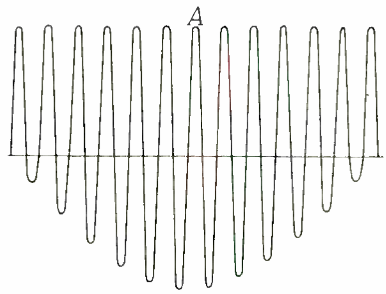


Fig. 34.—Representative appearance of the output wave of a limiter.

figure by using a rather small amplitude of input wave and/or high screen and plate voltages, which move the cutoff point to the left in Fig. 33. If—as is normally the case—the amplitude is sufficiently great, and the plate and screen voltages are fairly low, the negative half cycles will be clipped by plate-current cutoff, and the output wave will look like that shown in Fig. 34.

Note that the largest peaks, such as A, are very thin, because most of the wave lies beyond cutoff (shown in dotted lines). One bad effect of this is that the successive peaks coming through the limiter are not similar in shape.

Since these waves are distortions of assumed sinusoid coming in, they contain a fundamental component corresponding in frequency to the input wave, plus harmonics. The latter are bypassed by the tuned i.f. transformer in the output of the limiter, and only the fundamental is passed on.

If, however, the successive pulses differ in shape, then the amplitude of the fundamental component (as well as other components) will vary in magnitude somewhat; in short, the peak-to-peak value of the wave will remain constant, but the amplitude of the fundamental component that is passed on to the next stage through the tuned circuit will vary to some extent. Thus, the limiter will not entirely eliminate the amplitude modulation present in the input wave.

If a second limiter is employed, the residual amplitude modulation in the output of the first limiter is almost completely removed. Thus, the higher grade f.m. receivers will employ two limiters and thereby derive the maximum benefit from this method of modulation. The cheaper models, however, employ in general one limiter stage; considerable improvement in signal-to-noise ratio is obtained thereby, although not the maximum results.

Since a limiter essentially clips off and discards most of its incoming wave, its gain is in general very low, usually about unity. Moreover, it requires a considerable signal input (from 1 to 5 volts) to deliver the requisite 1/2 volt or thereabouts to the discriminator. Hence, appreciable i.f. gain is required ahead of it, and in practice two normal high-gain i.f. stages precede it.

Thus, an input voltage at the antenna terminals of but 5 μv may be sufficient to saturate the limiter, and the receiver has therefore considerably higher gain than an a.m. receiver. This is also necessitated by the fact that the relatively small half-wave dipole used at the higher f.m. frequencies picks up much less energy than the standard broadcast antenna (except the loop types) and hence requires more gain in the set.

In Fig. 35 is shown a method of obtaining a.v.c. from the limiter

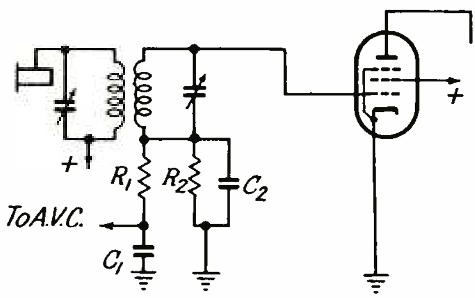


Fig. 35.—Method of obtaining a.v.c. from a limiter stage.

stage. In this figure, R_2 C_2 constitutes the time-constant circuit which produces the limiting action. Since the peaks of the incoming wave drive the grid momentarily a volt or so positive, the rest of the wave drives the grid in the negative direction. A large-amplitude input voltage thus drives the grid considerably negative over most of the cycle; and the *average* or d.c. potential of the grid is considerably negative.

A small-amplitude input voltage, on the other hand, does not drive the grid as far negative, and its average or d.c. potential in this case is less. This can be checked by reference to Fig. 33, where the average potential, which is halfway between the negative and positive peaks, weaves back and forth as the amplitude varies. This phenomenon is representative of the manner in which an oscillator establishes its grid bias by grid current, and automatically increases the bias as the plate-to-grid regenerative coupling is increased.

Hence, the *negative* potential of the top end of R_2 in Fig. 35 will increase if the incoming carrier wave increases, and decrease if it decreases. Consequently, if another R-C network, composed of R_1 and C_1, is connected as shown, only the slow variations in the carrier owing to fading or stations of different strengths will appear as a slowly changing negative potential at the junction of R_1 and C_1, and this negative potential can be employed for a.v.c. purposes. (As will be shown subsequently, a.v.c. can also be obtained from the discriminator output.)

Two final points about limiters are:

1. If the signal is very weak, the limiter is not saturated and fails to function, just like a delayed a.v.c. network in an a.m. receiver. Indeed, a limiter in a sense is also an a.v.c. circuit. The receiver must be designed to have sufficient gain so that such weak signals as fail to saturate the limiter are too weak to "capture" the receiver from the noise present.

2. After a sudden high-amplitude noise impulse, the receiver may be "paralyzed" by the biasing action of this impulse at the limiter input. This means that the grid time constant, such as $R_2 C_2$ in Fig. 35, must be sufficiently small (about 1 to 6 μsec.) to discharge sufficiently rapidly to follow the sudden decrease in amplitude after the impulse has passed, and yet not too small, so as not to follow the individual i.f. cycles. The receiver will then "punch out" the desired signal to a minimum extent when the impulses occur, and thus they may hardly be noticeable to the ear.

DEMODULATORS.—The f.m. detector or demodulator must translate frequency variations into the modulating wave once more. At the same time it should be as nearly insensitive to amplitude variations as possible, although the limiters preceding it normally take care of this consideration. One form of demodulator, known as the ratio detector, is so insensitive to amplitude modulation as to require no limiter; the other forms in general do not possess this advantage, except to a slight degree.

The usual method of detection is to impress the f.m. wave on a frequency-sensitive circuit, which converts the constant-amplitude variable-frequency input wave into a variable-amplitude variable-frequency wave. An ordinary diode detector can then be used to furnish the audio output, since it is sensitive to amplitude variations even though it is practically insensitive to frequency variations.

Elementary F.M. Detector.—One of the earliest and most elementary of f.m. detectors is shown in simple schematic form in Fig. 36. The last i.f. stage feeds a diode detector from whose output is obtained the modulation frequencies.

At first glance this might appear to be an ordinary a.m. detector, but there is an important difference: the last i.f. transformer is tuned off the carrier

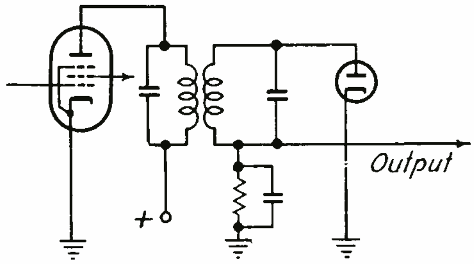


Fig. 36.—Elementary f.m. detector.

frequency, so that the signal "rides" up and down on the sloping part of the resonance curve.

This is shown in Fig. 37(A). The carrier f_c is on the sloping attenuating part of the curve. As the carrier deviates over its maximum range $2F_D$, the output of the filter varies correspondingly,

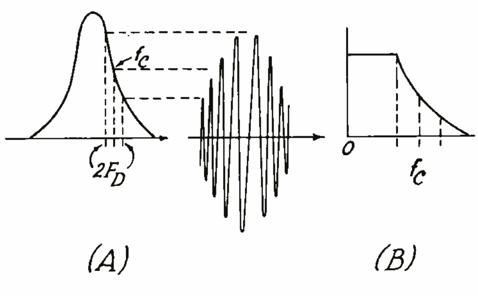


Fig. 37.—Use of resonant circuit and also cutoff portion of a band-pass filter to obtain a.m. from f.m.

and the frequency modulation is thereby converted into simultaneous amplitude modulation, as shown by the r.f. wave at the right of the resonance curve.

An ordinary diode detector is then able to demodulate the amplitude modulation and thus furnish the audio output. However, the side of a resonance curve is by no means straight, and so distortion in the amplitude-modulated wave can be expected, especially for wide deviations. A somewhat straighter slope can be obtained by the use of a low-pass filter in place of the resonant circuit, and operating the filter along its cutoff slope, as shown in (B). Neither arrangement, however, is satisfactory for high-fidelity reception.

Standard Discriminator.—The arrangement in Fig. 36 gave way to an arrangement having two tuned circuits; one tuned to a frequency somewhat higher than the carrier frequency, and the other somewhat lower. The two produced amplitude modulation 180° out of phase with one another. Two diodes, each fed from one of the above tuned circuits, had their outputs connected in push-pull arrangement; this cancelled even-harmonic distortion just like in other push-pull circuits.

The main difficulty with this circuit was the difficulty in adjusting the two tuned circuits to

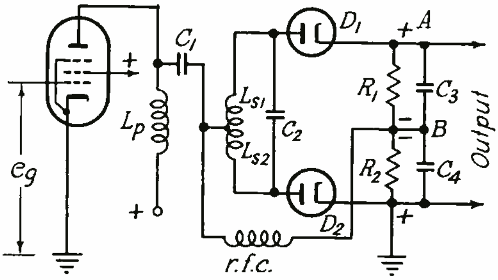


Fig. 38.—Typical discriminator circuit.

proper balance, and so the arrangement gave way to the Seeley-Foster circuit shown in Fig. 38. Here the primary L_p (which is often also tuned) is coupled magnetically to the center-tapped secondary L_{s1}/L_{s2} as well as capacitively coupled.

Thus, a current flowing through L_p induces by mutual induction two voltages, one in L_{s1} and an equal magnitude in L_{s2}. *With respect to the center tap*, however, these two voltages are in phase opposition, just like in the secondary of a push-pull input transformer.

At the same time, the coupling network C_1 in series with the radio-frequency choke (r.f.c.) impresses practically the entire voltage developed across L_p between the secondary center tap and ground (since at intermediate frequencies the reactance of C_1 and C_4 compared to r.f.c. are both negligible). Therefore the voltage impressed across D_1 (whose cathode is essentially at ground potential owing to the negligible reactance of C_3) is the primary voltage vectorially added to that induced in L_{s1}; the voltage across D_2 is the vector sum of the primary voltage and that induced in L_{s2}.

The secondary circuit L_{s1} L_{s2} C_2 is resonant at the carrier intermediate frequency. Hence at this frequency (representing zero deviation), the vector conditions are as follows:

1. Refer to Fig. 39. Let the input grid voltage e_g be the reference vector. Then, for a pentode tube, the a.c. component i_p, of the plate current flowing through L_p will be in phase with e_g, since its magnitude and phase are determined practically entirely by the high R_p of the tube.

2. The voltage e_p developed across L_p, an inductance, should *lead* i_p by 90°. However, owing to the 180° phase reversal in a tube, it *lags* i_p by 90°, and is so shown in Fig. 39.

3. The voltage e'_s induced in the secondary by i_p will lag the latter by 90°, and hence be in phase with e_p.

4. At resonance, the secondary current i_s will be in phase with e'_s.

5. It will develop across C_2, and hence the terminals of the secondary coil, a voltage which *lags* i_s by 90° (since the current *leads* the voltage by 90° in a capacitor).

6. Half of this voltage, or e_{s1}, developed across L_{s1}, will lag i_s, but the other half, or e_{s2}, developed across L_{s1}, will—as explained previously—be 180° out of phase with e_{s1}, and will

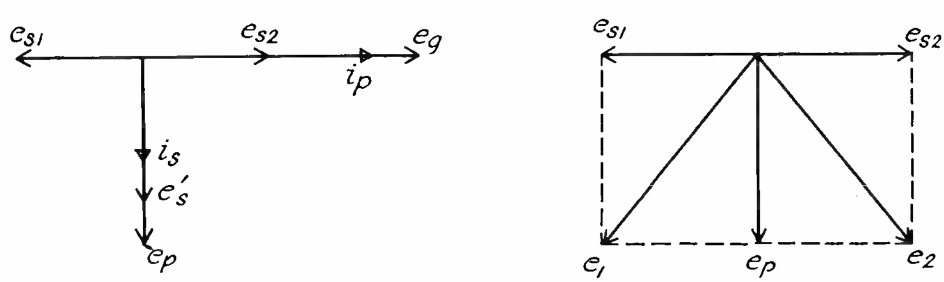


Fig. 39.—Vector phase relations at resonance in a discriminator circuit.

therefore lead i_s. This is also shown in Fig. 39.

7. The right-hand diagram of Fig. 39 then shows the two resultants e_1 and e_2 that are impressed across diodes D_1 and D_2 respectively. Thus e_1 is the resultant of e_{s1} and e_p; e_2 is the resultant of e_{s2} and e_p; from the figure it is clear that they are equal.

This means that the two diodes have equal voltages impressed across them, and hence produce equal *but oppositely flowing* currents in their respective load resistors R_1 and R_2. As a result, the voltage between terminal A and ground (the output) is *zero*. Note, however, that the voltage from the center point B to ground, is not zero, but is some negative value depending upon the magnitude of the carrier, hence the magnitude of e_g at the input of the driving tube. Therefore, an a.v.c. voltage of the correct negative polarity can be obtained between terminal B- and ground.

Thus, when the circuit is adjusted to resonate at the carrier frequency, the output is zero when the wave is unmodulated. Suppose the wave now rises in frequency from its unmodulated value. The circuit is now being operated above its resonant frequency, and develops an *inductive* reactance. Owing to the losses in the coil and the damping effect of the diodes, the circuit has a mixed resistive and inductive impedance, of the form R + jX.

As a result, i_s now lags e'_s, the induced voltage, by some angle θ. This is shown in Fig. 40. Since the voltage across capacitor C_2 always lags i_s by 90°, the voltage e_{s1} across L_{s1} lags i_s by 90°, and the voltage e_{s2} across L_{s2} leads i_s by 90°, and the two are therefore at an angle θ to e_g, the reference vector.

The two resultants e_1 and e_2 are now of unequal length, with e_2 greater than e_1. Diode D_2 therefore develops a greater d.c. voltage across R_2 than D_1 develops across R_1, with the result that ground is more positive to B than A is to B. This in turn means that A is *negative* to ground.

As the frequency of the wave increases, B becomes at first more negative to ground, but at the same time the impedance of the secondary circuit increases more and more, so that i_s decreases. This causes the two voltages e_{s1} and e_{s2} to decrease, hence, ultimately also the resultants e_1 and e_2, until their differential effect through the diodes increases less rapidly, reaches a maximum, and then begins

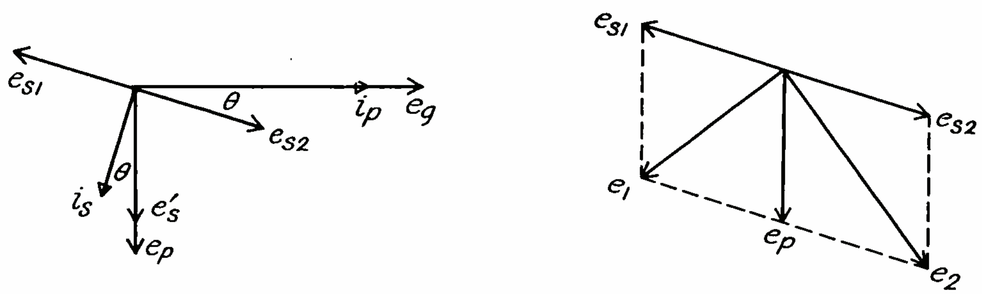


Fig. 40.—Vector phase relations above resonance in a discriminator circuit.

to decrease. This represents the region beyond which rise in output voltage is directly proportional to frequency deviation; i.e., this is a nonlinear region.

If the frequency of the wave decreases from the carrier value, the conditions are reversed; the secondary circuit becomes capacitive, or rather of the form of $R - jX$, i_s leads e'_s by some angle, and the conditions are as represented in Fig. 41. Now e_1 exceeds e_2; the potential from A to B is more positive than from ground to B, so that A is now *positive* to ground. Thus the polarity of A with respect to ground varies from negative to positive as the frequency varies above and below resonance, respectively. A maximum in output voltage is also reached for a sufficient decrease in frequency of the wave, similar to the maximum described above.

The complete discriminator characteristic is shown in Fig. 42. Note the linear portion between A and B. If this exceeds 2 × 75 kc, distortionless output can be expected for the loudest sound, *provided the secondary is tuned to* the undeviated carrier frequency f_c. If the circuit is mistuned, then distortion can occur either for deviations in a positive or negative direction, depending upon the direction of mistuning.

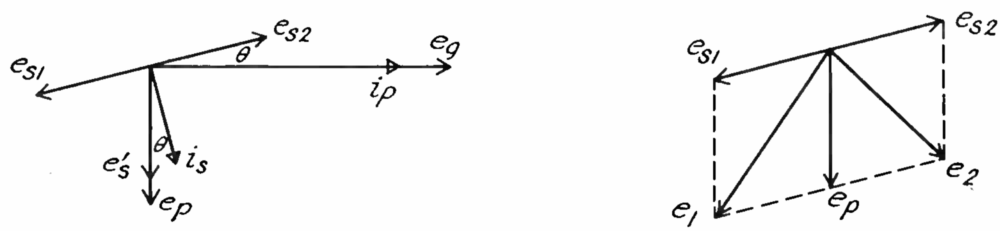


Fig. 41.—Vector phase relations below resonance in a discriminator circuit.

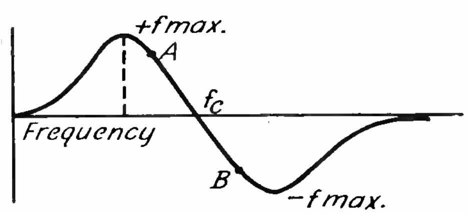


Fig. 42.—Discriminator characteristic: output versus frequency deviation.

Discriminator Alignment.—This suggests a simple method of aligning the discriminator. An a.m. signal generator is connected to the grid of the last limiter, and adjusted to deliver the intermediate frequency, say 10.7 mc, amplitude-modulated with some audio frequency—usually 400 c.p.s. It will be advisable to raise the G_m and hence gain of the limiter tube. This can be done by shunting the resistor in series with the plate load with a low resistor of about 2,000 ohms. This raises the plate voltage, and hence the gain of the stage.

Now connect an output meter across the voice coil of the loudspeaker. If the discriminator is tuned to 10.7 mc, the output will be zero, since the circuit is balanced at this frequency regardless of the amplitude of the input wave, and hence is unresponsive to amplitude modulation.*

Adjustment for zero output is obtained by tuning the secondary capacitor or inductance (if the latter has an iron core tuning slug). Three points of zero output can be obtained, corresponding to slope AB in Fig. 42, or the opposite slopes to the left or right of AB. Slope AB has the correct minimum point, and can be distinguished from the other two slopes by the fact that varying the capacitor, for example, in either direction, increases the output.

The adjustment is then set slightly off the null position so as to obtain some output. If the primary is also tuned with a capacitor, this circuit can now be adjusted for maximum output across the voice coil. The same procedure can then be followed for each preceding i.f. stage, just as in the case of an a.m. receiver (described in a previous assignment). Once the receiver is aligned, the discriminator can be readjusted for zero output once more.

Another method, the so-called visual method, employs a frequency-modulated i.f. signal generator. This has an internal reactance tube, actuated by a built-in audio oscillator, usually of 400 cycles, although an external audio oscillator can also be connected. The main dial is set to the desired carrier frequency, and an auxiliary dial can then be set to any amount of frequency deviation wanted. A sweep of ±200 kc is advisable so as to trace out the complete discriminator characteristic shown in Fig. 42.

The high side of the sweep signal generator is connected through a .05-μf capacitor to the grid of the last limiter tube, and the low side to ground. An oscilloscope, preferably one employing a 5-inch tube, is used as a visual indicator; its vertical deflection amplifier is connected across the discriminator output, and the horizontal deflection amplifier to the audio modulation, a pair of synchronizing terminals being provided on the sweep generator for this purpose.

The test setup is shown in Fig. 43. The modulating voltage

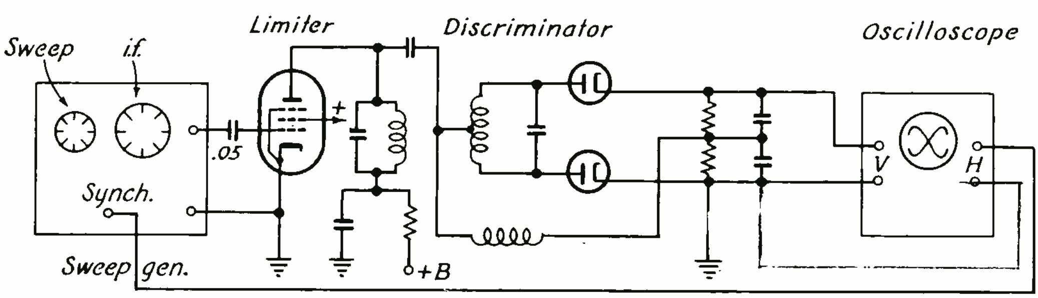


Fig. 43.—Test setup for visual adjustment of f.m. receiver.

*It is also possible to use an unmodulated signal, and connect a high-resistance d.c. voltmeter across the output terminals of the discriminator itself. A zero reading indicates proper adjustment of the discriminator. This method, although simpler, is less sensitive because no advantage can be taken of the gain of the audio system.

in the signal generator produces a frequency variation or sweep that varies in a triangular fashion with time; i.e., during one-half of the sweep cycle the frequency of the signal generator increases linearly with time from a minimum to a maximum value, and during the next half-cycle it varies uniformly from the maximum back to the minimum value.

The horizontal deflection voltage for the oscilloscope beam is sawtooth in shape, and has a relatively slow forward stroke (left to right), and a quick return stroke. The frequency of the deflection voltage is made *twice* that of the modulating voltage and hence sweep rate.

As the audio voltage rises from its negative to its positive peak, assume the sweep generator varies from its minimum to its maximum frequency. The discriminator output will vary from maximum negative to maximum positive voltage, and the cathode ray beam will therefore move upward on the screen.

At the same time the beam moves from the left to the right and then back, so that a curve similar to AB in Fig. 44 will be seen on the cathode ray screen. When the audio wave varies from a positive to a negative maximum, the frequency of the sweep generator will drop from its highest to its lowest value, and the beam will simultaneously make another sweep from left to right, and then

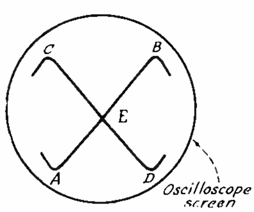


Fig. 44.—Appearance of signal traces on the cathode ray screen.

back. The result is the trace CD in Fig. 44 (the return strokes are faint and not shown).

The correct adjustment of the discriminator is obtained when the crossing E of the two traces occurs exactly midway between the peaks. This method of adjustment is more sensitive than the one described previously in that an absolute null adjustment does not always occur, whereas the correct crossover E in Fig. 44 is readily obtained.

Alignment of the preceding i.f. stages is also readily obtained. Thus, the shape of the i.f. selectivity curve can be seen visually by this means, and the necessary adjustments made to the tuning capacitors or slugs in the coils, as the case may be. The vertical amplifier of the oscilloscope should be connected across the grid resistor of the limiter stage. A common ground point should be used for all components to prevent oscillation.

As in the case of the discriminator test, two curves are seen on the screen: one for the forward and

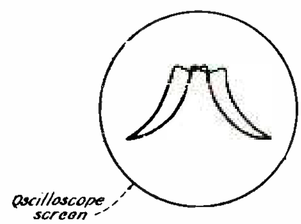


Fig. 45.—Appearance of i.f. resonance curves on oscilloscope screen.

one for the backward sweep. This is indicated in Fig. 15. By adjusting the primary and secondary circuits, the two curves can be made to coincide; this represents the correct adjustment for the stage. This method can, of course, also be used to align an a.m. i.f. amplifier, and a high-frequency sweep generator can be employed in similar manner to adjust the r.f. circuits.

THE RATIO DETECTOR.—The ordinary discriminator just described is insensitive to amplitude variations only at the carrier or resonant frequency. When the frequency shifts during modulation, an unbalance occurs in the discriminator and produces a differential output between the two diodes; the magnitude of this output depends upon the amount of frequency shift, *and also upon the amplitude of the i.f. wave.*

Hence, if the wave undergoes amplitude as well as frequency modulation, the output from the discriminator will have the effects of both rather than of the frequency variation alone, and thus audio distortion will result. It is therefore necessary to use a limiter to prevent such distortion as well as to reduce the effects of noise interference.

An ingeniously simple modification of the ordinary discriminator circuit, developed by S. W. Seeley of R.C.A., and known as the *ratio detector*,* makes it insensitive to amplitude variations, and hence eliminates the need for limiter stages. The result is an important economy in the manufacture of the receiver.

The theory upon which the operation of the ratio detector is based is quite simple. Let e_1 and e_2 be the two unbalanced voltages developed by the phasing network shown in Fig. 46. These are the voltages developed between the ends of the tuned secondary and ground, and are impressed on the two diode circuits. The latter

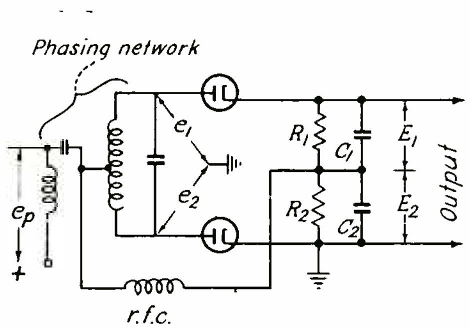


Fig. 46.—Discriminator circuit showing a.c. and d.c. voltages.

produce, as a result, two corresponding d.c. voltages E_1 and E_2 across the diode resistors R_1 and R_2. Neglecting voltage drops in the phasing network and diodes, E_1 and E_2 are approximately equal to the peak values of e_1 and e_2, owing to capacitors C_1 and C_2 shunting the diode load resistors. Hence, the ratio of E_1 to E_2 is equal to that of e_1 to e_2.

The phasing network is composed of linear circuit elements. This means that the currents flowing in it, and the voltage drops produced in it, are directly proportional to the impressed voltage e_p.

Therefore, as e_p deviates from the carrier frequency, and e_1, for example, exceeds e_2, then if the amplitude of e_p is doubled, both e_1 and e_2 will double in magnitude, but their ratio e_1/e_2 will remain unchanged. This is merely a statement of the behavior of any linear network. It means, for example, in Fig. 41, that as

*See, for example, "The Ratio Detector," Seeley and Avins, *RCA Review*, June 1947.

e_p changes in length, the vector diagram changes in size, but the relative magnitudes of the vectors, i.e., their ratios, remain unchanged.

The corresponding d.c. output voltages E_1 and E_2 will also remain in the same constant ratio as e_1 and e_2 regardless of variations in the amplitude of the input voltage e_p, so long as the rectification efficiency of the diodes does not materially decrease. (This will be discussed farther on.) Thus, for a given frequency deviation of the input voltage e_p, a certain unbalance is produced in the two a.c. voltages e_1 and e_2 and hence in their rectified outputs E_1 and E_2, such that e_1/e_2 and E_1/E_2 are independent of the magnitude of e_p.

Now suppose that the sum $E_1 + E_2$ can be kept constant at some desired value, regardless of the variation in the amplitude of e_p. Then it will be shown below that E_1 and E_2 individually will be constant, and independent of variations in e_p, so that the audio output, which is $E_1 - E_2$, will not be affected by variations in e_p. The mathematical relationships are quite simple.

Let $e_1/e_2 = E_1/E_2 = k_1$, a constant. Next, assume that by suitable means, $E_1 + E_2$ is maintained constant at some value k_2. Thus

$$E_1 + E_2 = k_2 \qquad (41)$$

$$E_1/E_2 = k_1 \qquad (42)$$

or

$$E_1 = k_1 E_2 \qquad (43)$$

Substitute Eq. (43) in Eq. (41) and obtain

$$k_1 E_2 + E_2 = k_2$$

or

$$E_2(1 + k_1) = k_2$$

or

$$E_2 = k_2/(1 + k_1) \qquad (44)$$

and from Eq. (43)

$$E_1 = k_1 k_2/(1 + k_1) \qquad (45)$$

Eqs. (44) and (45) state that both E_2 and E_1 are equal to constants, since $k_2/(1 + k_1)$ and $k_1 k_2/(1 + k_1)$ are constants if k_1 and k_2 are constants. Therefore their difference, $E_1 - E_2$, the audio output, will be a constant, too, independent of the *magnitudes* of the a.c. input voltage e_1 and e_2, but dependent only upon their ratio.

It therefore appears that (1) since the ratio e_1/e_2 is inherently independent of the magnitude of the input voltage e_p, but dependent only on its frequency deviation, and (2) since the ratio E_1/E_2 is essentially equal to e_1/e_2; then it is only necessary to hold the sum $E_1 + E_2$ constant in order that the difference or audio voltage $E_1 - E_2$ be independent of the magnitude of e_p and therefore also of e_1 and e_2, and dependent only on the frequency shift of e_p.

The method of holding $E_1 + E_2$ constant is to connect a d.c. (battery or its equivalent) across the two diode load resistors. Then, if e_p and/or e_1 and e_2 increase, excess current will flow through the battery from the phasing network, and the excess voltage will be used up as impedance drop in the network, so that $E_1 + E_2$ will remain constant and equal to the bridging d.c. battery potential.

Other changes must be made in the circuit, because if the voltage

$E_1 + E_2$ across the two diode load resistors is maintained constant, no audio voltage will be produced across them. The necessary modifications are shown in Fig. 47.

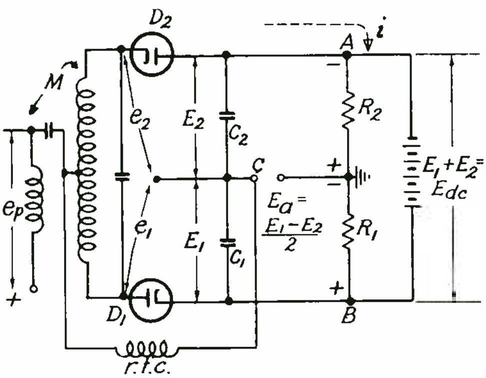


Fig. 47.—Basic circuit of ratio detector.

Note the three most important changes:

1. The diodes are connected in opposite polarity instead of in the same polarity.

2. The stabilizing d.c. battery across the diode load resistors R_1 and R_2. This battery keeps the sum $E_1 + E_2$ equal to its generated voltage E_{dc}.

3. The audio voltage

$$E_a = \frac{E_1 - E_2}{2}$$

is taken off between tap C of the two diode load capacitors C_1 and C_2, and the grounded tap of the two load resistors R_1 and R_2.

At any instantaneous shift in frequency of e_p from its carrier value, a.c. voltages e_2 and e_1 become unequal and their ratio changes from unity (at balance) to some other value. The output voltages E_1 and E_2 have a similar ratio to one another, but their sum remains equal to E_{dc}. Suppose e_2 becomes greater and e_1 smaller. Then E_1 and E_2 become unequal in the same ratio.

Owing to the polarity of connection of the two diodes, electron current can flow through the tubes and R_2 and R_1 in series in a clockwise direction, so that after the two capacitors have charged up, d.c. flows through the above-mentioned circuit path, and since $R_1 = R_2$, produces equal voltage drops in the two resistors.

This means that ground is halfway in potential between terminals A and B, or the potential of ground with respect to terminal A is $(E_1 + E_2)/2$. Terminal C is positive to terminal A by E_2 volts. Therefore, the potential of C with respect to ground is the difference between their potentials with respect to A, or

$$E_a = \frac{E_1 + E_2}{2} - E_2 = \frac{E_1 - E_2}{2} \qquad (46)$$

This indicates that the audio output voltage of a ratio detector is inherently half of that of a discriminator circuit, which is $E_1 - E_2$.

The next thing to investigate is what happens if $e_2 + e_1$, the a.c. voltage acting around the circuit, changes. There are two factors that can make $e_2 + e_1$ change:

1. The input voltage e_p changes in magnitude, or

2. The sum $e_2 + e_1$ changes as e_p shifts in frequency but remains constant in amplitude.

Factor (1) is obvious: if e_p changes in magnitude, both e_1 and e_2 will change correspondingly in magnitude, and hence their sum will change. Factor (2) is not obvious. However, consider Fig. 48, which is a repetition of the right-hand vector diagrams of Figs. 39, 40, and 41. The d.c. output voltages E_1 and E_2 are proportional solely to the magnitudes of e_1 and e_2, regardless of their phase relationships. As e_p deviates in frequency, the arithmetic sum of $e_1 + e_2$ will vary as shown in Fig. 49 even though e_p remains constant in amplitude.

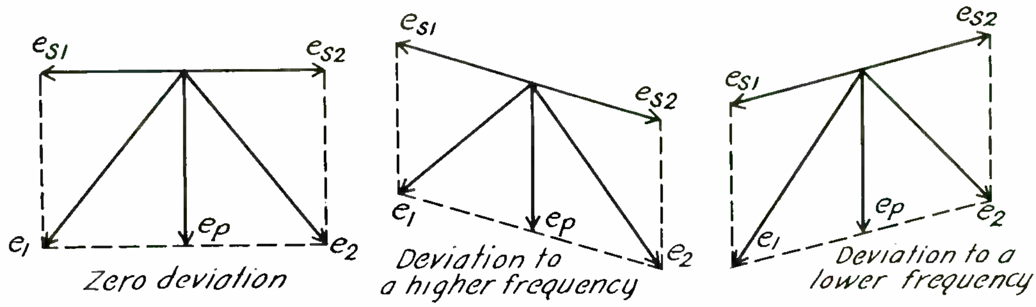


Fig. 48.—Vector diagrams showing phasing voltages for zero deviation, deviation to a higher frequency, and deviation to a lower frequency.

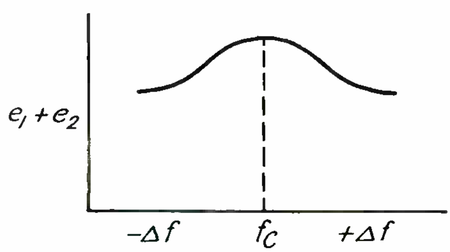


Fig. 49.—Variation in $(e_1 + e_2)$ with frequency deviation $\pm \Delta f$ from the carrier frequency f_c.

It will be observed that the sum is a maximum at zero deviation, and decreases either for positive or negative deviation; this is a consequence of the vector relations in Fig. 48.

The variation in $(e_1 + e_2)$ during the modulation cycle tends to produce a similar variation in the output voltage $(E_1 + E_2)$. Since, however, this voltage is stabilized by the bridging battery, it cannot vary. As a result, the output characteristic is straighter and extends over a greater frequency range, as is indicated in Fig. 50.

The characteristic shown in solid line in Fig. 50 remains unchanged if e_p varies during the modulation cycle, or increases to a higher value and remains at that value.

On the other hand, if e_p decreases and remains indefinitely

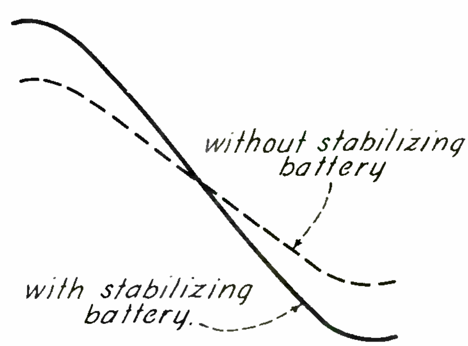


Fig. 50.—Effect of using a stabilizing battery.

at a lower value, the lower a.c. voltages e_1 and e_2 are incapable of forcing current through the diodes against the higher biasing potential of the bridging battery, and so the action stops. In order to permit the detector to operate once more, the voltage of the bridging battery must be reduced to permit the diodes to conduct and produce the variation in $E_1 - E_2$. The resulting characteristic now has a lower slope than that shown in Fig. 50. However, for *rapid variation*, such as occur during a modulation cycle, the circuit is operative and produces a straighter characteristic.

It is apparent that a battery is not a practical means of producing the ratio detector action.

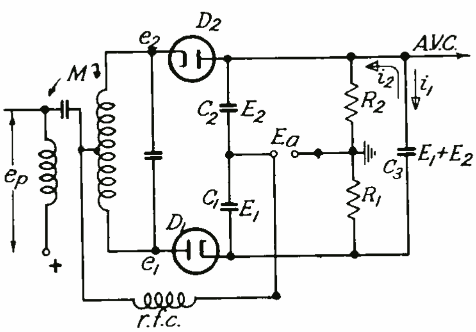


Fig. 51.—Use of a stabilizing capacitor C_3 instead of a battery.

Instead, a large capacitor (about 5 μf) is employed, as shown in Fig. 51 by C_3. The time constant $C_3(R_1 + R_2)$ should be at least 0.1 sec., which is large compared to the lowest audio modulating frequency. It then acts like a battery, once it has charged up to the *average* value of $E_1 + E_2$ over the modulating cycle.

If, for example, $e_1 + e_2$ increases, then $E_1 + E_2$ tries to increase over the potential of C_3. Instead, a charging current i_1 flows down through C_3 (which acts as a short-circuit for such a momentary current) and this excess current over that normally flowing through $R_2 + R_1$ produces an additional voltage drop in the tuned circuits and the diodes, thus absorbing the excess voltage of e_1 and e_2, and permitting $E_1 + E_2$ to remain unchanged.

If, on the other hand, e_1 and e_2 decrease, a discharge current i_2 flows out of C_3 into $R_1 + R_2$, thus holding up $E_1 + E_2$ to its previous value. So long as the change in e_1 and e_2 is momentary, C_3 can accomplish such charging or discharging without changing its potential appreciably; i.e., without $E_1 + E_2$ changing appreciably. This in turn means that C_3 acts like an ideal d.c. source that has negligible internal resistance.

The action of C_3 therefore produces the following rather curious and interesting result. If e_1 and e_2 increase, an excess charging current flows through C_3, as was just described. To the phasing network, looked upon as a generator, the load represented by the diodes and their associated circuits, appears to have lowered in resistance. This means that the damping on the a.c. circuit has increased, and the Q of this circuit has decreased. This in turn reacts on the coupled a.c. circuits involved to lower the voltages impressed upon the diode circuits.

On the other hand, if e_1 and e_2 decrease, the discharge current from C_3 takes the place, in part, of the current coming through the diodes from the a.c. network. To the latter, then, the diode circuit appears as a higher resistance, the damping is less, and the a.c. circuit Q's involved increase. This in turn enables the coupled a.c. circuits to impress higher voltages on the diodes.

The above effects are equivalent to the action of a generator having a high internal impedance and hence a high regulation, when feeding a constant load impedance. In particular, if e_p tends to decrease, the action of C_3 in producing less damping on the coupled a.c. circuits permits the a.c. voltages e_1 and e_2 to decrease to a lesser extent than they otherwise would.

This, in turn, enables e_1 and e_2 to cause current to flow to at least some extent through the diodes, until the voltages decrease to a relatively low value, whereupon the detection efficiency of the diodes becomes very low. Thus, for *inward* amplitude modulation, the diodes are not biased off, and the detector is therefore operative up to an inward modulation factor of 50 per cent or thereabouts. Since inward modulation is the critical variation in the amplitude of e_p or e_1 or e_2, successful operation of the ratio detector is possible over a wider range without the use of a limiter if the voltage regulating properties of coupled tuned circuits are properly employed. This will be discussed at greater length farther on.

If the input voltage e_p changes permanently to a different value, such as when tuning in to another station, $E_1 + E_2$ cannot be held constant indefinitely by C_3 at its former value. For example, if e_p increases, C_3 will slowly charge to the necessary higher value, and $E_1 + E_2$ will increase accordingly to a higher value. This action is in contrast to that of the fixed potential of a battery.

If e_p decreases, C_3 will decrease slowly in potential until $E_1 + E_2$ is sufficiently reduced to allow the diodes to conduct once more. Thus the ratio detector can automatically adjust itself to any given carrier level: there is no permanent "threshold" effect, as in delayed a.v.c. circuits. Indeed, a.v.c. voltage can be readily obtained from the negative end of C_3, as is indicated in Fig. 51.

Design Considerations.—The ratio detector can be represented by the basic circuit shown in Fig. 52(A). The various resistances have the following significance:

R_p represents the shunt resistance equivalent to the actual

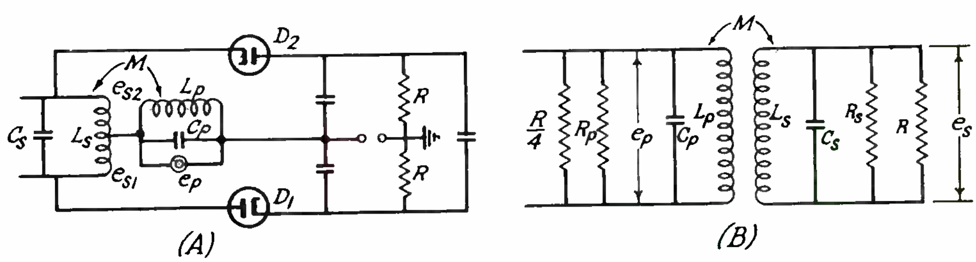


Fig. 52.—Elementary basic circuit of the ratio detector.

series resistance r_p of the primary coil L_p.

R_s represents the shunt resistance equivalent to the actual series resistance r_s of the secondary coil L_s.

R/4 represents the load which the two diode load resistors R present to the primary L_p.

R represents the load which the same two resistors present to the secondary L_s.

It has been shown in the previous assignment that the series resistance of a coil r_p may be represented by an equivalent shunt resistance R_p as follows:

$$R_p = \frac{(\omega L_p)^2}{r_p} \qquad (47)$$

Similarly,

$$R_s = \frac{(\omega L_s)^2}{r_s} \qquad (48)$$

With reference to the diode load resistors, each of value R in Fig. 52(A), it was shown that these appear as R/2 to the a.c. source. More accurately, they appear as $R/2\eta$, where η is the detection efficiency. However, η is generally so close to unity that the simpler expression R/2 can be used.

The primary coil L_p acts individually on each diode and its load resistor, owing to it being situated in the center leg of the circuit. Hence, each resistor R appears as R/2 to it, and the two resistors therefore appear in parallel to it (like two soldering iron loads on an a.c. line); hence, their joint apparent impedance is R/4.

To the secondary coil L_s, the two diodes and load resistors appear in series, so that the apparent load resistance is 2R/2 = R. The same relations hold for the ordinary discriminator circuit, too, in spite of the difference in polarity of the diodes.

In order to obtain the benefits of high regulation mentioned previously, the loaded secondary Q_s should be 1/4 of the unloaded Q_s. The primary unloaded Q should also be high compared to the Q when it is damped by the diode load. As a representative example,

Q_s (unloaded) = 90 (range from 75 to 150)

Q_s (loaded) ≐ 22.5

Q_p (unloaded) = 70

Q_p (loaded) = 40

Consider the secondary circuit. If the unloaded Q is 90, the equivalent shunt resistance R_s is also given by

$$\omega C_s R_s = Q_s \qquad (49)$$

A representative value for C_s is 50 $\mu\mu f$ (the optimum range for 10.7 mc is 25 to 75 $\mu\mu f$). Then

$$L_s = \frac{1}{\omega_1^2 C_s}$$

$$= \frac{1}{(2\pi \times 10.7 \times 10^6)^2 \times 50 \times 10^{-12}}$$

$$= 4.42\ \mu h$$

From Eq. (49), there is obtained

$$R_s = Q_s/\omega C_s$$

$$= 90/(2\pi \times 10.7 \times 10^6 \times 50 \times 10^{-12})$$

$$= 26{,}700 \text{ ohms}$$

The diode load, equivalent to R, shunts R_s, and reduces Q_s to 90/4 = 22.5. From Eq. (49) it can be seen that this is because R and R_s in parallel are equal to one-quarter of R_s itself. Thus

$$\frac{R\,R_s}{R_s + R} = \frac{26700}{4} = 6,675 \text{ ohms}$$

Solving for R, there is obtained

$$R(26,700) = 6,675\,R + (6,675)\,(26,700)$$

or

$$R = 8,900 \text{ ohms}$$

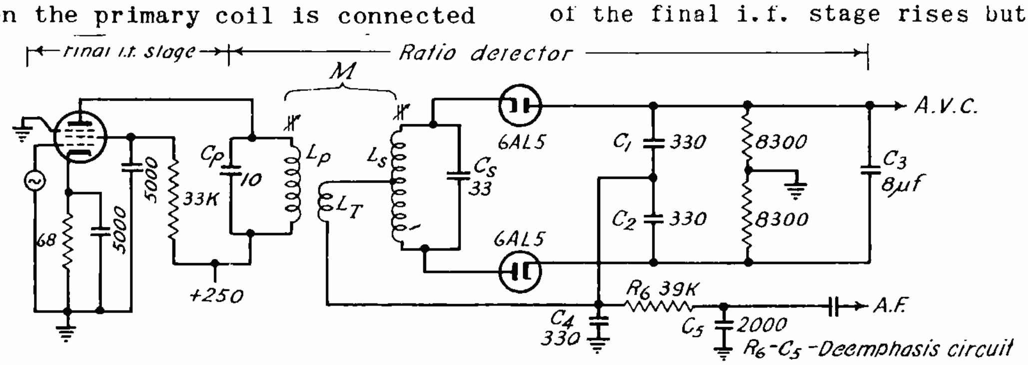
This is the value each diode load resistor should have.

Consider now the primary coil. If it is directly connected into the circuit, then the damping of the diode load alone will be 8,900 ÷ 4 = 2,225 ohms. This is such a low value, that a large value for C_p would be required [see Eq. (49)], in order to obtain any appreciable Q. The primary inductance L_p would therefore be very low, and the gain of the stage driving the detector would be low.

For this reason, it is desirable to employ some form of impedance matching. Either a tap on the primary coil is connected to the center tap of the secondary coil, or a third or tertiary winding may be employed (considering the secondary itself as two windings). The tertiary winding has fewer turns than the other coils, and a load resistance of 2,225 ohms across it will appear as a much higher resistance across the primary itself. This permits the latter to have a higher L/C ratio as well as a higher Q, and thus permits more gain in the driving stage.

A circuit employing a tertiary coil and representative values for the circuit elements is illustrated in Fig. 53. (A variation of this is shown in the diagram of the RCA Model 612V1 receiver in the next section.) The tertiary winding L_T has 8,300/4 = 2,075 ohms across it. It in turn reflects a much higher impedance to L_p, the primary winding, so that the latter has a high Q and gain. Thus, Q_p drops from 70 (unloaded) to 40 (loaded), instead of to a much lower value.

On the other hand, on inward modulation, when the action of C_3 is to make the diode load look like a much higher resistance, Q_p rises from 40 toward 70. The increase is but moderate, so that the gain of the final i.f. stage rises but



(Adapted from RCA Review.)

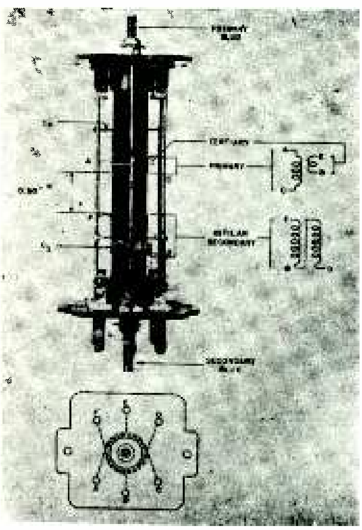
Fig. 53.—Modified ratio-detector circuit to afford impedance matching.

from 100 to 130. This is important; if the gain rose to too high a value on inward modulation, instability and oscillation could easily result.

The construction and winding data of the ratio-detector transformer is of interest, and one model will be given here. As shown in Fig. 54, the tertiary winding of 4 $^1/_2$ turns is wound at the B+ ("cold") end of the 24-turn primary, and is closely coupled to it, so that its induced voltage unbalance in the circuit performance.

The ratio of half-secondary voltage to tertiary voltage is 0.65 : 1, but in general, a ratio more nearly equal to unity is desirable. The ratio depends upon the coupling between the secondaries and the primary. An optimum value appears to be 0.5 of critical coupling k_c. It will be recalled that

$$k_c = \frac{1}{\sqrt{Q_p\ Q_s}} \qquad (50)$$



FORM: 13/32″ O.D. 3/8″ I.D. Bakelite XXX

SHIELD CAN: 1 3/8″ Aluminum, Square

TUNING SLUGS: Stackpole G-2, SK-124

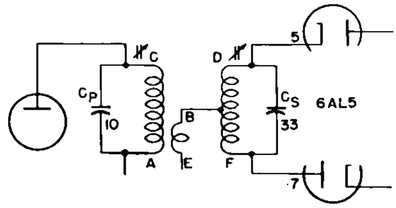
PRIMARY: 24 Turns #36E close wound

TERTIARY: 4½ Turns #36E close wound on layer paper electrical tape (0.004) over B + end of primary

SECONDARY: Bifilar, 18 Turns (total) #28. Form double grooved 20 TPI. Length of winding = 0.45″

All windings counter clockwise starting at far end

* *Measured adjacent to side rod E*



(Courtesy RCA Review, by Seeley and Avins.)

Fig. 54.—Details of construction of ratio-detector transformer.

is practically in phase with the primary voltage.

The secondary is wound in two coils of 9 turns each, so arranged that the turns of one half lie between the turns of the other half secondary. In this way, a single tuning slug inserted into the bottom end of the coil form affects equally the inductance of each half secondary, and thereby prevents an

Since Q_p and Q_s are variable if there is amplitude modulation present (as explained previously) k_c is not a fixed quantity; hence the actual coupling should be adjusted after the other factors are fixed.

Although a 6H6 double-diode will function in a ratio detector, a newer double-diode of higher perveance (lower resistance at any applied voltage) designated as the 6AL5 is preferable. The higher the perveance, the less the residual unbalances in the circuit, and the greater the rejection of amplitude modulation in the output.

Note in Fig. 53 the deemphasis circuit consisting of a 39,000-ohm resistor and a 2,000-$\mu\mu$f capacitor, which gives rise to a time constant of 78 μsec. As explained previously, the high frequencies are attenuated or deemphasized more than the low frequencies, and thus the emphasis occurring in the transmitter is compensated for. Observe also the a.v.c. take-off, the a.v.c. voltage *to ground* is one-half of $E_1 + E_2$.

The ratio detector can be readily adjusted in a manner similar to a discriminator circuit. The signal generator is set at the center or carrier frequency, a high-resistance vacuum-tube voltmeter is connected across the audio output, and the primary tuning first adjusted for maximum gain and output. Then the secondary tuning is adjusted for zero output.

Locked-Oscillator Detector.—A radically different form of f.m. detector employs an oscillator that can be synchronized or locked to the incoming i.f. wave. The frequency of this oscillator then becomes identical with that of the f.m. wave, and follows its deviations during modulation. The advantages of this method are that:

1. The synchronizing range of the oscillator can be restricted to that of the given channel, so that it is unresponsive to adjacent channel interference and noise components beyond the range, even if these are quite strong.

2. A small synchronizing i.f. signal can control the relatively large output of the oscillator. The latter then acts as an unconventional amplifier in that its output is many times the synchronizing input.

3. The output of the oscillator can be made of fixed amplitude regardless of variation in amplitude of the incoming i.f. signal. Thus the oscillator acts as a limiter, but with this important difference: the stage exhibits a gain instead of unity amplification, at most.

It is to be noted, however, that the oscillator may require a discriminator or detector of some sort following it in order to convert the frequency deviation of its output into corresponding amplitude variations. Thus, in some designs, the locked-oscillator may in a sense be regarded more as an i.f. amplifier and limiter than as an f.m. detector.

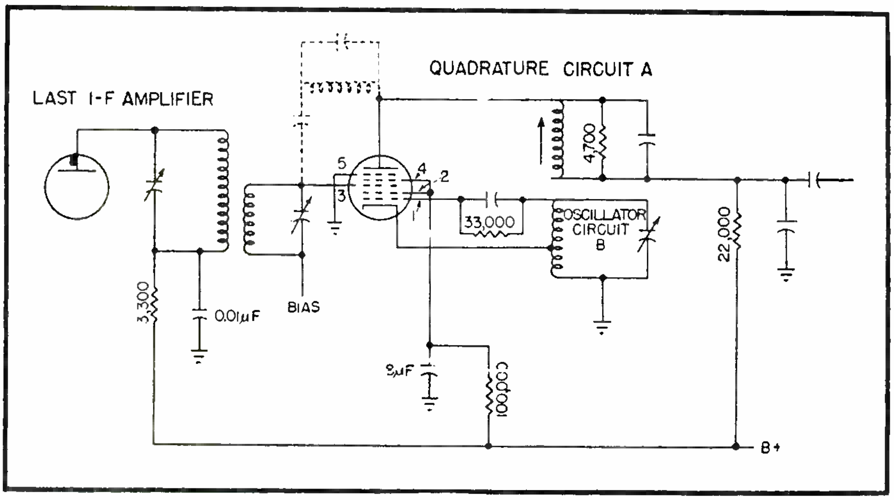
One form has been described by G. L. Beers.* Another type, employed by Philco, does not require a discriminator circuit in that the plate current varies with the frequency deviation so that it contains the modulation intelligence.

The circuit is illustrated in Fig. 55.** A special heptode tube, having a high G_m and sharp cutoff

*Patent #2,356,201—"Frequency Modulation Signal Receiving System," G.L. Beers, August 22, 1944.

**See also, "Single-stage F-M Detector," W.E. Bradley, *Electronics*, October 1946.

is employed. Grid #3 is shielded by a.c. grounded grids #2 and #4 from the other tube elements, can be exercised in the circuit layout. (The series capacitor is large and acts merely as a blocking



(Courtesy *Electronics*, by William E. Bradley.)

Fig. 55.—**Basic circuit of the f.m. detector.**

and acts as the synchronizing grid fed from the i.f. amplifier.

The oscillator is of the grounded-plate type; grids #2 and #4 acting as the plate, and grid #1 as the oscillator grid. Audio output is obtained from the plate circuit.

As will be observed from the figure, there is also a tuned circuit connected to the heptode plate, and tuned—as is also the oscillator tank—approximately to the intermediate frequency. Note the 4,700-ohm resistor in the plate tuned circuit; such heavy damping makes this circuit have a band width approximately six times the deviation band.

The dotted-line components act as a parallel resonant circuit of high impedance and thereby prevent stray capacity coupling between grid #3 and the plate. This is not required if sufficient care capacitor.)

The oscillating portion of the circuit is so designed (Class C) that the plate current flows in very short pulses (compared to the oscillating period). The synchronizing grid #3 determines the fraction of the space current, originating at the cathode, that flows through grids #2 and #4 and arrives at the plate; i.e., it determines the *amplitude* of the current pulses reaching the plate.

It does not, however, affect the *phase* of the *fundamental* component of the current pulses, but merely the *amplitude* of this component. Hence, depending upon the combination of bias on grid #3, and the instantaneous value of the incoming i.f. signal at the moment when the oscillating portion of the heptode sends a current pulse to the plate, the amplitude of the pulse and of the fundamental component take on certain values.

The fundamental component is fed back by the plate tuned circuit into the oscillator tank. The plate tuned circuit is adjusted to introduce a reactive feedback component into the oscillator tank. If the current through the former is large, the reactive feedback is large; if it is small, the feedback is small. Since, however, the phase of the current is unchanging, the feedback can remain purely reactive for all amplitudes of the feedback plate current.

As a result, the feedback reactance changes the tuning of the oscillator tank, thereby varying its frequency, without varying the amplitude of the oscillator voltage. The oscillator frequency changes until an equilibrium condition is reached whereby the oscillator frequency is equal to that of the incoming signal. For example, if the incoming signal changes its frequency by 25 kc, it begins to advance in phase with respect to the oscillator pulses at a rate of 25 kc per second. Thus, the incoming signal first increases the magnitude of the plate pulses and then decreases them, at a rate of 25,000 variations per second.

The reactions in the oscillator circuit is first to make it increase in frequency, then decrease, at a 25-kc rate. The former effect is greater than the latter, so that the oscillator frequency increases until it equals the incoming signal frequency. Thereafter it oscillates at this frequency, but maintains a certain fixed phase angle with respect to the latter so that the plate current pulses then have the correct amplitude to provide the requisite amount of reactive feedback to maintain this higher frequency. Of course, in normal operation the signal frequency is continually changing; the oscillator can react sufficiently fast to stay locked to the signal frequency.

However, as just indicated above, for any instantaneous value of signal frequency, the oscillator must so adjust itself in phase to this frequency as to permit the correct pulse amplitude to flow to the plate. Therefore, as the signal frequency varies, the pulse amplitude must vary accordingly, as is indicated in Fig. 56. The average value or so-called d.c. component of the pulse will vary in proportion to the peak amplitude, and will therefore vary in proportion to the frequency deviation.

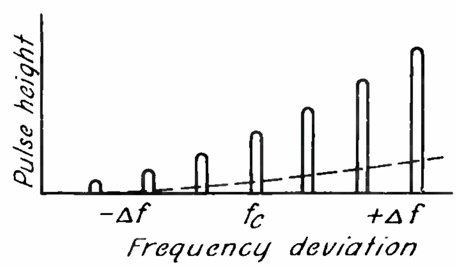


Fig. 56.—**Relation between pulse amplitude and frequency deviation.**

Such variation represents the audio signal or modulation; hence the plate current contains the modulation frequencies and no additional discriminator is required. The audio component may then be impressed upon the input grid of the audio amplifier for subsequent amplification.

If the amplitude of the incoming signal changes, the phase angle of the oscillator relative to that of the signal changes to a different value so as to pass the same amplitude of pulse to the plate and therefore maintain the same oscillator frequency as before, namely, that of the incoming signal. In short, the a.f. output is not affected appreciably by any amplitude modulation in the incoming signal. Some effect, however, may occur owing to stray capacity coupling between grid #3 and the plate; the dotted-line circuit of Fig. 55 tends to minimize this effect.

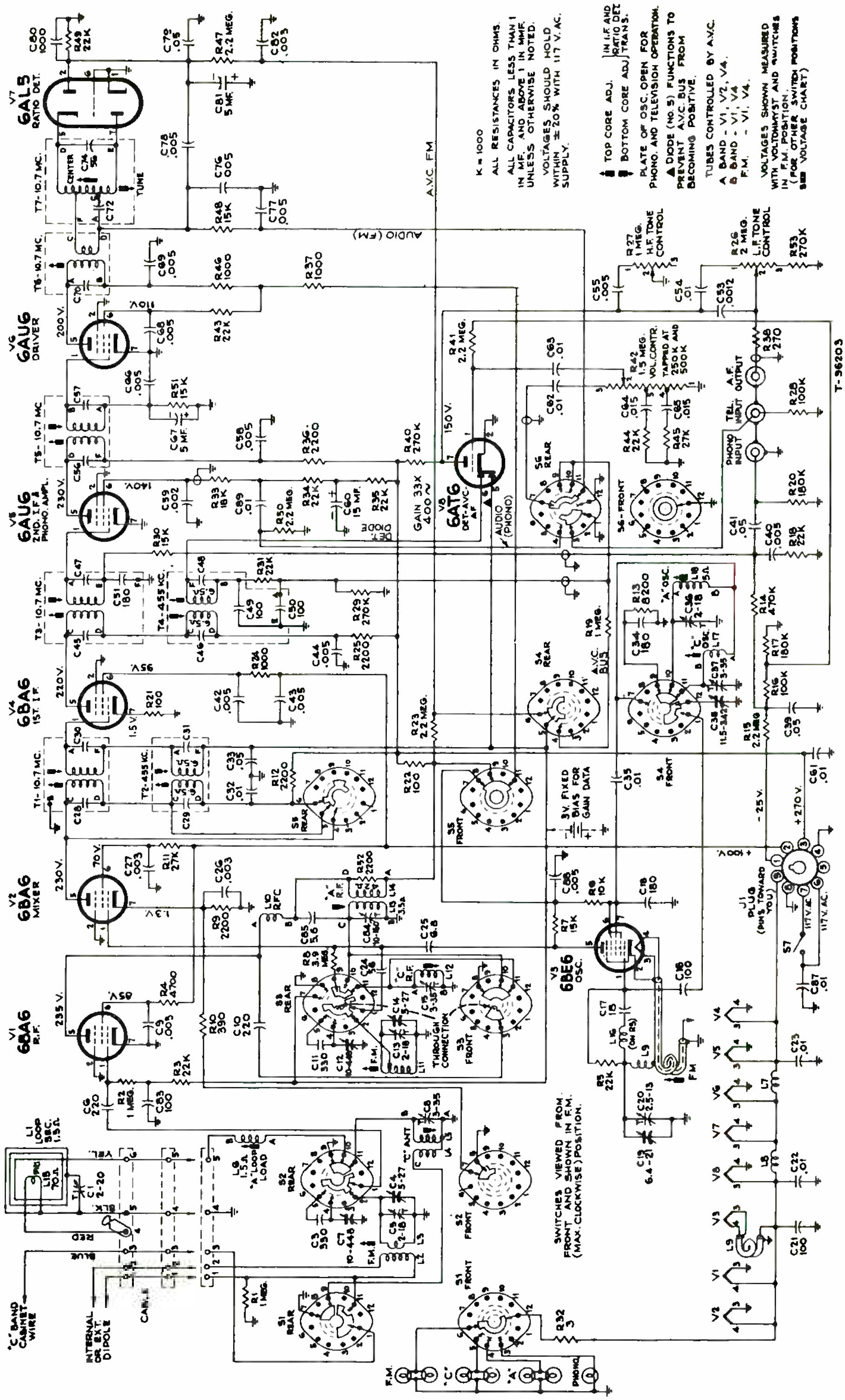
TYPICAL RECEIVER CIRCUIT.—In Fig. 57 is shown a typical f.m. receiver, the RCA Model 612V1. This receiver covers the standard broadcast band, the short wave "C" band (9.2 - 16.0 mc) and the f.m. band. A dipole antenna must be employed for the f.m. band; it feeds through the tuned-secondary transformer to the 6BA6 r.f. stage. The plate is shunt fed from +270 volts, and is capacitively coupled to the tuned circuit L_{11}, C_{13} and C_{14}. The grid of the 6BA6 mixer tube is also coupled to this circuit, thereby obtaining its signal voltage.

A 6BE6 tube is used as the local oscillator. This is of the grounded anode type, the anode being grids #2 and #4. An interesting feature is the manner of feeding the heater wires through the tubular coil comprising the lower end of the tank coil L9. This permits the heater and cathode to be at the same r.f. potential, and minimizes the possibility of the introduction of hum modulation. The output of the oscillator is coupled to the mixer 6BA6 control grid through a 6.8-$\mu\mu f$ capacitor (C25) and a 56-$\mu\mu f$ capacitor (C24), in conjunction with R8, a 3.9-megohm resistor. Owing to the a.c. grounded grids #2 and #4, the plate circuit is isolated from the oscillator circuit, specifically grid #1, so that the incoming signal cannot cause "pulling" of the local oscillator.

The i.f. output of the mixer and of the two i.f. amplifiers are fed through 10.7 mc i.f. transformers, below which are connected in series with the above the 455 kc standard and band "C" i.f. transformers. The latter units act as by-pass capacitors at 10.7 mc and so have no appreciable reactance at this higher i.f. frequency. On the other hand, the 10.7 mc i.f. transformers act substantially as direct connections through their coils (of relatively few turns) to the 455-kc transformers below them, at the lower intermediate frequency.

Thus both types of transformers can be permanently connected together without requiring any switching; the i.f. amplifier functions at either intermediate frequency. The 10.7 mc transformers are connected next to the plates because their stray capacity to ground has less effect on the tuning of the 455 kc i.f. transformers below them, as compared to a reverse sequence of connections.

The second i.f. stage has a 6AU6 tube, which is connected to a 6AU6 driver stage. This drives the ratio detector containing the 6AL5 double diode, whose output then feeds the audio amplifier (located on another chassis and not shown here). For a.m. the



(Courtesy RCA)

Fig. 57.—Complete schematic for Radio Chassis—Range Switch shown in f.m. position. Audio power amplifier is on a separate chassis (not shown).

second 455-kc i.f. stage feeds one of the diode plates of the 6AT6 detector-a.v.c.-audio amplifier stage. Switch S6 can be set so that either the a.m. detector or the f.m. ratio detector feeds the 1.5-megohm volume control and the grid of the 6AT6 tube, whose output feeds the audio plug (a.f. output).

The design of the ratio detector in this receiver is of interest. In the first place, the primary coil T6 is not inductively coupled to the secondary coil T7, but instead is coupled to it through a step-down coil in T6. The latter coil acts as the tertiary winding.

This portion of the receiver circuit is reproduced in Fig. 58.

center or carrier frequency, C72 tunes the loop to resonance.

There is no center tap on the secondary; instead, a center tap can be obtained at the junction of the step-down coil and C72. This is the "high" side of the audio output, and feeds the volume control potentiometer R42 through the deemphasis circuit R48 and C77, in conjunction with coupling capacitor C62.

The "low" side of the audio output is ground, but ground is not located at the midpoint of the diode load resistor R49; instead, it is at the bottom end. Recall, however, that stabilizing capacitor C81 maintains a constant or d.c. voltage across R49. Hence, there is no audio voltage between any

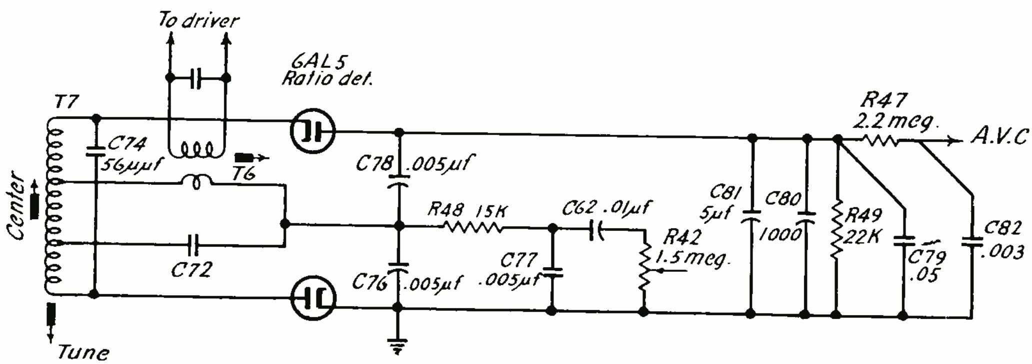


Fig. 58.—Ratio-detector circuit used in RCA Model 612V1 receiver.

The step-down winding is connected through C72 to the secondary as shown. The portion of the secondary included between the two points of connection acts very much like a separate winding inductively coupled to the entire secondary, so that this mesh of the circuit is essentially a link coupling circuit. Adjustment of the iron slug tunes the primary to resonance at the

two points of R49; so far as a.c. is concerned, any point of R49 may be grounded, and the same audio output obtained between such a point and the junction of C78 and C76 as between any other point and the junction.

The stabilizing time constant is C81 × R49 = $5 \times 10^{-6} \times 22{,}000 = 0.11$ sec. This is about the optimum value; the voltage will be stabilized against low audio frequencies, but will be able to follow the slower variations that occur owing to fading or other similar cause. The absence of a threshold carrier level makes this circuit superior to a limiter circuit in this respect: Amplitude modulation is rejected by the ratio-detector no matter how low the average carrier level may be, whereas the limiter ceases clipping when the carrier level drops below a certain minimum value.

Observe that a.v.c. voltage is obtained from the top end of the circuit. The a.v.c. output is double that of a ratio-detector in which the center point of R49 is grounded. Although it is therefore an advantage to ground the bottom end of the diode-load resistors instead of the center point, the latter connection has the advantage of much better rejection of low audio frequencies.

CONCLUSION

This concludes the assignment on frequency modulation. The general theory was first covered, and then a brief description given of the Armstrong and reactance-tube type of modulation. Following this, an analysis was made of the reaction of such a system to thermal and impulse noise, as well as to co-channel and adjacent channel interference. It was shown that by the use of wide frequency deviation, as well as limiter stages or similar circuits, the signal-to-noise ratio could be improved over that of ordinary amplitude modulation.

However, there is reason to believe that a.m. can be improved to the point where it can seriously compete with f.m. on the basis of signal-to-noise ratio.* Thus there is no one system of communication, destined to supplant all others, but several competing types, all of which must be understood by the student.

From a practical viewpoint, the f.m. transmitter tubes, by operating at a constant power output, are utilized more efficiently than under a.m. operation, and from this viewpoint f.m. may show a definite superiority to the ordinary a.m. system. Moreover, certain u.h.f. tubes such as the magnetron and klystron, are more readily frequency-modulated than amplitude-modulated, and this practical consideration may determine the preference for f.m. over a.m. at ultra-high frequencies.

The design considerations for an f.m. receiver are essentially the same as for an a.m. receiver, at least so far as the r.f., oscillator, and i.f. stages are concerned. The use of limiters, however, is essentially peculiar to f.m. receivers, and the f.m. detector is decidedly different from the a.m. detector. In this assignment the tuned-circuit detector, the balanced discriminator, the ratio-detector, and the locked-oscillator type of detector are all analyzed in considerable detail, and the advantages and/or disadvantages of each discussed.

Finally, a typical f.m. re-

*"Comparison of Amplitude and Frequency Modulation," by M. G. Nicholson, *Wireless Engineer*, July 1947.

ceiver employing a ratio detector is studied, and various practical features analyzed. Of particular interest are the departures, in details, of the components from their normal design, since these indicate the individual beliefs and "know-how" of the designer.

APPENDIX I

F.M. SIDE BANDS

Given the frequency-modulated wave

$$i = I \sin [\omega_c t + m_f \sin \omega_m t] \quad (1)$$

The angle is $\omega_c t + m_f \sin \omega_m t$. From trigonometry it is known that the sine of the sum of two angles is:

$$\sin(x + y) = \sin x \cos y + \cos x \sin y \quad (2)$$

Let $x = \omega_c t$ and $y = m_f \sin \omega_m t$. Then Eq. (1) can be written, in view of Eq. (2), as

$$i = I [\sin \omega_c t \cos (m_f \sin \omega_m t) + \cos \omega_c t \sin (m_f \sin \omega_m t)] \quad (3)$$

In Eq. (3) are two rather unusual expressions, $\cos (m_f \sin \omega_m t)$ and $\sin (m_f \sin \omega_m t)$. These involve the cosine and sine of an angle that is in itself a sine function of time. Each of these can be expanded into an infinite series of sine and cosine terms somewhat like the Fourier series. Each sine and cosine term has a certain amplitude; the amplitudes in turn are also expressible in certain infinite series known as Bessel functions. The latter series involve various powers of m_f and certain numbers. They have been evaluated, just like the the more usual sine, cosine, and logarithmic functions, in the form of tables, and also in the form of curves, similar to those of Fig. 10 in the text.*

*The tables and curves can be found in Jahnke and Emde's "Tables of Functions," Dover Publications.

Thus

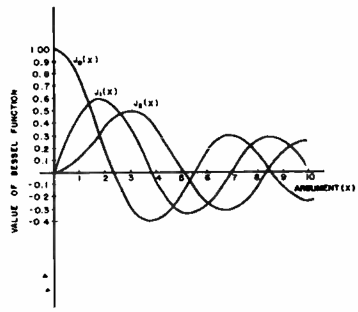
$$\sin (m_f \sin \omega_m t) = 2 J_1 (m_f) \sin \omega_m t + 2 J_3 (m_f) \sin 3 \omega_m t + 2 J_5 (m_f) \sin 5 \omega_m t + \ldots \quad (4)$$

and

$$\cos (m_f \sin \omega_m t) = J_0 (m_f) + 2 J_2 (m_f) \cos 2 \omega_m t + 2 J_4 (m_f) \cos 4 \omega_m t + \ldots \quad (5)$$

As mentioned previously in the text, the various J's are Bessel functions of the first kind; J_0 is of the first kind and zero order; J_1 is of the first kind and first order, and so on. For each J there is a corresponding table giving its value for different values of the argument m_f.

An inspection of Fig. 10, re-



(Courtesy of Communications, by Nathan Marchand.)

Fig. 10.—Plot of Bessel functions of the first kind and first three orders.

produced here, shows that $J_0(m_f)$ is equal to unity when m_f equals 0; i.e., $J_0(0) = 1$, and all the higher orders are zero for $m_f = 0$. When $m_f = 0$, there is no frequency modulation, and the J_0 term, which will be shown to represent the carrier amplitude, is a maximum (unity), and all the other J functions are 0; i.e., ($J_1(0) = 0$, $J_2(0) = 0$, $J_3(0) =$ 0, etc., so that all the side-band amplitudes are zero, as is to be expected. Also note that the successive peaks of any one function become progressively smaller as m_f increases, and also that *corresponding* peaks of the various functions are smaller, the higher the order of the function.

Substitute the relations of Eqs. (4) and (5) in Eq. (3), and group the terms so that the order of the J-coefficients are progressively higher:

$$\begin{aligned} i = I[&J_0(m_f)\sin\omega_c t \\ &+ 2\,J_1(m_f)\sin\omega_m t\cos\omega_c t \\ &+ 2\,J_2(m_f)\cos 2\,\omega_m t\sin\omega_c t \\ &+ 2\,J_3(m_f)\sin 3\,\omega_m t\cos\omega_c t \\ &+ \ldots.] \end{aligned} \tag{6}$$

Observe that each line of Eq. (6) involves the product of a sine and cosine term. It was shown earlier in the text that for similar products that occur in amplitude modulation, a trigonometric expansion can be made in terms of sum and difference frequencies (thereby giving rise to side bands). Thus,

$$\sin A\cos B = \frac{1}{2}\sin(A + B) + \frac{1}{2}\sin(A - B) \tag{7}$$

Take the second line of Eq. (6):

$$2\,J_1(m_f)\sin\omega_m t\cos\omega_c t$$

and substitute $\omega_m t$ for A and $\omega_c t$ for B in Eq. (7). There is obtained:

$$2\,J_1(m_f)\sin\omega_m t\cos\omega_c t = 2\,J_1(m_f) \times\left[\frac{1}{2}\sin(\omega_m + \omega_c)t + \frac{1}{2}\sin(\omega_m - \omega_c)t\right] \tag{8}$$

Since $\sin(\omega_m - \omega_c)t = -\sin(\omega_c - \omega_m)t$, Eq. (8) can be written (after multiplying through by 2) as

$$2\,J_1(m_f)\sin\omega_m t\cos\omega_c t = J_1(m_f) \times[\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] \tag{9}$$

These represent the first-order side bands of frequency $(\omega_c + \omega_m)$ and $(\omega_c - \omega_m)$.

The remaining terms can be expanded in similar fashion; they will involve higher order side bands of frequencies equal to the carrier frequency ω_c ± multiples of the audio frequency ω_m. The end result is

$$\begin{aligned} i = I\{&J_0(m_f)\sin\omega_c t \\ &+ J_1(m_f)[\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] \\ &+ J_2(m_f)[\sin(\omega_c + 2\,\omega_m)t + \sin(\omega_c - 2\,\omega_m)t] \\ &+ J_3(m_f)\ \sin(\omega_c + 3\,\omega_m)t - \sin(\omega_c - 3\,\omega_m)t] + \ldots.\} \end{aligned} \tag{10}$$

This is Eq. (22) of the text.

FREQUENCY MODULATION.—If the frequency shift (or RATE of angular deviation) is proportional to the amplitude of the sound source, the system is frequency modulated.

PHASE MODULATION.—If the phase shift (or MAGNITUDE of angular deviation) is proportional to the amplitude of the sound source, the system is phase modulated.

To determine whether a system is F.M. or P.M., consider the following:

Assume a sound source of *constant amplitude.* Suppose the frequency of the sound source is 1,000 c.p.s. The output of the system being modulated will contain both phase and frequency deviations, since the two must inherently occur together just like distance traversed and velocity. Thus, if an automobile has a certain velocity, it must be traveling a distance; likewise, if it is traversing a certain distance, it must have velocity.

Suppose the peak frequency deviation from the nominal carrier frequency is $f_D = \pm 30$KC. Then the accompanying peak phase deviation is $\theta_m = f_D/f_m = \pm 30{,}000/1{,}000 = 30$ radians. Note that if $f_m = 1{,}000$ c.p.s., and $f_D = \pm 30$KC, then θ_m must be ±30 radians; i.e., all three cannot be chosen indiscriminately.

Now suppose the audio frequency increases to 2,000 c.p.s., BUT THE AMPLITUDE OF THE WAVE DOES NOT CHANGE. Then if the system is to be F.M., the frequency deviation, f_D, *must remain* ±30KC, although, of course, this deviation now occurs 2,000 times per second instead of 1,000 times per second.

Owing to the greater number of times that f_D now varies about the carrier value, the peak phase deviation will have to be HALF as much, or ±15 radians instead of ±30 radians. Then, and only then, will f_D remain equal to ±30KC, and only then will we have FREQUENCY MODULATION.

Suppose, on the other hand, PHASE deviation is desired. Then, whether the audio tone is 1,000 c.p.s. or 2,000 c.p.s., θ_m must stay constant at ±30 radians, PROVIDED, ONCE MORE, THAT THE AMPLITUDE OF THE AUDIO TONE DOES NOT CHANGE. In this case the frequency deviation for the 2,000 cycle audio tone will be DOUBLE that for the 1,000 cycle audio tone; i.e., it will be ±60KC instead of ±30KC. Note that the phase deviation remained the same in spite of the fact that the audio frequency doubled, because the audio amplitude did not change, therefore, the system represents PHASE modulation.

EXERCISE PROBLEMS

F-M RECEIVERS

1. A 300KC carrier is amplitude - modulated with a 500-cycle tone. Give the frequencies of the side bands.

2. A 300mc carrier is frequency - modulated with a 300 cycle tone. Give the first five orders of side bands.

3. Phase shift is 800 degrees frequency - modulated with a 300 cycle tone. Suppose the carrier is now modulated with a 500 cycle tone of twice the amplitude. What is the phase shift in this case?

4. Reference problem 3. What would be the phase shift for 500 cycle tone if the carrier were phase - modulated?

5. A carrier frequency of 100mc, and two thermal noise components of equal amplitude, one aas a frequency of 100.002mc, and the other 100.004mc. What audio frequency will each produce in the output?

6. An f-m receiver is to amplify a maximum audio band width of 15,000c.p.s. maximum frequency deviation for 100 per cent modulation is to be 85KC. How does the signal-to-noise ratio compare with an a-m receiver of the same audio band width with 100 per cent modulation for a-m? Use (N/C) = .1 input.

7. An i-f amplifier for an f-m receiver employs a tube whose R_p = 2 megohms, and whose G_m = 6000 μmhos. The entire band width is to be 175KC, and the intermediate frequency is 14mc. The primary and secondary tuning capacitors are 40$\mu\mu$F each, and the primary and secondary coil Q'$_s$ are $Q_p = Q_s = 95$. The secondary coil drives the grid of another i-f amplifier tube.

   (a) Find the necessary coefficient of coupling K, and the critical coupling, K_c.

   (b) Find the variation in gain over the pass band.

   (c) Find the gain at the center frequency.

ANSWERS TO EXERCISE PROBLEMS

F-M RECEIVERS

1. 3,000,500 c.p.s.
   2,999,500 c.p.s.

2. 300,000,000 ± 300
   300,000,000 ± 300
   300,000,000 ± 900
   300,000,000 ± 1200
   300,000,000 ± 1500

3. 480°

4. 1600°

5. 2000 cycle tone
   4000 cycle tone

6. Signal Ratio 56.7:1
   For a-m 10:1

7. (a) $K = .0145$
   $K_c = .0106$

   (b) 1.18

   (c) 76.3

F.M. RECEIVERS

EXAMINATION

1. What characteristic of an amplitude-modulated wave remains unchanged?

2. What characteristic of a frequency-modulated wave remains unchanged?

3. A 1,000-kc carrier is amplitude-modulated with a 200-cycle tone. Give the frequencies of the side bands.

4. A 100-mc carrier is frequency-modulated with a 200-cycle tone. Give the frequencies of the first five orders of side bands.

5. (A) The phase shift or deviation is 720 degrees. Suppose the carrier is now frequency-modulated with a 400-cycle tone instead of 200 c.p.s. *of the same amplitude*. What is the phase shift in this case? (Refer back to problem 4).

5. (B) What would be the phase shift or deviation for the 400-cycle tone if the carrier were phase-modulated?

6. Consider a carrier frequency of 90 mc, and two thermal noise components of equal amplitude (small compared to that of the carrier), but one has a frequency of 90.001 mc; and the other 90.002 mc. What audio frequency will eacn produce in the output?

7. The f.m. receiver is to amplify a maximum audio band width of 60 to 12,000 c.p.s. The maximum frequency deviation for 100 per cent modulation is to be 75 kc. How does the signal-to-noise ratio of this receiver compare with that of an a.m. receiver of the same audio band width, receiving a 100 per cent amplitude-modulated wave? (Assume no preemphasis or deemphasis.)

8. From a practical viewpoint, what is the major feature of the suppression of the weaker of two waves by the stronger, in co-channel interference of an f.m. receiver?

9. An i.f. amplifier for an f.m. receiver employs a tube whose R_p = 1 megohm, and whose G_m = 4,000 μmhos. The entire band width is to be 141.4 kc, and the intermediate frequency is 10.7 mc. The primary and secondary tuning capacities are 30 μμf each, and the primary and secondary *coil* Q's are Q_p = Q_s = 80. The secondary coil drives the grid of another i.f. amplifier tube.

(A) Find the necessary coefficient of coupling, k, and the critical coupling kc.

(B) Find the variation in gain over the pass band.

9. (C) Find the gain at the center frequency.

(D) What effect will a limiter stage have upon the overall flatness of response if the signal is strong, and if it is very weak?

10. (A) What is the advantage of a ratio-detector over the ordinary balanced discriminator circuit?

(B) What are the advantages of using a capacitor instead of a battery in the ratio-detector?

(C) What advantage may the locked-oscillator type of f.m. detector have over the ratio detector?