MODULATION OF ALTERNATING CURRENT WAVES

BY

KARL M. DOERING

A Thesis Submitted to the Graduate Committee
For the Degree of

MASTER OF SCIENCE

in

Electrical Engineering

Approved:

W. A. Murray
Head of Department

E B Norris
Dean of Engineering

Louis O'Shaughnessy
Chairman, Graduate Committee

Virginia Polytechnic Institute

1939

125633

CONTENTS

ILLUSTRATIONS

I INTRODUCTION

1. Importance of the Problem

The importance of electrical communication is so well known that it needs no explanation in itself. It is obvious that any phase of the communication problem that offers hope for substantial improvement, in efficiency or effectiveness, is well worth investigating. In fact, any study that leads to a better understanding of the subject should be valuable to the electrical engineer.

2. Meaning of Modulation

All electrical communication is accomplished by modulation of some kind of electric current. Modulation consists of controlling this current in such a way that variations of the input will be present in the output of the circuit. If some method is used to detect these variations in the output, the original controlling force (or signal) can be reproduced, and can be used to convey intelligence from one end of the circuit to the other. The simplest example of such a system is a direct current telegraph circuit, in which a direct current is modulated with a key which merely starts and stops the current. An example of speech modulation is the ordinary d.c. telephone circuit. The telephone transmitter is a modulating device that changes its resistance as speech waves strike a diaphragm, thus changing the current through the circuit.

3. Modulation of High Frequencies

When radio transmission was developed, it became necessary to modulate high frequency alternating currents, because low frequency or direct currents could not be transmitted without wires. Simple telegraph modulation was first used to transmit wireless signals in code. Speech modulation was accomplished later so that the voice and other sounds could be transmitted. The radio telegraph is still used to a large extent because of its comparative simplicity and economy, and its effectiveness for communication under adverse conditions.

As an offspring of radio transmission, the high frequency carrier was adapted to wired communication in the form of carrier current telephone systems. This method makes it possible to transmit a relatively large number of separate and distinct signals over a single pair of wires. The greatest expense encountered by the telephone companies is building and maintaining outside lines, which represent a much larger investment than does the central office equipment. Therefore the use of carrier currents for some minimun distance will save enough to pay for the additional central office equipment, provided there is sufficient traffic demand, and for longer distances the carrier current systems can effect large savings in the cost of operation. Hence the problem of high frequency modulation is becoming more important in telephone practice as well as in radio.

4. Acknowledgments

The writer wishes to express appreciation of assistance received from Prof. W. A. Murray, and the staff of the electrical engineering department. In particular, Prof. R. E. Bailey has given much valuable help in this study.

II PROCEDURE

5. Object

The primary purpose of this thesis is to present, in clear and logical form, an analysis of high frequency modulation. This will be done in order to explain each of the various methods of obtaining modulation, and the characteristics of the resultant waves. The question of side bands will be considered in detail. The subject of circuits and associated apparatus is of secondary importance, emphasis being placed on the fundamental electric theory. From this point of view, possibilities for improvement and new developments will be investigated.

6. Method of Analysis

In dealing with alternating currents, the phenomena involved are not always easy to understand. Calculations become more complicated than in the case of direct currents. Mechanical analogies and physical explanations are less obvious, sometimes seeming utterly incomprehensible. This is particularly true in the study of modulation, so that most engineering students think that high frequency modulation is a difficult subject. The modulated carrier is not a simple sinusoidal wave, but rather the sum of several distinct waves which can be expressed as sine or cosine functions.

The current wave can, of course, be represented graphically, and it can be seen on the oscillograph. These pictures may be helpful in showing the composition of the complex wave, but they do not explain the cause or effect of the components. The difficulty of getting a physical picture of cause and effect accounts for much confusion on the subject.

For the reasons stated above, the problem has been investigated on a mathematical basis. Only in the mathematical equations do we find clear and logical answers to the following questions: (1) How can a high frequency current be modulated ? (2) What is the nature of the resultant wave, and why ? (3) How will it respond to electric circuits? The proof of the answers is given in the derivation of the equations.

7. Laboratory Investigation

Limitations of time and equipment prohibit any very complete or conclusive laboratory research within the scope of this work. The problem might perhaps provide material for a departmental research project, carried on over a period of years, but it is too extensive to be covered thoroughly in a thesis for a graduate degree, within a period of nine months.

Laboratory apparatus was constructed, however, to illustrate practical apparatus for obtaining each type of modulation. The circuit used has many interesting features, which will be discussed in Chapter V. Unfortunately, much time was wasted in the design and construction of an oscillator that was not needed. A commercial product (Western Electric) already available in the communications laboratory was found to be satisfactory for the purpose. The modulation circuit is not original, but the apparatus and circuit constants were selected by guess or experiment. Calculations of circuit design are not within the scope of this work.

8. Results

It is hoped that the information given in the following sections will be valuable as a clear and complete explanation of the modulation problem. Although much of the data has been published in periodicals and text books, the writer has not seen it assembled and correlated in this form. In fact, many of the articles leave the burden of proof to the reader, passing it off with the statement that "it can easily be shown". The authors do not always point out the meaning of results obtained in this manner.

No claim is made for complete or conclusive laboratory results; they serve to supplement and illustrate the theory. Perhaps the circuit shown will be of interest to students planning to study some phase of communication engineering.

As a result of the investigation, the writer has naturally formed certain opinions concerning future trends in electric communication, and possibilities for improvements. These will be discussed after the theory and data have been presented.

III DERIVATIONS

9. Fundamental Equations

In order to modulate an alternating current wave, it is necessary to vary some magnitude of the wave in accordance with a signal wave.* Since the current is a periodic function, it can be represented by an equation of the form,

$$i = A \sin b \qquad (1)$$

where i is the instantaneous value
A is the maximun value
b is a function of time (t)

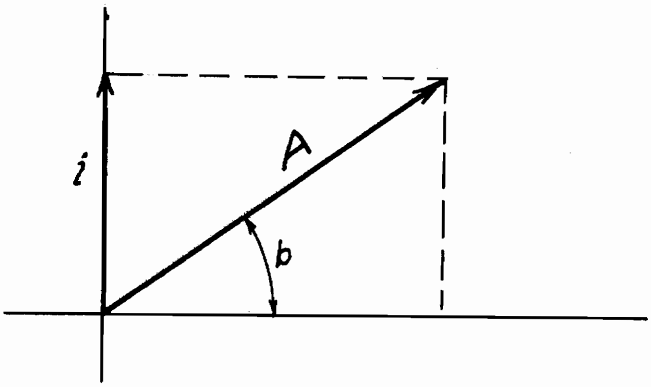

Fig. 1. Vector Representation of Periodic Function

---

*See Roder, "Amplitude, Phase, and Frequency Modulation", <u>Proceedings of the Institute of Radio Engineers</u>, Dec. 1931, page 2145.

By definition, the angular velocity of the vector (A) is the rate of change of the angle (b) with respect to time. That is,

$$w = \frac{db}{dt} \qquad (2)$$

where w is the angular velocity

Hence, $db = w\,dt$

And integrating, $$b = \int w\,dt + \phi \qquad (3)$$

where ϕ is the constant of integration

If the angular velocity (w) is constant, it follows that:

$$b = wt + \phi$$

Substituting this value of b in equation (1),

$$i = A \sin(wt + \phi) \qquad (4)$$

which is the equation for the current in terms of time, with constant frequency, phase, and amplitude. This is, of course, the sine wave, familiar to all electrical engineers. As Roder points out, there are three independent magnitudes, as follows:

A - amplitude

ϕ - phase angle

w - angular velocity = 2π x frequency

Modulation can be obtained by means of a periodic change of any one of the three independent magnitudes.

10. Amplitude Modulation

Practically all modulation systems in use at the present time make use of a change in the amplitude of the carrier to transmit the signal.

If the amplitude (A) of the current vector is subjected to a periodic change, it can be represented by:

$$A = A_0 + B \sin ut \tag{5}$$

where A_0 is the average value of the vector
B is the amplitude of the modulating signal
u is the angular velocity of the modulating frequency

The amplitude (B) of the audio signal must not be larger than A_0; otherwise the carrier current vector (A) will take on a negative value for some negative values of sin ut. This negative magnitude has no physical meaning and does not exist. It is essential to distinguish between negative sense, which merely indicates a change in the direction of flow, and negative magnitude, which is physically impossible. Since the negative value of (A) does not exist, the function becomes discontinuous and the equation does not hold when B is greater than A_0. Such a condition indicates that the signal will be distorted, and it must be avoided in operation.

Since B and A_o are both constants, the value of B can be written in terms of A_o, so that:

$$B = m_a A_o$$

where m_a is the constant of proportionality, and must not be greater than unity.

The constant m_a may be called the amplitude factor of modulation. Putting the above value of B into (5):

$$A = A_o + m_a A_o \sin ut$$

or: $$A = A_o(1 + m_a \sin ut) \qquad (6)$$

Substituting (6) into equation (4):

$$i = A_o(1 + m_a \sin ut)\sin(wt + \phi) \qquad (7)$$

Since ϕ is an arbitrary constant of integration, it can be evaluated for any given conditions. By proper choice of the axes, ϕ can be made equal to zero, and (7) becomes:

$$i = A_o(1 + m_a \sin ut)\sin wt$$

or: $$i = A_o(\sin wt + m_a \sin wt \sin ut) \qquad (8)$$

From the trigonometric relation,

$$\sin a \sin b = \tfrac{1}{2}\cos(a - b) - \tfrac{1}{2}\cos(a + b)$$

Equation (8) can be evaluated as follows:

$$i = A_o\left[\sin wt + \tfrac{1}{2}m_a\cos(w - u)t - \tfrac{1}{2}m_a\cos(w + u)t\right] \qquad (9)$$

Equation (9) may be called the side band equation, because it includes terms for the side bands which exist as components of the modulated wave. In these derivations, it is assumed that the modulating current is a single wave of sinusoidal form. If the modulating current contains a number of different frequencies, as it almost always does in actual practice, it will be a complex wave but it can be reduced to a series of sine and cosine functions. Although there will be more terms in the equations, the derivation can be carried through in the same manner and the basic relations will be the same. As indicated in equation (9), there will be two side band frequencies for each audio frequency; one side band frequency equal to the sum of the carrier and audio frequencies, and another equal to the difference of the carrier and audio frequencies.

The intelligence must be conveyed in the side bands, because the carrier term alone is a function of time only, and cannot recreate the audio variations in the receiver. Therefore it is desirable to keep the modulation factor as near unity as possible in order to secure the maximum possible transmission of the signal.

Another important point illustrated by equation (9) is that the carrier frequency must be higher than the highest signal frequency. If it is not, then the lower side band will produce a hypothetical negative frequency, which is quite as impossible as a negative magnitude. It is true, of course, that the lower side band can be written as a positive frequency, because cos(a - b) is equal to cos(b - a). However, it will be noted that the upper side band (a cosine term also) is subtracted algebraically from the sum of the first two terms, and that under certain conditions a negative magnitude of the current will be indicated. In any event, the equation becomes discontinuous and distortion occurs if the signal frequency is higher than the carrier frequency.

11. Phase Modulation

In this case, the amplitude and frequency of the carrier current are constant, but the phase angle ϕ is assumed to change its value periodically. Then ϕ can be expressed as a function of time, thus:

$$\phi = \phi_0(1 + k \sin ut) \qquad (10)$$

where u represents the signal frequency
k is the constant of proportionality between ϕ_0 and the signal amplitude

which is similar to equation (6).

The value of ϕ given in equation (10) can now be substituted into (4) to find the current:

$$i = A \sin\left[wt + \phi_0(1 + k \sin ut)\right]$$

or: $$i = A \sin(wt + \phi_0 + m_\phi \sin ut) \qquad (11)$$

where $m_\phi = k\phi_0$

The value of ϕ_0 which remains in equation (11) represents an arbitrary constant, and can be made equal to zero without changing the function. Note that ϕ_0 does not have the same significance that A_0 has in the case of amplitude modulation; it merely represents an angle with respect to some arbitrary reference axis, whereas A_0 is actually the amplitude of the unmodulated carrier. Hence the current equation can be written as:

$$i = A \sin(wt + m_\phi \sin ut) \qquad (12)$$

The constant m_ϕ represents the maximum angle through which the current wave is varied. It is theoretically unlimited but is subject to limitations for practical operation similar to those which will be considered in the next chapter with respect to frequency modulation. Expanding (12),

$$i = A\left[\sin wt \cos(m_\phi \sin ut) + \cos wt \sin(m_\phi \sin ut)\right] \qquad (13)$$

This equation cannot be evaluated by simple trigonometric relations as before. However, by use of infinite series and the theory of Bessel functions* the following relations are known:

$$\sin(x \sin r) = 2\,J_1(x)\sin r + 2\,J_3(x)\sin 3r + 2\,J_5(x)\sin 5r + \ldots\ldots \qquad (14)$$

$$\cos(x \sin r) = J_0(x) + 2\,J_2(x)\cos 2r + 2\,J_4(x)\cos 4r + \ldots\ldots \qquad (15)$$

In these equations, the symbol $J_n(x)$ means the Bessel function of the first kind, and n^{th} order, for the argument x.

By means of the relations given above, equation (13) can be written in terms of Bessel functions,

$$i = A\Big[J_0(m_\phi)\sin wt + 2\,J_1(m_\phi)\sin ut \cos wt + 2\,J_2(m_\phi)\cos 2ut \sin wt + 2\,J_3(m_\phi)\sin 3ut \cos wt + \ldots\ldots\Big] \qquad (16)$$

Expanding the terms of equation (16) by trigonometric relations, the side band equation is obtained; it is similar in form to equation (9) for amplitude modulation, just as

---

*See Appendix I; For the definition and derivation of Bessel functions, see Sokolnikoff, I.S. and E.S., "Higher Mathematics for Engineers and Physicists"; Also Byerly, "Fourier Series and Spherical Harmonics".

equation (16) corresponds to (8):

$$i = A\Big\{J_0(m_\phi)\sin wt + J_1(m_\phi)\big[\sin(w + u)t - \sin(w - u)t\big] + J_2(m_\phi)\big[\sin(w - 2u)t - \sin(w + 2u)t\big] + J_3(m_\phi)\big[\sin(w + 3u)t - \sin(w - 3u)t\big] + \cdots\cdots\Big\} \qquad (17)$$

Note that the phase modulated wave contains an infinite number of side bands. It includes not only the sum and difference terms of the carrier and signal waves, but also terms for the sum and difference of the carrier and all integral multiples of the signal.

12. Frequency Modulation

The third method of modulation is accomplished by keeping the amplitude and phase of the wave constant while the frequency is varied periodically so that:

$$w = w_0(1 + k \cos ut) \qquad (18)$$

The frequency is written as a cosine function rather than a sine function to simplify the derivation. The change could be made, of course, in evaluating the phase angle (ϕ) for the particular conditions required. Equation (18) cannot be substituted directly into (4) as in previous cases, because (4) was derived with the assumption that w was constant. Substituting into (3), integrating, and putting the arbitrary constant of intègration equal to zero,

$$i = A \sin(w_o t + \frac{w_o k}{u} \sin ut)$$

The equation can be written in a simplified form by dropping the subscript "o" and defining $w = w_o$; also the coefficient of the audio term can be defined as the constant of modulation m, and the equation becomes:

$$i = A \sin(wt + m \sin ut) \qquad (19)$$

$$\text{where } w = w_o$$
$$m = \frac{w\ k}{u}$$

Since equation (19) is identical to (12) except for the different definition of the constant terms, it can be written immediately as:

$$\begin{aligned} i = A\Big[J_0(m)\sin wt &+ 2\ J_1(m)\sin ut \cos wt \\ &+ 2\ J_2(m)\cos 2ut \sin wt \\ &+ 2\ J_3(m)\sin 3ut \cos wt \\ &+ \ldots\ldots\ldots\Big] \end{aligned} \qquad (20)$$

Although equation (20) is written exactly like (16), note that the factor of modulation (m) is not the same in the two equations. It does show, however, that the characteristics of the frequency modulated wave are practically the same as for a wave produced by phase modulation.

The side band equation, like that for phase modulation, can be written as:

$$\begin{aligned} i = A\Big\{ J_0(m)\sin wt &+ J_1(m)\big[\sin(w + u)t - \sin(w - u)t\big] \\ &+ J_2(m)\big[\sin(w - 2u)t - \sin(w + 2u)t\big] \\ &+ J_3(m)\big[\sin(w + 3u)t - \sin(w - 3u)t\big] \\ &+ \cdots\cdots\cdots \Big\} \end{aligned}$$

As before, there are an infinite number of side bands. The significance of the side bands for practical operation, and the relation of their amplitudes for various values of m, will be discussed next.

The equations also show that for a single audio frequency there is no differency between phase and frequency modulation. If more frequencies are added or if the one frequency is changed, the two cases will give different results.

IV FREQUENCY MODULATION

13. The Noise Problem

The effect of extraneous noises on the receiver is perhaps the most serious defect of radio transmission at the present time. Since amplitude modulation is used, the receiving system must be sensitive to all variations of the input voltage in order to detect the signal in the amplitude modulated wave. Unfortunately, the radio transmitter is not the only source of radio frequency waves; irregular and unpredictable electromagnetic waves at radio frequencies can, and do, originate from a great many sources, both natural and artificial. Such waves cover the entire frequency spectrum, and have amplitudes comparable to that of the radio signals. The noise, or "static", created in the receiver by such disturbances, is always annoying to the listener, and often makes the reception of weak radio signals impossible.

Armstrong* has suggested the use of frequency modulation to reduce noise disturbances. His outstanding

---

*Armstrong, Edwin H., "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation", Proceedings of the Institute of Radio Engineers, May 1936, page 689 work in this field has attracted wide attention among radio engineers who are always looking for a solution to the noise problem. It is apparent that, if the receiver is designed to respond to variations in the frequency of the radio carrier, it can be made insensitive to changes in amplitude. The probability that electrical disturbances will produce an audio signal in such a system is much less than in an amplitude system.

14. Transmission Requirements

Obviously the transmission system must undergo considerable change to produce a frequency modulated signal. Armstrong sums up the requirements thus:

> "The conditions which must be fulfilled to place a frequency modulation system upon a comparative basis with an amplitude modulated one are the following:
>
> 1. It is essential that the frequency deviation shall be about a fixed point. That is, during modulation there shall be a symmetrical change in frequency with respect to this point, and over periods of time there shall be no drift from it.
>
> 2. The frequency deviation of the transmitted wave should be independent of the frequency of the modulating current and directly proportional to the amplitude of that current."

He goes on to say that the frequency modulator must produce a minimum phase shift of about forty-five degrees with relation to the unmodulated carrier when the audio amplitude is a maximum, in order to obtain the equivalent of 100% modulation. He defines "100% modulation" as that degree of modulation necessary to produce a signal capable of being handled effectively by the receiver in the presence of disturbing currents.

Armstrong's system should not be confused with phase modulation. He specifies that the <u>frequency deviation</u>, the actual change in frequency produced by modulation, must be directly proportional to the amplitude of the modulating current. It was shown in the derivation (pages 18 to 20) that under those conditions the phase shift will depend upon the audio frequency and amplitude both. The 45 degree minimum phase shift may be considered as an index to indicate when full modulation is adequate to produce a good signal in the receiver

15. Side Band Considerations

Upon glancing at the equations for the frequency modulated wave (pages 19 and 20), it would seem that the infinite number of side bands prohibits any practical application of such a wave. It will be shown, however, that satisfactory results can be obtained by using only the first pair of side bands, and eliminating all others.

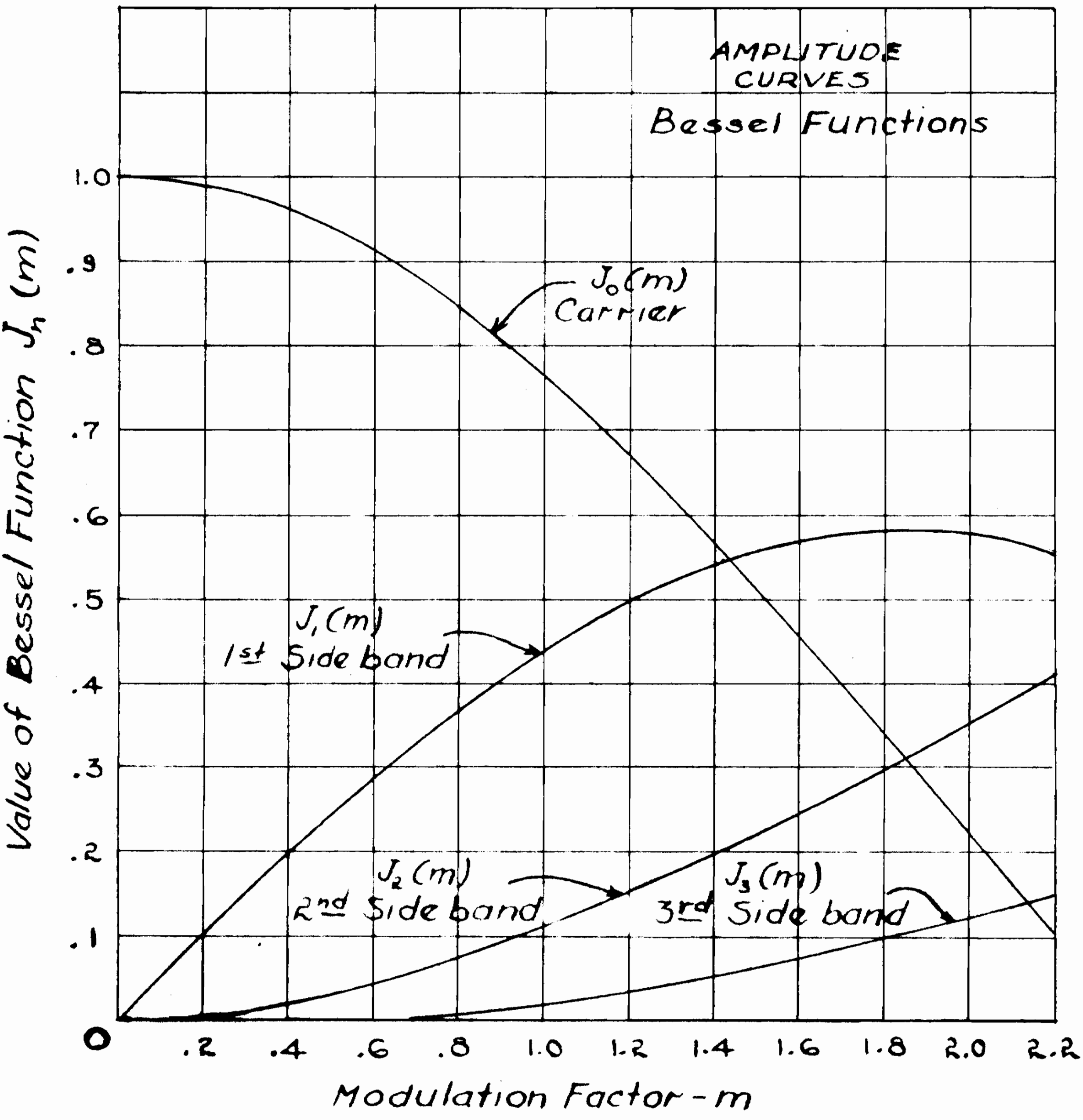

Fig. 2. Amplitude Curves - Bessel Functions

Before proceeding with this discussion, it might be well to call attention to the fact that the subject is not easy to understand at a quick glance. The difficulty arises, not because any of the points presented are hard to understand, but because of the large number of factors concerned. Therefore the meaning of the symbols, and the relations between them, become very confusing unless the subject is studied carefully. The frequency modulation equation (20) on page 19, which may be called the vector equation, and the side band equation on page 20 will be referred to in some of the explanations. The reader may find it convenient to copy those equations on a separate paper to save trouble in referring to them.

Now consider the side band equation. It will be observed that the amplitudes of the carrier and respective side band terms depend upon certain Bessel functions of the modulation factor, m. These Bessel functions are of integral order, from zero to infinity. The relative magnitudes of the first four terms are illustrated in Fig. 2 page 24, where the corresponding Bessel functions are plotted for various values of m. The curves show that for values of m less than about one-half, the second side band term is no more than about 10% of the first, and the third term is practically negligible. Within that range, Bessel functions of higher order do not appear in the first two decimal places. From a study of the conditions, it seems probable that, if m is limited to an arbitrary maximum of about 0.5, satisfactory transmission can be obtained by using only one pair of side bands, all others being eliminated in transmission.

16. Modulation Curve

After selecting the limiting value for the modulation factor, curves can be plotted to show the relation between the frequency deviation and the amplitude of the audio signal current. For perfect modulation, the frequency deviation should of course be directly proportional to the audio amplitude; that is, the relation between them should be linear. To illustrate the relation obtained with only one pair of side bands, a modulation curve for that condition is shown on page 27.

In calculating the modulation curve, the phase shift angle was taken as a measure of the frequency deviation, as indicated on the ordinate scale. That is not strictly true for frequency modulation; the relation shown is really a phase modulation curve. In the derivation, it was proven that the modulation factor (m) must vary inversely with the signal frequency, in order that the frequency deviation shall be independent of that frequency.

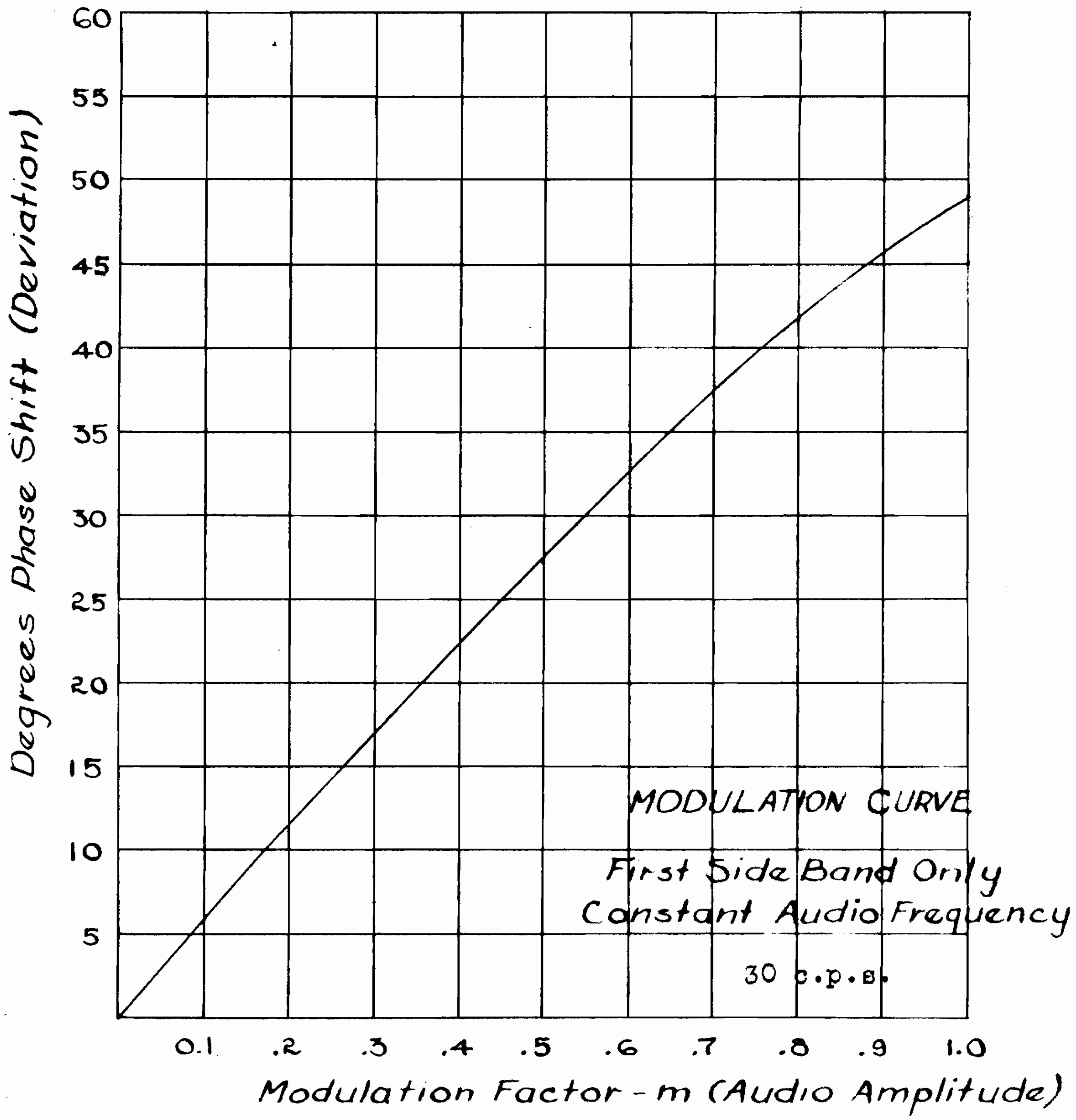

Fig. 3. Modulation Curve - First Side Bands Only

Thus Fig. 3 holds for frequency modulation only when the audio modulating frequency is constant, and that condition was assumed in calculating the curve. The problem is not quite so simple because m and ∅ are functions of u (the modulating frequency), and the relationship is further complicated by the absence of the higher order side band terms. However, the curve does give a fairly good picture of the linearity achieved under the conditions stated. A glance at Fig. 3 reveals that the curve is almost straight and it appears to have only a smooth general curvature. There are actually harmonic variations present, but their magnitude is negligible, being so small that it was impractical to show them on the scale used.

Another complication arises from the variations in the carrier term within the range selected for m. Referring to the vector equation for the current, and to the amplitude curves on page 24, it can be seen that the carrier amplitude will vary somewhat with changes of m. Keep in mind the fact that m is a function of the frequency as well as the amplitude of the modulating current, both of which relations are complicated by the absence of the higher order side bands, and that the changes in amplitude are not sinusoidal. Hence there will be a small degree of amplitude modulation, producing some rather complex terms.

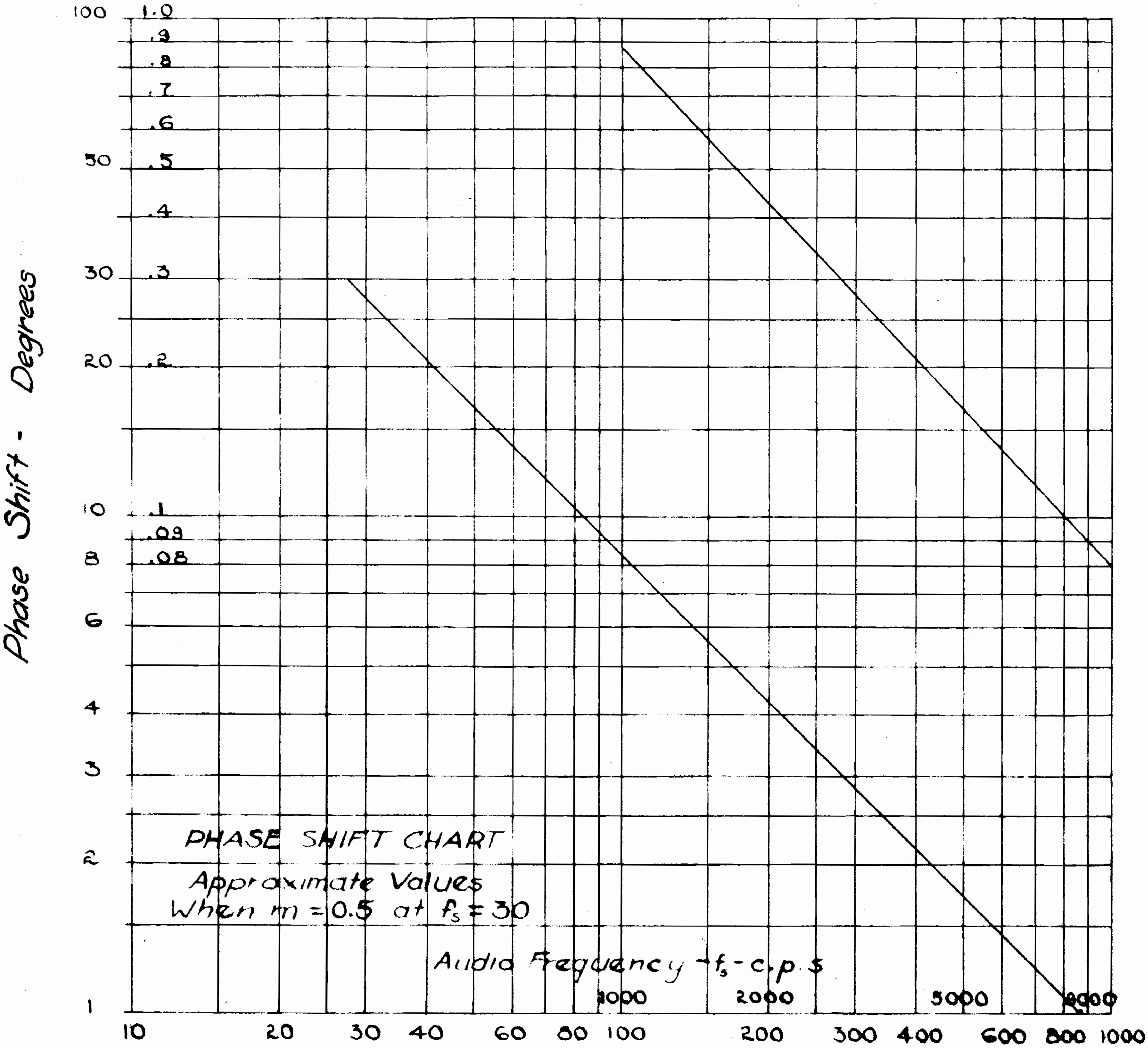

Fig. 4. Phase Shift Chart

Fortunately, the magnitude of such terms will be small, and they may be of no importance in practical operation. The amplitude modulation is caused, of course, by side band elimination. Inspection of the vector equation shows that the coefficient of each term which is in phase with the carrier is an even order Bessel function of m, multiplied by two. Calculations reveal that the summation of all such coefficients, plus that of the carrier, will always be unity, regardless of the value of m. Therefore, there would be no amplitude modulation if all terms were present; a condition which was required in the derivation in the first place.

17. Phase Shift

An important requirement for satisfactory frequency modulated transmission is that the degree of modulation shall be sufficient to produce a good signal in the receiver. Armstrong states that a minimum of 45 degrees phase shift is necessary for the equivalent of 100% modulation; that is, at least 45 degrees phase shift with a maximum amplitude of the modulating current. This requirement must be fulfilled by the transmission system without exceeding the limits which have been set for the modulation factor. Therefore it is necessary to consider the amount of phase shift produced by modulation under the conditions established for operation.

Referring to the vector equation, note that the greatest angle of phase shift will be, approximately, the inverse tangent $2J_1(m)/J_0(m)$. Since m is a function of the audio modulating frequency, it is necessary to decide upon the audio range to be transmitted. For high fidelity transmission, a range of 30 to 10,000 cycles per second is desirable; then the maximum value of m will occur at 30 cycles. Remember that the value of m is limited to about 0.5 because of other considerations. Hence the phase shift at the carrier fundamental frequency is limited to about thirty degrees at 30 cycles, and will decrease to about 0.08 degree at 10,000 cycles. The relation between the audio frequency and the phase shift angle has been plotted in the phase shift chart (Fig. 4) on page 29. In this case, the linear relation on logarithmic scales does not hold exactly because of the Bessel function terms, but it represents a fairly good approximation.

With the conditions established, it is impossible to obtain 45 degrees phase shift at the carrier fundamental frequency. Armstrong's work with receivers has shown that a minimum of 45 degrees, or at least something close to it, is essential in order to secure a fully modulated signal in the receiver without recourse to methods which would have an adverse effect on the signal-to-noise ratio, and destroy the only advantage of frequency modulation. Therefore the only way that a sufficiently large phase shift can be transmitted is by multiplying the fundamental frequency and thus multiplying the phase shift angle. Consequently the frequency of the modulated wave must be multiplied more than 500 times in order to secure adequate phase shift for the condition of full modulation at all audio frequencies.

18. Carrier Frequency

It has been shown that the phase shift depends upon the modulation factor, m. Referring to the definition of m under equation (19) page 20, note that m varies directly with the carrier frequency, represented by w. However, w is a constant under operating conditions, and it is possible to obtain the same value of m for any value of w by adjusting the audio amplitude, which is represented by k. Hence the carrier fundamental frequency may be considered independent of m, and is not limited by previous considerations. Since the carrier frequency must be higher than the highest modulating frequency, the absolute minimum for the carrier fundamental will be ten kilocycles. After multiplying 500 times, the absolute minimum frequency for the transmitted wave will be 5000 kilocycles. The intermediate range of radio frequencies, such as the present broadcast band, cannot be used by a frequency modulation system. For practical operation, the transmitted frequency may be well up into the "ultra-high" range, where the distance of radio transmission is limited by the distance of vision.

19. Band Width

In an amplitude modulation system the band width is equal to twice the highest audio frequency because the side bands from adjacent stations must be kept separate. If the side bands over-lap they will beat against each other, causing changes in amplitude that will be audible in the receiver as interference between stations. In the case of frequency modulation, Armstrong has demonstrated that overlapping side bands are practically never capable of causing frequency changes that would be audible as interference in a frequency modulated receiver. Therefore Armstrong maintains that the effective band width for a frequency system is determined by the frequency deviation and not by the side band frequencies. On that basis, he demonstrates that a band width or "channel" 150 kilocycles wide is desirable.

V CIRCUIT OPERATION

20. Amplitude Modulation

Circuits for amplitude modulation are fairly simple and very well known; some of them are described in every radio text book. In general, modulation is obtained by varying the voltage on some electrode of a vacuum tube amplifier in such a way that the amplification of the tube will change in accordance with the amplitude of the audio signal. It can be accomplished in either the plate or control grid circuit, or by means of other electrodes such as screen or supressor grids. With proper adjustment, it is possible to secure 100% modulation with practically linear response.

21. Frequency Systems

The frequency of a wave amplified by a vacuum tube is not affected by changes in electrode voltages. The frequency of the oscillator can be varied in that way, but such systems do not meet the requirements for satisfactory operation because of instability and non-linear response. It is possible to modulate an oscillator by means of mechanical variations of certain circuit constants such as the capacity of the tuning condenser, but all of the mechanical modulation systems have numerous other disadvantages as well as instability.

As a general rule, it is best to modulate an amplifier so that the oscillator frequency can be accurately fixed, and independent of modulation. If frequency modulation is to compete with amplitude systems, the need is apparent for some circuit which will accomplish this with practically linear response and consistently dependable performance in every respect.

22. Armstrong Circuit

The functional diagram for a frequency modulation circuit developed by Armstrong is shown on page 36. In this system, the oscillator is independent of the modulation circuit, and no mechanical operations are employed. The circuit was set up in the laboratory, using the schematic diagram shown on page 38, but the construction and the parts used were not satisfactory and it was not possible to obtain an accurate experimental check of performance. Therefore the circuit will be discussed on a theoretical basis.

Inspection of the diagrams will show that the carrier is fed from the oscillator into two separate

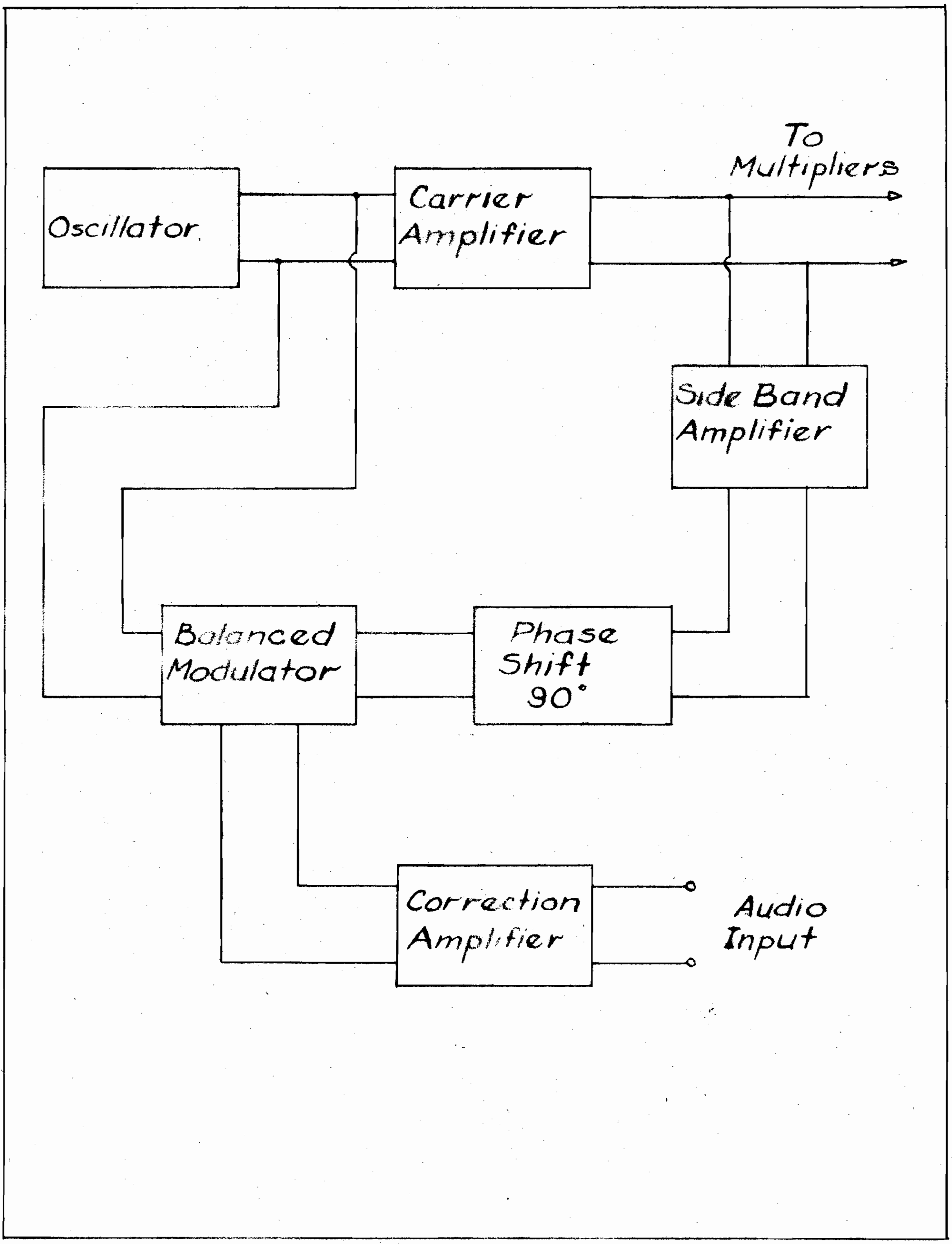

Fig. 5. Armstrong Circuit - Functional Diagram

units: an amplifier, and a balanced modulator. The balanced modulator is a special type of amplitude modulator, in which the carrier is eliminated from the output, and only the side bands are passed on to the next stage. Its operation will not be discussed in detail here, as it has been described in several text books* and periodicals.

The 90 degree phase shift is obtained by the relation between the plate circuit of the balanced modulator and the grid circuit of the side band amplifier. Note that the plate resonant circuit is a series arrangement shunted across a non-inductive resistance. At resonance, the tuned circuit will have zero reactance, so that most of the a.c. component of the plate current will flow through the coil and will be in phase with the modulator input. The plate coils are inductively coupled to the amplifier grid coil, which has high inductive reactance and negligible resistance. It follows that the voltage on the grid of the amplifier will lag 90 degrees behind the input to the balanced modulator.

At this stage, the modulator has produced a pair of side bands by means of amplitude modulation, balancing

---

* See Glasgow, R. S., "Principles of Radio Engineering" Sec. 135, page 331. For application to carrier current systems, see: Affel, Demarest, and Green, "Carrier Systems on Long Distance Telephone Lines", Bell System Technical Journal, July 1928, page 564.

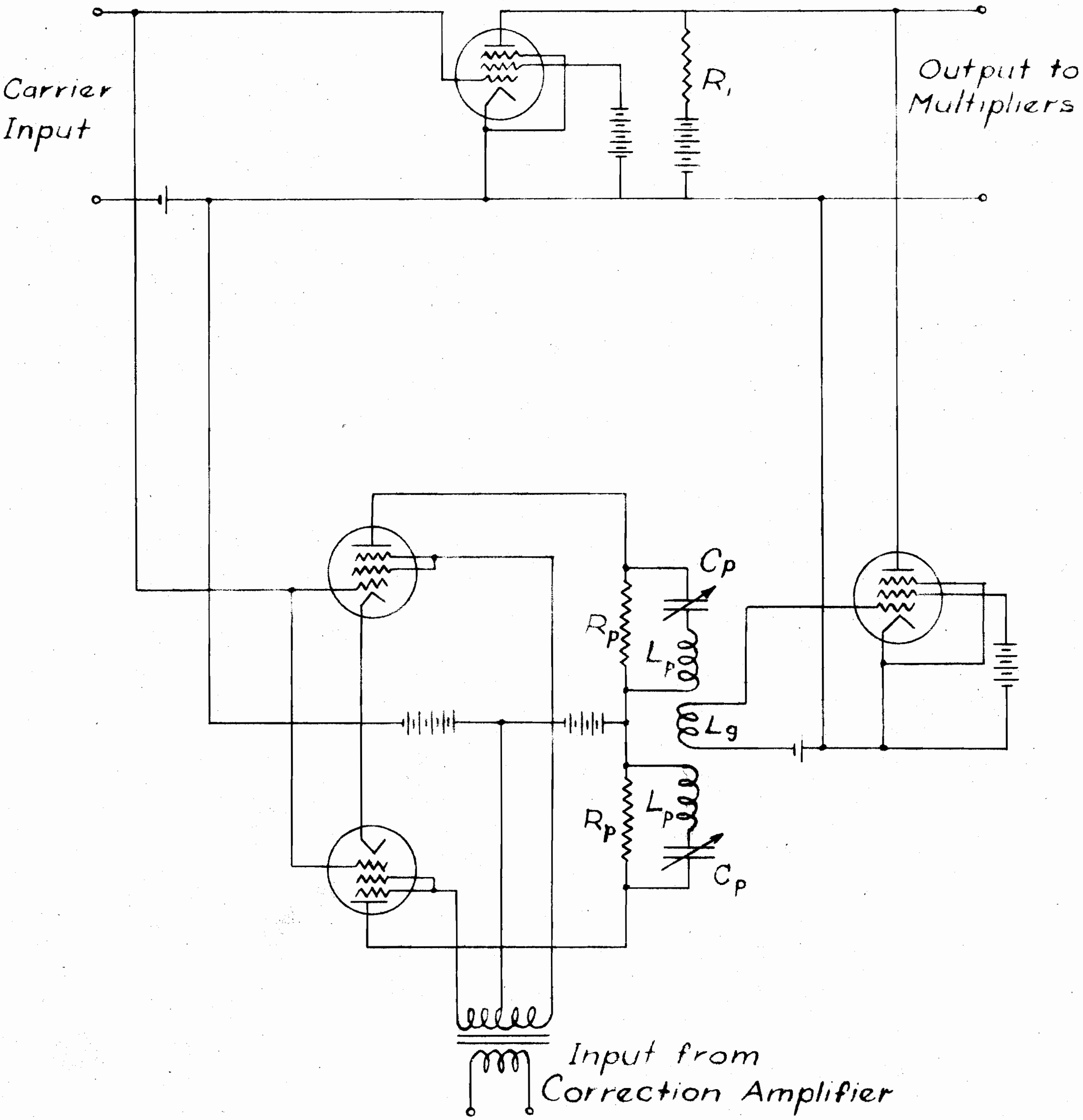

Fig. 6. Armstrong Circuit - Schematic Diagram

out the carrier. The side bands have been shifted 90 degrees from their original phase relation to the carrier, and amplified as required; the carrier has been amplified separately. Now the two current waves, carrier and the shifted side bands, are mixed by means of a common plate resistance load, so that the voltage on the grid of the next amplifier will be the sum of the two waves.

23. Current Equations

Having observed the operation of the various units, it is now possible to write the equations for the currents. In the balanced modulator stage, side bands are obtained from amplitude modulation; from equation (8) page 13, the vector form can be written as:

$$i = A_o m_a \sin ut \sin wt$$

where m_a is the amplitude modulation factor
u represents modulating frequency
A_o is the carrier amplitude

Shifting phase 90 degrees from the carrier, it becomes:

$$i = A_o m_a \sin ut \cos wt$$

Adding this current to the carrier current, write:

$$i = A(\sin wt \neq m_a \sin ut \cos wt)$$

where $A = A_o$

The preceeding equation corresponds in form to the first two terms of equation (20) page 19. The value of m_a will approximate the ratio $2J_1(m)/J_0(m)$. A phase shift will be produced, but there will be no amplitude modulation.

24. Modulation Curve

It is now possible to plot a modulation curve for the Armstrong circuit in terms of phase shift, assuming the same conditions as for the purely theoretical curve presented on page 27. In this case, the side band amplitude is directly proportional to the audio amplitude, so the curve is a tangent function. Note that the phase shift relation is substantially linear for angles of less than about 30 degrees. The curve (page 41) reveals that results obtained with the Armstrong circuit are very similar to the theoretical results discussed previously; in fact, the circuit gives a better approximation of linear frequency modulation than the previous condition with one pair of side bands.

25. Correction Amplifier

It has been shown that, in order to make the frequency deviation independent of the frequency of the audio current, the modulation factor must vary inversely with

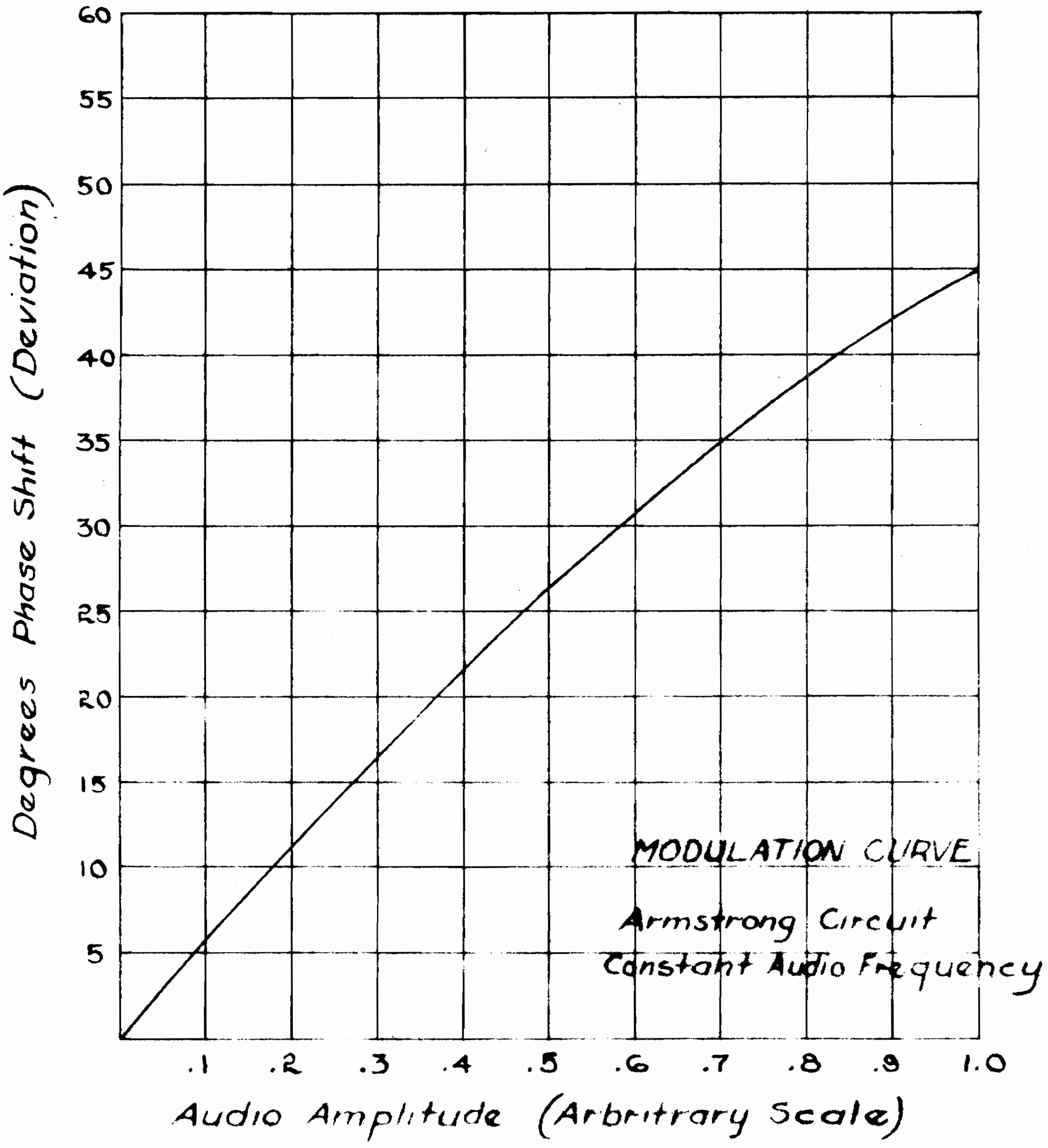

Fig. 7. Modulation Curve - Armstrong Circuit

the audio frequency. This requirement is fulfilled by means of a correction amplifier, which is designed so that the amplification will be inversely proportional to the audio frequency of the input.

26. Laboratory Difficulties

The laboratory apparatus developed a great many defects, and did not produce enough frequency modulation to be apparent on the oscillograph. For the benefit of future workers, some of the defects and some suggestions for improvement are listed.

(1). Type 6K7 tubes were used, because they were the only triple-grid tubes available in sufficient quantity. Their high gain, low power, and remote cut-off characteristics were unsatisfactory. Adjustments were critical, and the tubes introduced distortion before sufficient output could be obtained. Larger tubes with lower gain should be used, and it might be advisable to add several more amplifier stages to provide complete control of the various amplitudes for laboratory purposes.

(2). The properties of the plate and grid inductance coils were not adequate to provide the phase shift required. The plate coils were wound with a large number of turns on low quality insulation material, with the assumption that losses would be small at 50 kilocycles. However, it was practically impossible to obtain a resonant point in the modulator plate circuit, probably because of high losses in the coils.

(3). Numerous "stray" effects were encountered, because of coupling between various circuit elements. In order to avoid such effects, the units should be carefully isolated from each other. Construction should be rigid, and high frequency technique should be followed to avoid many unpredictable factors.

VI SUMMARY AND CONCLUSIONS

27. The Question

The principal question that has been raised in connection with this subject concerns the value of frequency modulation. Armstrong has conclusively proved that such a system can greatly reduce noise disturbances; his work and results can hardly be challenged on that basis. He has demonstrated, in actual operation, a system of high fidelity transmission and reception that has reduced noise disturbances to an almost negligible minimum. The important question may be stated in this way:

Can frequency modulation ever replace amplitude systems in commercial application ?

Naturally, the question of commercial use is vital in considering the value of frequency modulation. Unless it can compete with present systems on equal terms and prove some margin of superiority, any new system will have little value other than academic interest.

28. Advantages

The only concrete advantage that frequency modulation has to offer is noise reduction; all others are derived from it. It is, however, a tremendous advantage, and one that cannot be ignored. It affects every phase of the communication problem. As a result of noise reduction, effective communication is possible with less power. The entertainment value of broadcast programs would be increased immeasurably. In many respects, the system seems to be an answer to the favorite dream of radio workers, both amateur and professional.

29. Disadvantages

Before rushing too fast toward a seeming Utopia, it is surely advisable to consider whether the gain is worth the cost of the trip. The difficulties are numerous, so perhaps it is best to list some of them individually:

(1). Design and construction of both transmitters and receivers would become much more complicated than they now are. Compromises in theory are involved, as well as compromises in actual practice.

(2). Operation and maintenance also become more difficult, though in a lesser degree than design.

(3). The service range of the transmitters is limited because of the need for using frequencies so high that they are not reflected back to the earth in any consistent manner. Unless new possibilities are uncovered, rural areas seem to be necessarily excluded from service by frequency modulated transmitters.

(4). A defect of major proportions lies in the possibility of increased cost. The reduction in power requirements may be more than offset by the need for more equipment (Remember the 500 times multiplication of the fundamental). The economic problem is hard to solve, but there are many factors which tend to increase cost.

(5). An almost crushing blow to frequency modulation comes from the simple fact that amplitude systems are now in universal use. Large investments in equipment, training of personnel, methods of manufacture, operation procedure, etc., will not be discarded without a struggle.

30. The Answer - Maybe

Prophecy is a risky business, but at this stage of the game it does not seem likely that frequency modulation will ever gain anything like a competitive status with amplitude systems. This should not be construed as a futile attempt to discredit Armstrong's work; the writer has neither the ability nor desire to make such an attempt. The point is simply this: in the writer's opinion, the advantages of the new system probably do not have as much economic importance as the advantages of the present system; and if there is any margin of superiority, it is too slight to force a change in the status quo. Prospects of large financial gains would be necessary to overthrow the established methods, and those gains are not now evident; it seems improbable that the system could be operated even at the same cost.

Suppose, now, that frequency modulation had been developed first, and had been in common use for the same length of time as present methods. If amplitude modulation were now in the experimental stage, could the frequency system resist the attack of the simple, cheap "5-and-10" newcomer ?

Since frequency modulation cannot utilize the present broadcast frequency range, and since it may be unable to provide rural service, it probably could never replace amplitude modulation for that type of service. If amplitude systems must maintain rural service, then the frequency system would be restricted to metropolitan coverage. Could the two systems exist and operate separately at the same time ? Problems of receiver utility, interchangeability of parts, service information, etc., would all be multiplied. At the same time, there would be growing difficulties concerning broadcast station coverage and program distribution.

Another factor that should be mentioned is the question of band width. Of course, a channel 150 kilocycles wide is only a "drop in the bucket" within the very high range of 50 megacycles and higher. However, television is now struggling for recognition and promises that some day it may occupy large slices of the high frequencies. It would be very unfortunate to adopt a transmission system that might eventually limit the availability of the television service because of interference with speech transmission.

As an incidental side-light, the writer would like to see television transmitted with frequency modulation. No doubt it is possible, but from this angle it looks like quite a problem.

Very often new developments have been declared impractical or even impossible but proved valuable in actual use later on. As a result some people now think that any such statement is untrue. That does not necessarily follow, however, and should never be assumed.

31. New Methods

An interesting feature of Armstrong's work is his proof that wider transmission channels can be used to reduce noise disturbances. Therefore it seems probable that new methods can be developed to utilize wider channels for reducing noise, and at the same time avoiding some of the undesirable features of frequency modulation. It should not be necessary to make the channels quite as wide as 150 kilocycles, and it should be possible to use intermediate as well as very high frequencies.

The writer would like to propose a double channel amplitude system. Suppose all transmission channels were increased to 40 kilocycles; it would then be possible to operate two high fidelity amplitude modulated transmitters at separate frequencies within each channel. Receivers could be tuned to either transmitter, or to both by use of separate tuned circuits. Suppose one of the transmitters is designed so that its wave lags in phase 180 degrees behind the other. If the receiver is made to feed both signals into a detector with the same phase relation, no audio signal will be reproduced; but if one of the r.f. amplifiers causes a 180 degree phase shift so that the two signals are back in phase the audio signal will be heard. The point is, of course, that the noise amplitudes should then be 180 degrees out of phase and cancel each other.

The system proposed has obvious drawbacks for actual use, and conceivably it would not work at all.

However, it seems to offer possibilities, and the writer regrets lack of time and facilities to do more work on the subject.

In addition to the system mentioned, there are many ways that amplitude modulated signals could be combined by using broad channels. Perhaps there would be some value in using an inverted or "scrambled" signal. At any rate, the modulation and transmission problem offers many features of academic interest, and possibilities of development for commercial application of some system better than anything now in use.

APPENDIX I

To expand the expression:

$$\sin(x \sin r)$$

First expand into a power series,

$$\sin(x \sin r) = x \sin r - \frac{(x \sin r)^3}{3!} + \frac{(x \sin r)^5}{5!}$$

$$- \frac{(x \sin r)^7}{7!} + \dots\dots\dots$$

$$= x \sin r - \frac{x^3}{3!} \sin^3 r + \frac{x^5}{5!} \sin^5 r$$

$$- \frac{x^7}{7!} \sin^7 r + \dots\dots\dots$$

Each power term of the sine can be evaluated in terms of multiples of the angle, using trigonometric relations:

$$\sin^3 r = \sin r \sin^2 r = \sin r \left[\tfrac{1}{2}(1 - \cos 2r)\right]$$

$$= \tfrac{1}{2}\sin r - \tfrac{1}{2} \sin r \cos 2r$$

$$= \tfrac{1}{2}\sin r - \tfrac{1}{2}\left[\tfrac{1}{2}\sin 3r - \tfrac{1}{2}\sin r\right]$$

$$= \tfrac{1}{2}\sin r - \tfrac{1}{4}\sin 3r + \tfrac{1}{4}\sin r$$

$$= \tfrac{3}{4}\sin r - \tfrac{1}{4}\sin 3r$$

$$\sin^5 r = \sin^3 r \sin^2 r = (\tfrac{3}{4}\sin r - \tfrac{1}{4}\sin 3r)\sin^2 r$$

$$= \tfrac{3}{4}\sin^3 r - \tfrac{1}{4}\sin 3r\left[\tfrac{1}{2}(1 - \cos 2r)\right] \qquad \text{etc...}$$

$$= 10/16 \sin r - 5/16 \sin 3r + 1/16 \sin 5r$$

The expansion can be continued as long as desired to determine the series which will result. Then substitute for the power terms, and collect coefficients for the sine terms, each of which coefficients will be a series:

$$\begin{aligned} \sin(x \sin r) = {} & \sin r\left[x - \frac{x^3}{3!}\,\frac{3}{4} + \frac{x^5}{5!}\,\frac{10}{16} - \ldots\ldots\right] \\ & + \sin 3r\left[\ldots\ldots\ldots\ldots\right] \\ & + \ldots\ldots\ldots\ldots \text{ etc.} \end{aligned}$$

Each coefficient can be written as a Bessel function by making the proper changes in form to show the relation. For example, the sin r term is changed as follows:

$$\begin{aligned} & x - \frac{x^3}{3!}\,\frac{3}{4} + \frac{x^5}{5!}\,\frac{10}{16} - \ldots\ldots \\ & \qquad = 2\left[\frac{x}{2} - \frac{x^3}{3!}\,\frac{3}{8} + \frac{x^5}{5!}\,\frac{10}{32} - \ldots\ldots\ldots\right] \\ & \qquad = 2\left[\frac{x}{2} - \frac{x^3}{2!2^3} + \frac{x^5}{2!3!2^5} - \ldots\ldots\ldots\right] \end{aligned}$$

By the definition of Bessel functions, the above expression is equal to twice the Bessel function, of the first kind and first order, for the argument x. That is,

$2\,J_1(x)$ where J denotes the Bessel function of the first kind for x.

By continuing the expansion in the same manner to obtain any desired number of terms, the coefficient for each term containing a sine expression (sin nx) can be written as a Bessel function of x. The series finally becomes:

$$\sin(x \sin r) = 2J_1(x)\sin r + 2J_3(x)\sin 3r + 2J_5(x)\sin 5r + \cdots$$

The expression,

$$\cos(x \sin r)$$

can be evaluated in the same manner.

$$\cos(x \sin r) = J_0(x) + 2\,J_2(x)\cos 2r + 2\,J_4(x)\cos 4r + \cdots$$

APPENDIX II

Reference Bibliography

BOOKS

Burington, R. S. Handbook of Mathematical Tables and Formulas.
Handbook Publishers, Inc., Sandusky, Ohio. 1933

Byerly, W. E. Fourier's Series and Spherical Harmonics
Ginn and Company, Boston. 1893

Glasgow, R. S. Principles of Radio Engineering
McGraw-Hill Book Company. New York. 1936

Jahnke, Eugen, and Fritz Emde. Funktionentafeln
B. G. Teubner. Leipzig. 1933.

Sokolnikoff, I. S. and E. S. Higher Mathematics for Engineers and Physicists
McGraw-Hill Book Company. New York. 1934

PERIODICALS

Affel, Demarest, and Green. "Carrier Systems on Long Distance Telephone Lines". _Bell System Technical Journal_, July 1928, page 564.

Armstrong, Edwin H. "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation". _Proceedings of the Institute of Radio Engineers_, May 1936, page 689.

Roder, Hans. "Amplitude, Phase, and Frequency Modulation". _Proceedings of the Institute of Radio Engineers_, December 1931, page 2145.