FREQUENCY MODULATION

Volume I

FREQUENCY MODULATION

Volume I

Edited by

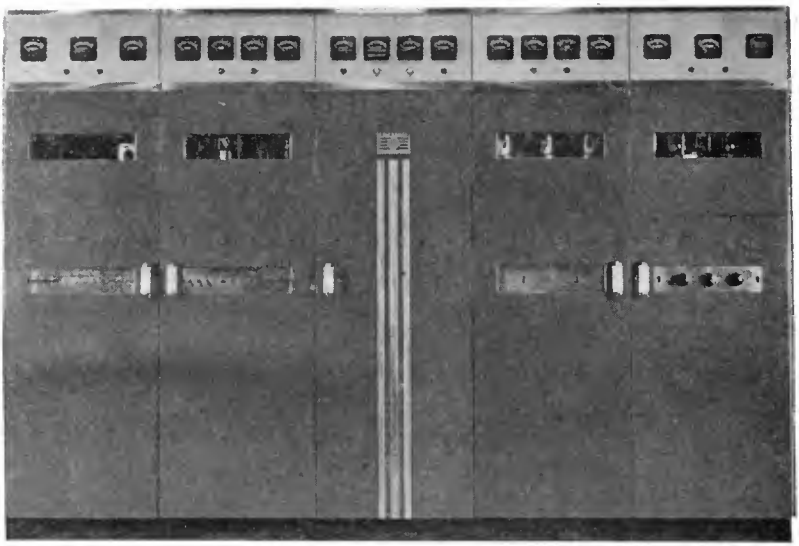
ALFRED N. GOLDSMITH
ARTHUR F. VAN DYCK
ROBERT S. BURNAP
EDWARD T. DICKEY
GEORGE M. K. BAKER

JANUARY, 1948

Published by

RCA REVIEW
RADIO CORPORATION OF AMERICA
RCA LABORATORIES DIVISION
Princeton, New Jersey

Printed in U.S.A.



1000-3000 watt FM transmitter—1947

Table-model AM-FM receiver—1947

FREQUENCY MODULATION, Vol. I

PREFACE

FREQUENCY MODULATION, Volume I is the seventh volume in the RCA Technical Book Series and the first on the subject of FM. Previous volumes have included material on television, radio facsimile, and ultra-high-frequency radio.

* * * * *

The papers in this volume cover the period 1936-1947 and are presented in four sections: general, transmission, reception, and miscellaneous. Each section includes papers reprinted in full and several reproduced in summary form only, with appropriate original publication data indicated in each case. As additional sources of reference, Appendix I includes a frequency modulation bibliography and Appendix II lists papers dealing with frequency modulation station placement and field survey techniques.

* * * * *

RCA REVIEW gratefully acknowledges the courtesy of the Institute of Radio Engineers *(Proc. I.R.E.)*, the McGraw-Hill Publishing Company, Inc. *(Electronics)*, the Society of Motion Picture Engineers *(Jour. Soc. Mot. Pic. Eng.)*, the Bryan-Davis Publishing Company *(Communications)*, and *FM and Television* in granting to RCA REVIEW permission to republish material which has appeared in their publications. The appreciation of RCA REVIEW is also extended to all authors whose papers appear herein.

* * * * *

FREQUENCY MODULATION, Volume I is being published for scientists, engineers and others interested in frequency modulation, with the sincere hope that the material here assembled may help to speed developments and enable frequency modulation to find its true place among companion arts and services.

RCA Laboratories,
Princeton, N. J.
December 15, 1947

The Manager, RCA REVIEW

FREQUENCY MODULATION

Volume I

CONTENTS

PAGE

FRONTISPIECE ..... —v—
PREFACE ..... *The Manager*, RCA REVIEW —vii—

GENERAL

Frequency Modulation Noise Characteristics ..... M. G. CROSBY 1
Frequency Modulation ..... S. W. SEELEY 44
Band Width and Readability in FM ..... M. G. CROSBY 57
Variation of Bandwidth with Modulation Index in Frequency Modulation ..... M. S. CORRINGTON 65

Summary:
Generalized Theory of Multitone Amplitude and Frequency Modulation ..... L. J. GIACOLETTO 84

TRANSMISSION

Frequency Modulation Propagation Characteristics ..... M. G. CROSBY 85
A Cathode-Ray Frequency Modulation Generator ..... R. E. SHELBY 101
NBC Frequency-Modulation Field Test ... R. F. GUY and R. M. MORRIS 111
Generation and Detection of Frequency-Modulated Waves ..... S. W. SEELEY, C. N. KIMBALL and A. A. BARCO 147
A New Exciter Unit for Frequency-Modulated Transmitters ..... N. J. OMAN 165
A Pretuned Turnstile Antenna ..... G. H. BROWN and J. EPSTEIN 178
Characteristics of the Pylon FM Antenna ..... R. F. HOLTZ 194

Summaries:
Reactance-Tube Frequency Modulators ..... M. G. CROSBY 202
A Transmitter for Frequency Modulated Broadcast Service Using a New Ultra-High-Frequency Tetrode. A. K. WING and J. E. YOUNG 202
Drift Analysis of the Crosby Frequency-Modulated Transmitter Circuit ..... E. S. WINLUND 203
Antennas for F-M Stations ..... J. P. TAYLOR 203
F-M Audio Measurements with an A-M Receiver ..... R. J. NEWMAN 204
A Square Loop F-M Antenna ..... J. P. TAYLOR 204
Audio Frequency Response and Distortion Measuring Techniques for F-M Transmitting Systems ..... R. J. NEWMAN 205
F-M Antenna Coupler ..... J. P. TAYLOR 205
Isolation Methods for F-M Antennas Mounted on A-M Towers ..... R. F. HOLTZ 206
Slot Antennas ..... N. E. LINDENBLAD 206

RECEPTION

The Service Range of Frequency Modulation ..... M. G. CROSBY 207
Impulse Noise in F-M Reception ..... V. D. LANDON 230

PAGE

Intermediate-Frequency Values for Frequency-Modulated-Wave Receivers ..... D. E. FOSTER and J. A. RANKIN 245
A Frequency-Dividing Locked-in Oscillator Frequency-Modulation Receiver ..... G. L. BEERS 257
Frequency-Modulation Distortion Caused by Multipath Transmission ..... M. S. CORRINGTON 275
Input Impedance of Several Receiving-Type Pentodes at FM and Television Frequencies ..... F. MURAL 305
Frequency Modulation Distortion Caused by Common- and Adjacent-Channel Interference ..... M. S. CORRINGTON 319
The Ratio Detector ..... S. W. SEELEY and J. AVINS 358

Summaries:

The Development of a Frequency-Modulated Police Receiver for Ultra-High-Frequency Use ..... H. E. THOMAS 394
Development of an Ultra Low Loss Transmission Line for Television ..... E. O. JOHNSON 394
Input Circuit Noise Calculations for F-M and Television Receivers ..... W. J. STOLZE 395

MISCELLANEOUS

Duplex Transmission of Frequency-Modulated Sound and Facsimile ..... M. ARTZT and D. E. FOSTER 396
Use of Subcarrier Frequency Modulation in Communication Systems ..... W. H. BLISS 410
The Transmission of a Frequency-Modulated Wave Through a Network ..... W. J. FRANTZ 419
Push-Pull Frequency Modulated Circuit and Its Application to Vibratory Systems ..... A. BADMAIEFF 452
Frequency Modulation and Control by Electron Beams ..... L. P. SMITH and C. I. SHULMAN 467

Summaries:

Carrier and Side-Frequency Relations with Multi-Tone Frequency or Phase Modulation ..... M. G. CROSBY 502
Some Recent Developments in Record Reproducing Systems ..... G. L. BEERS and C. M. SINNETT 502
Frequency Modulation of Resistance-Capacitance Oscillators ..... M. ARTZT 503
An F-M Calibrator for Disc Recording Heads ..... H. E. ROYS 503
Frequency-Modulation Mobile Radiotelephone Services ..... H. B. MARTIN 504
A Frequency-Modulated Magnetron for Super-High-Frequencies ..... G. R. KILGORE, C. I. SHULMAN, and J. KURSHAN 504
A 1-Kilowatt Frequency-Modulated Magnetron for 900 Megacycles ..... J. S. DONAL, JR., R. R. BUSH, C. L. CUCCIA and H. G. HEGBAR 505

APPENDIX I — FREQUENCY MODULATION — A Bibliography of Technical Papers by RCA Authors, 1936-1947 507
APPENDIX II—FM Station Placement and Field Survey Techniques—Summaries 513

FREQUENCY MODULATION NOISE CHARACTERISTICS*†

By

Murray G. Crosby

RCA Communications, Inc., Riverhead, L. I., New York

Summary.—*Theory and experimental data are given which show the improvements in signal-noise ratio effected by frequency modulation over amplitude modulation. It is shown that above a certain carrier-noise ratio in the frequency modulation receiver which is called the "improvement threshold," the frequency modulation signal-noise ratio is greater than the amplitude modulation signal-noise ratio by a factor equal to the product of a constant and the deviation ratio (the deviation ratio is equal to the ratio between the maximum frequency deviation and the audio modulation band width). The constant depends upon the type of noise, being slightly greater for impulse than for fluctuation noise. In frequency modulation systems with high deviation ratios, a higher carrier level is required to reach the improvement threshold than is required in systems with low deviation ratios; this carrier level is higher for impulse than for fluctuation noise. At carrier-noise ratios below the improvement threshold, the peak signal-noise ratio characteristics of the frequency modulation receiver are approximately the same as those of the amplitude modulation receiver, but the energy content of the frequency modulation noise is reduced.*

An effect which is called "frequency limiting" is pointed out in which the peak value of the noise is limited to a value not greater than the peak value of the signal. With impulse noise this phenomenon effects a noise suppression in a manner similar to that in the recent circuits for reducing impulse noise which is stronger than the carrier in amplitude modulation reception.

When the power gain obtainable in certain types of transmitters by the use of frequency modulation is taken into account, the frequency modulation improvement factors are increased and the improvement threshold is lowered with respect to the carrier-noise ratio existing in a reference amplitude modulation system.

Introduction

IN A previously published paper,[1] the propagation characteristics of frequency modulation were considered. Prior to, and during these propagation tests, signal-noise ratio improvements effected by frequency modulation were observed. These observations were made at an early stage of the development work and were investigated by experimental and theoretical methods.

It is the purpose of this paper to consider that phase of the frequency modulation development work by RCA Communications,

* Decimal Classification: R630.3 × R270.

† Reprinted from *Proc. I.R.E.*, April, 1937.

[1] Murray G. Crosby, "Frequency Modulation Propagation Characteristics," Proc. I.R.E., vol. 24, pp. 898–913; June, (1936).

Inc., in which the signal-noise characteristics of frequency modulation are studied. The theory and experimental work consider the known systems of frequency modulation including that independently developed by E. H Armstrong.[2]

Table of Symbols

C = carrier peak voltage.

C/N = theory: Ratio between peak voltage of carrier and instantaneous peak voltage of the noise in the frequency modulation receiver. Experiment: Ratio between peak voltage of carrier and *maximum* instantaneous peak voltage of the noise

C/n = ratio between the peak voltage of the carrier and the peak voltage of the noise component.

C_a/N_a = carrier-noise ratio in the amplitude modulation receiver.

F_a = maximum audio frequency of modulation band.

F_c = carrier frequency

F_d = peak frequency deviation due to applied modulation.

F_{dn} = peak frequency deviation of the noise.

F_1 = intermediate-frequency channel width.

F_m = modulation frequency

F_n = frequency of noise resultant or component.

F_d/F_a = deviation ratio.

K = slope filter conversion efficiency.

M = modulation factor of the amplitude modulated carrier.

M_f = modulation factor at the output of the sloping filter.

M_{fn} = modulation factor at the output of the sloping filter when noise modulates the carrier.

N = instantaneous peak voltage of the noise.

n = peak voltage of the noise component.

N_a = noise peak or root-mean-square voltage at amplitude modulation receiver output.

N_f = noise peak or root-mean-square voltage at frequency modulation receiver output.

$p = 2\pi F_m$.

S_a = signal peak or root-mean-square voltage at amplitude modulation receiver output.

S_f = signal peak or root-mean-square voltage at frequency modulation receiver ouput.

$\omega = 2\pi F_c$.

[2] Edwin H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation," Proc. I.R.E., vol. 24, pp. 689–740; May, (1936).

$\omega_n = 2\pi F_n$.
$\omega_{na} = (\omega - \omega_n) = 2\pi(F_c - F_n) = 2\pi F_{na}$.
$Z = C/n + n/C$.
$\phi(t)$ = phase variation of noise resultant as a function of time.

Theory

In the following analysis, frequency modulation is studied by comparing it with the familiar system of amplitude modulation. In order to do this, the characteristics of frequency modulation reception are analyzed so as to make possible the calculation of the signal-noise ratio improvement effected by frequency modulation over amplitude modulation at various carrier-noise ratios.[3] The amplitude modulation standard of comparison consists of a double side-band system having the same audio modulation band as the frequency modulation system and producing the same carrier at the receiver. Differences in transmitter power gain due to frequency modulation are then considered separately. The frequency modulation reception process is analyzed by first considering the components of the receiver and the manner in which they convert the frequency modulated signal and noise spectrum into an output signal-noise ratio.

The Frequency Modulation Receiver

The customary circuit arrangement used for the reception of frequency modulation is shown in the block diagram of Fig. 1. The intermediate-frequency output of a superheterodyne receiver is fed through a limiter to a slope filter or conversion circuit which converts

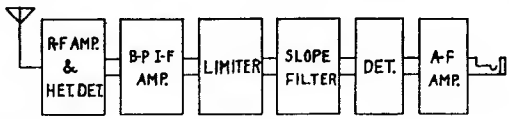


Fig. 1—Block diagram of a frequency modulation receiver.

the frequency modulation into amplitude modulation. This amplitude modulation is then detected in the conventional amplitude modulation manner. The audio-frequency amplifier is designed to amplify only the modulation frequencies; hence it acts as a low-pass filter which rejects noise frequencies higher than the maximum modulation frequency.

[3] Throughout this paper, carrier-noise ratio will refer to the ratio measured at the output of the intermediate-frequency channel. Signal-noise ratio will refer to that measured at the output of the receiver and will depend upon the depth of modulation as well as upon the carrier strength.

The purpose of the limiter is to remove unwanted amplitude modulation so that only the frequency modulation component of the signal will be received. It may take the form of an overloaded amplifier tube whose output cannot rise above a certain level regardless of the input. Care must also be exercised to insure that the output of the overloaded amplifier does not fall off as the input is increased since this would introduce amplitude modulation of reverse phase, but of equally undesirable character.

The main requirement of the conversion circuit for converting the frequency modulation into amplitude modulation is that it slope linearly from a low value of output at one side of the intermediate-frequency channel to a high value at the other side of the channel. To do this, an off-tuned resonant circuit or a portion of the characteristic of

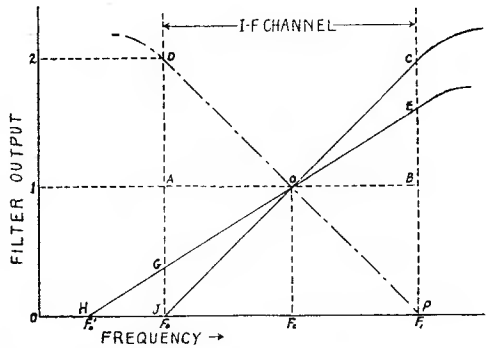


Fig. 2—Ideal sloping filter characteristics.

one of the many forms of wave filters may be utilized. The ideal slope filter would be one which gave zero output at one side of the channel, an output of one voltage unit at carrier frequency, and an output of two units at the other side of the channel. Such a characteristic is given by the curve *JOC* of Fig. 2. From this curve it is easily seen that if the frequency is swung between the limits F_0 and F_1, about the mean frequency F_c, the output of the filter will have an amplitude modulation factor of unity. The modulation factor for a frequency deviation, F_d, will be given by

$$M_f = \frac{F_d}{(F_1 - F_c)} = \frac{F_d}{(F_c - F_0)} = \frac{2F_d}{F_i} \qquad (1)$$

where F_i = intermediate-frequency channel width.

When the converting filter departs from the ideal characteristic in the manner of the filter of curve *HGOE* of Fig. 2, the modulation factor produced by a given frequency deviation is reduced by a factor equal

to the ratio between the distances AG and AJ or BE and BC. A convenient term for this reduction factor of the filter is "conversion efficiency" of the filter. Taking into account this conversion efficiency, the modulation factor for a frequency deviation F_d becomes

$$M_f = \frac{2KF_d}{F_i} \tag{2}$$

where $K = AG/AJ = BE/BC =$ conversion efficiency of the filter.

A low conversion efficiency may be used as long as the degree of limiting is high enough to reduce the amplitude modulation well below the level of the converted frequency modulation. This is true since lowering the conversion efficiency reduces the output of the noise in the same proportion as the signal as long as no amplitude modulation is present. Hence the signal-noise *ratio* is unimpaired and the only effect is a reduction of the audio gain by the factor K. If insufficient limiting is applied so that the output of the limiter contains appreciable amplitude modulation, a filter with a high conversion efficiency is desirable so that the amplitude modulation noise will not become comparable to the frequency modulation noise and thereby increase the resultant noise.

A push-pull, or "back-to-back" receiver may be arranged by providing two filters of opposite slope and separately detecting and combining the detected outputs in push-pull so as to combine the audio outputs in phase. Another slope filter having a characteristic as shown by the dot-dash line DOP in Fig. 2 would then be required.

A further type of receiver in which amplitude modulation is also balanced out may be arranged by making one of the slope filter circuits of the above-mentioned back-to-back type of receiver a flat-top circuit for the detection of amplitude modulation only. The sloping filter channel then detects both frequency and amplitude modulation; the flat-top channel detects only amplitude modulation. When these two detected outputs are combined in push-pull, the amplitude modulation is balanced out and the frequency modulation is received. This type of detection, as well as that in which opposite slope filters are used, has the limitation that the balance is partially destroyed as modulation is applied. However, if a limiter is used, the amplitude modulation is sufficiently reduced before the energy reaches the slope filters; consequently, for purposes of removing amplitude modulation, the balancing feature is not a necessity.

Noise Spectrum Analysis

The first step in the procedure to be followed here in determining the noise characteristics of the frequency modulation receiver will be

to determine mathematically the fidelity with which the noise is transmitted from the radio-frequency branch, in which it originates, to the measuring instrument as a function of frequency. To do this, the waves present at the receiver input will be assumed to be the frequency modulated carrier and the spectrum of noise voltages. This wave and spectrum will be combined into a single resultant whose amplitude and phase are functions of the constants of the component waves. The resultant will then be "mathematically" passed through the limiter to remove the amplitude modulation. From a determination of the instantaneous frequency of the resultant, the peak frequency deviation effected by the noise will be found. A single noise component of arbitrary frequency will then be substituted for the resultant of the noise spectrum, and the modulation factor at the output of the converting filter will be obtained. This noise component will then be varied in frequency to determine the over-all transmission of the receiver in terms of the modulation factor at the sloping filter output. The area under the curve representing the square of this over-all transmission will then be determined. By comparing this area with the corresponding area for an amplitude modulation receiver under equivalent conditions, and taking into consideration the pass band of the intermediate- and audio-frequency channels, a comparison will be obtained between the average noise powers, or the average root-mean-square noise voltages from the two receivers.[4]

The peak voltage characteristics of the two receivers will be compared for fluctuation noise by a correlation of known crest factors with the root-mean-square characteristics. (Crest factor = ratio between the peak and root-mean-square voltages.) The peak voltage characteristics of impulse noise will be determined by a separate consideration of the effect of the frequency modulation over-all transmissions upon the peak voltage of this type of noise.

After a comparison between the noise output voltages from the frequency and amplitude modulation receivers has been obtained, the respective signal output voltages will be taken into consideration so that the improvement in signal-noise ratio may be determined.

In the process of determining the over-all transmission of the noise, the frequency modulated wave may be expressed by

$$e_s = C \sin \{\omega t + (F_d/F_m) \cos pt\} \tag{3}$$

[4] Stuart Ballantine, "Fluctuation noise in radio receivers," PROC. I.R.E., vol. 18, pp. 1377–1387; August, (1930). In this paper, Ballantine shows that the average value of the square of the noise voltage " . . . is proportional to the area under the curve representing the square of the over-all transimpedance (or of the transmission) from the radio-frequency branch in which the disturbance originates to the measuring instrument as a function of frequency . . . . "

where C = carrier peak voltage, $\omega = 2\pi F_c$, F_c = carrier frequency, F_d = applied frequency deviation, $p = 2\pi F_m$, and F_m = modulation frequency. The noise spectrum may be expressed by its resultant,[5]

$$e_n = N \sin (\omega_n t + \phi(t)) \tag{4}$$

where N = instantaneous peak voltage of the noise (a function of time). $\phi(t)$ takes into account the fact that the noise resultant is phase modulated, as would be the case with the resultant of a spectrum of many noise voltages. $\omega_n = 2\pi F_n$, F_n = frequency of the noise resultant.

The signal voltage given by (3) and the noise voltage given by (4) may be combined by vector addition to give

$$e = \sqrt{C^2+N^2+2CN\cos\left\{(\omega-\omega_n)t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}$$

$$\sin\left[\omega t+\frac{F_d}{F_m}\cos pt+\tan^{-1}\frac{\sin\left\{(\omega-\omega_n)t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}{\frac{C}{N}+\cos\left\{(\omega-\omega_n)t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}\right]. \tag{5}$$

When the resultant wave given by (5) is passed through the limiter in the frequency modulation receiver, the amplitude modulation is removed. Hence the amplitude term is reduced to a constant and the only part of consequence is the phase angle of the wave. The rate of change of this phase angle, or its first derivative, is the instantaneous frequency of the wave. Taking the first derivative and dividing by 2π to change from radians per second to cycles per second gives

$$d\left[\omega t+\frac{F_d}{F_m}\cos pt+\tan^{-1}\frac{\sin\left\{\omega_{na}t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}{\frac{C}{N}+\cos\left\{\omega_{na}t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}\right]\Big/dt\times\frac{1}{2\pi}$$

$$=f=F_c-F_d\sin pt-\frac{\left(F_{na}-\frac{1}{2\pi}\frac{d\phi(t)}{dt}-F_d\sin pt\right)}{\frac{C}{N}+\cos\left\{\omega_{na}t-\phi(t)+\frac{F_d}{F_m}\cos pt\right\}}\Big/\frac{\frac{N}{C}+\cos\left\{\omega_{na}t-\phi(t)+\frac{F_d}{F_n}\cos pt\right\}}{\frac{N}{C}+\cos\left\{\omega_{na}t-\phi(t)+\frac{F_d}{F_n}\cos pt\right\}}+1 \tag{6}$$

[5] John R. Carson, "The reduction of atmospheric disturbances," PROC. I.R.E., vol. 16, pp. 967–975; July, (1928).

in which $\omega_{na}=(\omega-\omega_n)=2\pi(F_c-F_n)=2\pi F_{na}$.

Equation (6) gives the instantaneous frequency of the resultant wave consisting of the signal wave and the noise resultant voltage. From this equation the signal and noise frequency deviations may be obtained. In order to determine the over-all transmission with respect to the various components in the noise spectrum, a single component of noise, with constant amplitude and variable frequency, will be substituted for the resultant noise voltage given by (4). This makes N equal to n, which is not a function of time, and $\phi(t)$ equal to zero. Making these changes in (6) gives

$$f = F_c - F_d \sin pt - \frac{(F_{na} - F_d \sin pt)}{\dfrac{\dfrac{C}{n} + \cos\left\{\omega_{na}t + \dfrac{F_d}{F_m}\cos pt\right\}}{\dfrac{n}{C} + \cos\left\{\omega_{na}t + \dfrac{F_d}{F_m}\cos pt\right\}} + 1}. \tag{7}$$

The equations for the instantaneous frequency, given by (6) and (7), show the manner in which the noise combines with the incoming carrier to produce a frequency modulation of the carrier. From these equations the frequency deviations of the noise may be determined, and from the frequency deviations the modulation factor at the output of the sloping filter may be found. Hence the over-all transmission may be obtained in terms of the modulation factor at the output of the sloping filter for a given carrier-noise ratio. When the carrier-noise ratio is high, (6) and (7) simplify so that calculations are fairly easy. When the carrier-noise ratio is low, the equations become involved to a degree which discourages quantitative calculations.

High Carrier-Noise Ratios

When C/n is large compared to unity, and the applied modulation on the frequency modulated wave is reduced to zero ($F_d=0$), (7) reduces to

$$f = F_c - \frac{n^2}{C^2} F_{na} - \frac{n}{C} F_{na} \cos \omega_{na}t. \tag{8}$$

From (8) the effective peak frequency deviation of a single noise component of the spectrum is

$$F_{dn} = \frac{nF_{na}}{C}\left(\frac{n}{C} + 1\right). \tag{9}$$

But, since n/C is negligible compared to unity,

$$F_{dn} = \frac{nF_{na}}{C}. \tag{10}$$

When this value of frequency deviation is substituted in (1) to find the modulation factor[6] of the energy at the output of the sloping filter, the following results:

$$M_{fn} = \frac{n}{C} \times \frac{2F_{na}}{F_i}. \tag{11}$$

Equation (11) shows that the modulation factor of the noise is inversely proportional to the carrier-noise ratio and directly proportional to the ratio between the noise audio frequency and one half the intermediate-frequency channel width. When this equation is plotted with the noise audio frequency, F_{na}, as a variable and the modulation factor as the ordinate, the audio spectrum obtained for the detector output is like that of the triangular spectrum OBA in Fig. 3. Such a spectrum would be produced by varying F_n through the range between the upper and lower cutoff frequencies of the intermediate-frequency channel. The noise amplitude would be greatest at a noise audio frequency equal to one half the intermediate-frequency channel width. At this noise audio frequency, the ratio $2F_{na}/F_i$ is equal to unity and the modulation factor becomes equal to n/C. If the detector output is passed through an audio system having a cutoff frequency F_a, the maximum frequency of the audio channel governs the maximum amplitude of the spectrum. The maximum amplitude of the detector output is therefore reduced by the ratio $F_i/2:F_a$. This can be seen by a comparison of the spectrum OBA for the detector output and the spectrum ODH for the audio channel output.

When the amplitude modulation reception process is analyzed with a carrier and noise spectrum present at the receiver input, the modulation factor of the energy fed to the detector is found to be equal to the reciprocal of the carrier-noise ratio for all of the noise frequencies in the spectrum. That is to say, the receiver transmission for amplitude modulation will be constant for all of the frequencies in the spectrum. Normally the upper cutoff frequency of the audio amplifier is equal to one half the intermediate-frequency channel width ($F_a = F_i/2$). Consequently the audio spectrum of the amplitude modulated noise fed to

[6] The ideal filter is used in this case since the use of a filter with a low conversion efficiency would merely require the addition of audio gain to put the frequency modulation receiver on an equivalent basis with the amplitude modulation receiver. The audio gain necessary would be equal to the reciprocal of the conversion efficiency of the sloping filter.

the detector will be the same as that at the receiver output and will be as portrayed by the rectangle *OCEH*.

The spectra of Fig. 3 show the manner in which the frequency modulation receiver produces a greater signal-noise ratio than the amplitude modulation receiver. The noise at the output of the detector of the frequency modulation receiver consists of frequencies which extend out to an audio frequency equal to one half the intermediate-frequency channel width, and the amplitudes of these components are proportional to their audio frequency. Hence in passing through the audio channel the noise is reduced not only in range of frequencies, but also in amplitude. On the other hand, the components of the signal wave are properly disposed to produce detected signal frequencies which fit into the audio channel. In the case of the amplitude modulation receiver, the amplitude of the components at the output of the audio

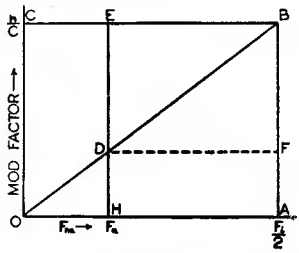


Fig. 3—Amplitude and frequency modulation receiver noise spectra. *OBA* = frequency modulation detector output. *ODH* = frequency modulation receiver output. *OCEH* = amplitude modulation receiver output.

channel is the same as that at the output of the detector since the spectrum is rectangular. Thus the frequency modulation signal-noise ratio is greater than the amplitude modulation signal-noise ratio by a factor which depends upon the relative magnitudes of the spectra *OCEH* and *ODH*. The magnitudes which are of concern are the root-mean-square and peak values of the voltage due to the spectra.

Root-Mean-Square Noise Considerations

The average noise power or average root-mean-square voltage ratio between the rectangular amplitude modulation spectrum *OCEH* and the trangular frequency modulation spectrum *ODH*, of Fig. 3, may be found by a comparison of the squared-ordinate areas of the two spectra. Thus,

$$\frac{W_a}{W_f} = \frac{\text{area } OCEH \text{ (ordinates)}^2}{\text{area } ODH \text{ (ordinates)}^2}$$

$$= \frac{\left(\frac{n}{C}\right)^2 F_a}{\int_0^{F_a} \left(\frac{n}{C} \times \frac{2F_{na}}{F_i}\right)^2 dF_{na}} = 3\left(\frac{F_i}{2F_a}\right)^2 \qquad (12)$$

where W_a/W_f is the ratio between the amplitude modulation average noise power and the frequency modulation average noise power at the receiver outputs. The root-mean-square noise voltage ratio will be

$$\frac{N_a}{N_f} \text{ (r-m-s fluctuation)} = \sqrt{\frac{W_a}{W_f}} = \sqrt{3}\,\frac{F_i}{2F_a}. \qquad (13)$$

Equation (13) gives the root-mean-square noise voltage ratio for equal carriers applied to the two receivers. The modulation factor of the frequency modulated signal due to the applied frequency deviation, F_d, is, from (3), equal to $2F_d/F_i$. The modulation factor of the amplitude modulated signal may be designated by M and has a maximum value of 1.0. Thus the ratio between the two signals will be given by

$$\frac{S_a}{S_f} \text{ (peak or r-m-s values)} = \frac{F_i M}{2F_d} = \frac{F_i}{2F_d} \text{ (for } M = 1.0). \qquad (14)$$

Dividing (13) by (14), to find the ratio between the signal-noise ratios at the outputs of the two receivers, gives

$$\frac{S_f/N_f}{S_a/N_a} \text{ (r-m-s values)} = \sqrt{3}\,\frac{F_d}{F_a}. \qquad (15)$$

It is apparent that the ratio between the frequency deviation and the audio channel, F_d/F_a, is an important factor in determining the signal-noise ratio gain effected by frequency modulation. A convenient term for this ratio is the "deviation ratio" and it will be designated as such hereinafter.

Equation (15) gives the factor by which the amplitude modulation root-mean-square signal-noise ratio is multiplied in order to find the equivalent frequency modulation signal-noise ratio. Since this factor is used so frequently hereinafter, it will be designated by the word "improvement." The improvement given by (15) has been developed under the assumption of zero applied frequency deviation (no modulation) and a carrier which is strong compared to the noise. However, as will be shown later, as long as the carrier is strong compared to the

noise, this equation also holds true for the case where modulation is present.

Peak Noise Considerations

In the ultimate application of signal-noise ratios, peak voltages are of prime importance since it is the peak of the noise voltage which seems to produce the annoyance. This is especially true in the case of impulse noise such as ignition where the crest factor of the noise may be very high. Thus the energy content of a short duration impulse might be very small in comparison with the energy content of the signal, but the peak voltage of the impulse might exceed the signal peak voltage and become very annoying. The degree of this annoyance would of course depend upon the type of service and will not be gone into here. In view of this importance of peak noise considerations, the final judgment in the comparison between the systems of frequency and amplitude modulation treated here will be based upon peak signal-noise ratios.

When the peak voltage or current ratio of the frequency and amplitude modulation spectra is to be determined, the characteristics of the different types of noise must be taken into consideration. There seem to be two general types of noise which require consideration. The first of these is fluctuation noise, such as thermal agitation and shot effect, which is characterized by a random relation between the various frequencies in the spectrum. The second is impulse noise, such as ignition or any other type of noise having a spectrum produced by a sudden rise of voltage, which is characterized by an orderly phase and amplitude relation between the individual frequencies in the spectrum.

Experimental data taken by the author have shown that the fluctuation noise crest factor is constant, independent of band width, when the carrier is strong compared to the noise. Thus the peak voltage of fluctuation noise varies with band width in the same manner that root-mean-square voltage does, namely, as the square-root of the band width ratio. Consequently, for the strong-carrier condition, the peak voltage characteristics of fluctuation noise may be determined by applying the experimentally determined crest factor to the root-mean-square characteristics. Hence, in the case of fluctuation noise, (15) applies for peak noise improvement as well as for average root-mean-square noise improvement.

Impulse Noise Characteristics

A simple way of visualizing the manner in which impulse noise produces its peak radio-frequency voltage is to consider the case of a

recurrent impulse. It is well known that a recurrent impulse, such as square-wave-form dots, may be expressed by a Fourier series which consists of a fundamental and an infinite array of harmonics. The amplitudes of these harmonics are inversely proportional to their frequencies. The components of the single impulse will be similar to those of the recurrent impulse since the single impulse may be considered as a recurrent impulse with a very low rate of recurrence. The part of this impulse spectrum that is received on a radio receiver is a small band of the very high order harmonics. Since the frequency difference between the highest and lowest frequencies of this band is small compared to the mid-frequency of the band, all of the frequencies received are of practically equal amplitude. These harmonics are so related to each other by virtue of their relation to a common fundamental that they are all in phase at the instant the impulse starts or stops. Hence, for the interval at the start or stop of the impulse, all of the voltages in the band add up arithmetically and the peak voltage of the combination is directly proportional to the number of individual voltages. Since the individual voltages of the spectrum are equally spaced throughout the band, the number of voltages included in a given band is proportional to the band width. Consequently, the peak voltage of the resultant of the components in the spectrum is directly proportional to the band width. Thus impulse noise varies, not as the square root of the band width, as fluctuation noise does, but directly as the band width.[7] Since the voltages in the spectrum add arithmetically, their peak amplitude is proportional to their average ordinate as well as proportional to the band width. This makes the peak voltage of impulse noise, not proportional to the square root of the ratio between the *squared-ordinates areas*, as is the case with root-mean-square noise, but proportional to the ratio between the *areas* of the two spectra. Hence, (referring to Fig. 3)

$$\frac{N_a}{N_f}\ (\text{peak values, impulse}) = \frac{\text{area } OCEH}{\text{area } ODH}$$

$$= \frac{(n/C) \times F_a}{F_a \times \frac{1}{2} \times \frac{2F_a}{F_i} \times \frac{n}{C}} = \frac{F_i}{F_a}. \quad (16)$$

[7] The fact that the peak voltage of impulse noise varies directly with the band width was first pointed out to the author by V. D. Landon of the RCA Manufacturing Company. The results of his work were later presented by him as a paper entitled "A study of noise characteristics," before the Eleventh Annual Convention, Cleveland, Ohio, May 13, 1936; published in the Proc. I.R.E., vol. 24, pp. 1514–1521; November, (1936).

Dividing (16) by (14) to obtain the ratio between the frequency and amplitude modulation output signal-noise ratios gives

$$\frac{S_f/N_f}{S_a/N_a} \text{ (peak values, impulse)} = 2\frac{F_d}{F_a}. \tag{17}$$

Equation (17) shows that the frequency modulation peak voltage improvement with respect to impulse noise is equal to twice the deviation ratio or about 1.16 times more improvement than is produced on fluctuation noise. The peak power gain would be equal to the square of the peak voltage gain or four times the square of the deviation ratio.

Low Carrier-Noise Ratios

When the expression for the instantaneous frequency of the wave modulated by the noise component and signal, given by (7), is resolved into its components by the use of the binomial theorem, the following is the result:

$$f = F_c - F_d \sin pt - \frac{(F_{na} - F_d \sin pt)}{Z}\left[\frac{n}{C} - \left(1 - \frac{2n}{ZC}\right)\left\{\frac{K_0}{Z} - K_1 \cos(\omega_{na}t - F_d \sin pt) + K_2 \cos 2(\omega_{na}t - F_d \sin pt) + K_3 \cos 3(\omega_{na}t - F_d \sin pt) \cdots\right\}\right] \tag{18}$$

in which $Z = \frac{C}{n} + \frac{n}{C}$ and

$$K_0 = K_1 = \left(1 + \frac{3}{Z^2} + \frac{10}{Z^4} + \frac{35}{Z^6} + \frac{126}{Z^8} + \frac{462}{Z^{10}} + \frac{1716}{Z^{12}} + \cdots\right) \tag{19}$$

$$K_2 = \left(\frac{1}{Z} - \frac{4}{Z^3} - \frac{15}{Z^5} - \frac{56}{Z^7} - \frac{210}{Z^9} - \frac{792}{Z^{11}} - \frac{3003}{Z^{13}} - \cdots\right) \tag{20}$$

$$K_3 = \left(\frac{1}{Z^2} + \frac{5}{Z^4} + \frac{21}{Z^6} + \frac{84}{Z^8} + \frac{330}{Z^{10}} + \frac{1287}{Z^{12}} + \frac{5005}{Z^{14}} + \cdots\right) \tag{21}$$

Additional terms of the series of (19), (20), and (21), as well as higher order series, may be found with the aid of a table of binomial coefficients.

Equation (18) shows that, as the carrier-noise ratio approaches unity, the effective signal-noise ratio at the receiver output is no longer directly proportional to the carrier-noise ratio. The effective frequency deviation produced by the noise has harmonics introduced and a constant frequency shift added. The effect of the harmonics and

constant shift is to make the wave form of a single noise component very peaked and of the nature of an impulse. Because of the selectivity of the audio channel, none of the harmonics are present for the noise frequencies in the upper half of the audio spectrum. As the frequency of the noise voltage is lowered, more and more harmonics are passed by the audio channel and as a consequence, the peak frequency deviation due to the noise is increased. This can be more easily understood from the following calculation of the wave form produced by the instantaneous frequency deviation of the single noise component.

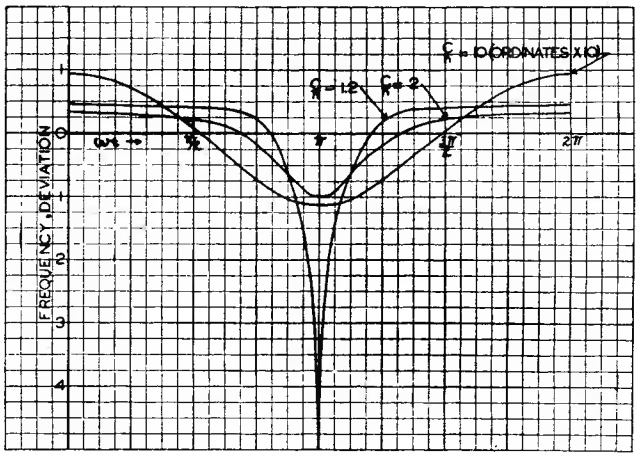


Fig. 4—Calculated wave forms showing the distortion produced on the instantaneous frequency deviation of the wave composed of the combination of the carrier and a single noise component. C/n = ratio between the peak voltage of the carrier and the peak voltage of the noise component.

The curves of Fig. 4 have been calculated from (7) and show how the instantaneous frequency deviation varies with time or the phase angle of the wave. A wave with the instantaneous frequency given by these curves would produce voltages in the output of the detector of the frequency modulation receiver which are proportional to the frequency deviations. It can be seen from these curves that, as the carrier-noise ratio approaches unity, the wave form becomes more and more peaked. The harmonics which enter in to make up this peaked wave form are given by (18) and are completely present for all noise frequencies only in the absence of audio selectivity.

In the presence of audio selectivity, the condition portrayed by (18) is approached as the audio frequency of the noise approaches zero. Thus the wave form of the noise is sinusoidal at a noise frequency high

enough to have its harmonics eliminated by the audio selectivity, but becomes more peaked as the frequency is made lower so that more harmonics are included. This effect tends to increase the peak voltage of the low-frequency noise voltages which have a large number of harmonics present. Thus, as the carrier-noise ratio approaches unity, the triangular audio spectrum is distorted by an increase in the amplitude of the lower noise frequencies.

The above gives a qualitative and partially quantitative description of the noise spectrum which results at the lower carrier-noise ratios. Further development would undoubtedly make possible the exact calculation of the peak and root-mean-square signal-noise ratio at the receiver output when the carrier-noise ratio at the receiver input is close to unity, but, because of the laborious nature of the calculations involved in evaluating the terms of (18), and pressure of other work, the author is relying upon experimental determinations for these data.

Noise Crest Factor Characteristics

The crest factor characteristics of the noise can be studied to an approximate extent by a study of (6). This equation portrays the resultant peak frequency deviation of the wave at the output of the limiter. From it, the crest factor characteristics of the output of the detector may be determined since in the frequency modulation receiver frequency deviations are linearly converted into detector output voltages. However, the crest factor characteristics of the receiver output are different from those at the detector output due to the effect of the selectivity of the audio channel. This is especially true in the case of the frequency modulation receiver with a deviation ratio greater than unity, that is, where the audio channel is less than one half the intermediate-frequency channel. Consequently, in order to obtain the final results, the effect of the application of the audio selectivity must be applied to the results determined from a study of (6).

From the curves of Fig. 4, it can be seen that the peak frequency deviation of the wave given by (7) occurs at a phase angle equal to 180 degrees. From the similarity of (6) and (7), it can be seen that the peak frequency deviation of (6) would also occur at a phase angle of 180 degrees. At this phase angle the noise peak frequency deviation from (6) is

$$f_{dn}(\text{peak}) = \frac{\left(F_{na} - \frac{1}{2\pi}\frac{d\phi(t)}{dt} - F_d \sin pt\right)}{\frac{(C/N) - 1}{(N/C) - 1} + 1}$$

$$= \frac{-\left(F_{na} - \frac{1}{2\pi}\frac{d\phi(t)}{dt} - F_d \sin pt\right)}{(C/N) - 1} \qquad (22)$$

Equation (22) shows that the peak frequency deviation of the noise, for any value of carrier-noise ratio, C/N, is proportional to the noise instantaneous audio frequency given by the quantity $\left(F_{na} - \frac{1}{2\pi}\frac{d\phi(t)}{dt} - F_d \sin pt\right)$, and to the quantity $1/\{(C/N) - 1)\}$.

C/N is the resultant instantaneous peak carrier-noise ratio which is present in the output of the frequency modulation intermediate-frequency channel. It is apparent that when this carrier-noise ratio is high, the peak frequency deviation of the noise is proportional to N/C. When the carrier-noise ratio is equal to unity, the peak frequency deviation becomes infinite and it is evident that the frequency modulation improvement, which is based on a high carrier-noise ratio, would be lost at this point. The term "improvement threshold" will be employed hereinafter to designate this point below which the frequency modulation improvement is lost and above which the improvement is realized. Theoretically this term would refer to the condition where the instantaneous peak voltage of the noise is equal to the peak voltage of the carrier. However, in the practical case, where only *maximum* peak values of the noise are measured, the improvement threshold will refer to the condition of equality of the *maximum* instantaneous peak voltage of the noise and the peak voltage of the carrier.

As the experimental characteristics will show, this increase in peak frequency deviation of the noise is manifested in an increase in crest factor of the noise. The crest factor cannot rise to infinity, however, due to the limitations imposed by the upper and lower cutoff frequencies of the intermediate-frequency channel. This selectivity limits the peak frequency deviation of the resultant of the noise and applied modulation to a value not greater than one half the intermediate-frequency channel width. Hence, in the absence of applied frequency modulation, the peak voltage of the noise at the detector output may rise to a value equal to the peak voltage due to the applied frequency modulation with maximum frequency deviation. In the presence of the applied frequency modulation, the total peak frequency deviation is limited so that the noise peaks depress the signal, that is, they punch holes in the signal, but do not rise above it. Thus a phenomenon which might be termed "frequency limiting" takes place. This frequency

limiting limits frequency deviations in the same manner that amplitude limiting limits amplitude deviations. The resulting effect is the same as though an amplitude limiter were placed at the detector output to limit the output so that the peak voltage of the noise or signal, or their resultant, cannot rise above a voltage corresponding to that produced by the signal alone at full modulation.

Since the frequency limiting limits the noise so that its maximum amplitude cannot rise above the maximum amplitude of the applied modulation, a noise suppression effect is present which is similar to that effected by the recent noise suppression circuits[8,9] used for reducing impulse noise which is stronger than the amplitude modulated carrier being received. The result of such limiting is a considerable reduction of the annoyance produced by an intermittent noise, such as ignition, where the duration of the impulses is short and the rate of recurrence is low. With such noise, the depression of the signal for the duration of the impulse reduces the presence of the signal for only a small percentage of the time; the resultant effect is a considerable improvement over the condition where the peaks of the noise are stronger than the signal. On the other hand, for steady noise such as fluctuation noise, as the carrier-noise ratio is made less than unity, the signal is depressed more and more of the time so that it is gradually smothered in the noise.

When the effect of the audio selectivity is considered in conjunction with the frequency limiting, it is found that the noise suppression effect is somewhat improved for the case of a deviation ratio greater than unity. The reason for this is as follows: The frequency limiting holds the peak voltage of the noise at the output of the detector so that it cannot rise above the maximum value of the signal. However, in passing through the audio channel, the noise is still further reduced by elimination of higher frequency components whereas the signal passes through without reduction. Consequently the over-all limiting effect is such that the noise is limited to a value which is *less* than the maximum value of the signal. The amount that it is less depends upon the difference between the noise spectra existing at the output of the detector and the output of the audio selectivity.

Experimental determinations, which will be shown later, point out that as unity carrier-noise ratio is approached, the frequency

[8] Leland E. Thompson, "A detector circuit for reducing noise interference in C.W. reception," *QST*, vol. 19, p. 38; April, (1935). A similar circuit for telephony reception is described by the same author in *QST*, vol. 20, pp. 44–45; February, (1936).

[9] James J. Lamb, "A noise-silencing I.F. circuit for superhet receivers," *QST*, vol. 20, pp. 11–14; February, (1936).

modulation audio noise spectrum changes from its triangular shape to a somewhat rectangular shape. Hence the noise spectrum at the output of the detector when frequency limiting is taking place would be approximately as given by *OCBA* of Fig. 3. When the audio selectivity is applied, the spectrum would be reduced to *OCEH* and the band width of the noise would be reduced by a ratio equal to the deviation ratio. This would reduce the peak voltage of fluctuation noise by a ratio equal to the square root of the deviation ratio and that of impulse noise by a ratio equal to the deviation ratio. Thus, the resultant effect of the frequency limiting is that the fluctuation noise output is limited to a value equal to the maximum peak voltage of the signal divided by the square root of the deviation ratio. The corresponding value of impulse noise is limited to a value equal to the maximum peak voltage of the signal divided by the deviation ratio. Consequently, with fluctuation noise, when the noise and signal are measured in the absence of each other, the signal-noise ratio cannot go below a value equal to the square root of the deviation ratio; the corresponding signal-noise ratio impulse noise cannot go below a value equal to the deviation ratio. However, these minimum signal-noise ratios are only those which exist when the noise is measured in the absence of the applied frequency modulation. When the applied modulation and the noise are simultaneously present, the noise causes the signal to be depressed. When this depressed signal, with its depression caused by noise composed of a wide band of frequencies, is passed through the audio selectivity, the degree of depression is reduced. The amount of the reduction will be different for the two kinds of noise. The determination of the actual magnitude of this reduction of the signal depression, as effected by the audio selectivity, will be left for experimental evaluation.

In comparing frequency modulation systems with different deviation ratios at the *low carrier-noise ratios*, the wider intermediate-frequency channel necessary for the high deviation ratio receiver gives that receiver a disadvantage with respect to the low deviation ratio receiver. Since this wider channel accepts more noise than the narrower intermediate-frequency channel of the low deviation ratio receiver, when equal carriers are fed to both such receivers equality of carrier and noise occurs at a higher carrier level in the high deviation ratio receiver. As a result, a higher carrier voltage is required to reach the improvement threshold in the case of the high deviation ratio system. Thus at certain low carrier levels, the carrier-noise ratio could be above the improvement threshold in the low deviation ratio system, but below in the high deviation ratio system; at this carrier level the

low deviation ratio system would be capable of producing a better output signal-noise ratio than the high deviation ratio system.

The difference between the improvement thresholds of receivers with different deviation ratios may be investigated by a determination of the carrier-noise ratio which exists in the reference amplitude modulation receiver when the improvement threshold exists in the frequency modulation receiver. This carrier-noise ratio may be found by a consideration of the relative band widths of the intermediate-frequency channels of the receivers. Thus, when the deviation ratio is unity, and the intermediate-frequency channel of the frequency modulation receiver is of the same width as that of the amplitude modulation receiver,[10] the two receivers would have the same carrier-noise ratio in the intermediate-frequency channels. When the deviation ratio is greater than unity, and the intermediate frequency channel of the frequency modulation receiver is broader than that of the amplitude modulation receiver, the carrier-noise ratio in the frequency modulation receiver is less than that in the amplitude modulation receiver. For the case of fluctuation noise, where the peak values vary as the square root of the ratio between the two band widths concerned, the carrier-noise ratio in the frequency modulation intermediate-frequency channel would be less than that in the amplitude modulation intermediate-frequency channel by a ratio equal to the square root of the deviation ratio. Thus, when equal carrier voltage is fed to both receivers,

$$C_a/N_a = (C/N)\sqrt{F_d/F_a} \text{ (fluctuation noise, peak or r-m-s values)} \quad (23)$$

in which C_a/N_a = carrier-noise ratio in the amplitude modulation intermediate-frequency channel and C/N = corresponding ratio in the frequency modulation intermediate-frequency channel.

In the case of impulse noise, where the peak values of the noise

[10] In order to assume that the frequency modulation receiver with a deviation ratio of unity has the same intermediate-frequency channel width as the corresponding amplitude modulation receiver, the assumption would also have to be made that the peak frequency deviation due to the applied frequency modulation is equal to one half the intermediate-frequency channel width. In the ideal receiver with a square-topped selectivity characteristic, this amount of frequency deviation would produce considerable out-of-channel interference and would introduce distortion in the form of a reduction of the amplitudes of the higher modulation of frequencies during the intervals of high peak frequency deviation. However, under actual conditions, where the corners of the selectivity characteristic are rounded, it has been found that the frequency deviation may be made practically equal to one half the normal selectivity used in amplitude modulation practice without serious distortion. Receivers with high deviation ratios are less susceptible to this distortion due to the natural distribution of the side bands for the high values of F_d/F_m which are encountered with such receivers.

vary directly with the band-width ratio, the carrier-noise ratios in the two receivers are related by

$$C_a/N_a = \frac{C}{N}\frac{F_d}{F_a} \text{ (impulse noise, peak values).} \qquad (24)$$

From (23), it can be seen that, with fluctuation noise, a carrier-noise ratio equal to the square root of the deviation ratio would exist in the amplitude modulation intermediate-frequency channel when the carrier-noise ratio is at the improvement threshold ($C/N=1$) in the frequency modulation intermediate-frequency channel. Likewise, from (24), with impulse noise, the frequency modulation improvement threshold occurs at a peak carrier-noise ratio in the amplitude modulation intermediate-frequency channel which is equal to the deviation ratio.

Effect of Application of the Modulation

For the condition of a carrier which is strong compared to the noise, the equation for the instantaneous frequency of the wave modulated by the noise and signal, given by (7), may be reduced to the following:

$$f = F_c - F_d \sin pt - \frac{n^2}{C^2}(F_{na} - F_d \sin pt) - \frac{n}{C}(F_{na} - F_d \sin pt)\cos\left\{\omega_{na}t + \frac{F_d}{F_m}\cos pt\right\}. \qquad (25)$$

By neglecting the inconsequential term proportional to n^2/C^2, applying the sine and cosine addition formulas, the Bessel function expansions, and the Bessel function recurrence formulas, (25) may be resolved into

$$\begin{aligned} f = F_c - (n/C)\Big[& J_0\left(\frac{F_d}{F_m}\right)F_{na}\cos\omega_{na}t \\ & - J_1\left(\frac{F_d}{F_m}\right)\{(F_{na}+F_m)\sin(\omega_{na}t+pt)+(F_{na}-F_m)\sin(\omega_{na}t-pt)\} \\ & - J_2\left(\frac{F_d}{F_m}\right)\{(F_{na}+2F_m)\cos(\omega_{na}t+2pt)+(F_{na}-2F_m)\cos(\omega_{na}t-2pt)\} \\ & + J_3\left(\frac{F_d}{F_m}\right)\{(F_{na}+3F_m)\sin(\omega_{na}t+3pt)+(F_{na}-3F_m)\sin(\omega_{na}t-3pt)\} \\ & + J_4\cdots\Big]. \end{aligned} \qquad (26)$$

This resolution shows that the application of frequency modulation

to the carrier divides the over-all transmission of the receiver into components due to the carrier and each side frequency. The amplitudes of these components are proportional to the frequency difference between the noise voltage and the side frequency producing the component. The frequency of the audio noise voltage in each one of these component spectra is equal to the difference between the side frequency and the noise radio frequency. Thus the application of the modulation changes the noise from a single triangular spectrum due to the carrier, into a summation of triangular spectra due to the carrier and side frequencies. In the absence of selectivity, the total root-mean-square noise would be unchanged by the application of the modulation since the root-mean-square summation of the frequency modulation carrier and side frequencies is constant; hence the root-mean-square summation of noise spectra whose amplitudes are proportional to the strength of the carrier and side frequencies would be constant. However, since selectivity is present, the noise is reduced somewhat. This can be seen by considering the noise spectrum associated with one of the higher side frequencies. The noise spectrum associated with this side frequency, which acts as a "carrier" for its noise spectrum, is curtailed at the high-frequency end by the upper cutoff of the intermediate-frequency channel. The region of noise below the side frequency is correspondingly increased in range, but yields high-frequency noise voltages which are eliminated by the audio-frequency selectivity. Consequently when modulation is applied, the noise is slightly reduced. The amount of this reduction may be calculated by a root-mean-square summation of the individual noise spectra due to the carrier and side frequencies. For the case of a deviation ratio of unity, an actual summation of the various spectra for full applied modulation has shown the root-mean-square reduction to be between two and three decibels depending upon the audio frequency of the noise. The same sort of summations also shows that the reduction becomes less as the deviation ratio is increased.

The weak-carrier root-mean-square noise characteristics in the presence of applied frequency modulation do not lend themselves to such straightforward calculation as the corresponding strong-carrier characteristics and will not be gone into here. The same can be said for the peak-noise characteristics in the presence of applied frequency modulation.

Transmitter Frequency Modulation Power Gain

The above considerations, which are based upon the equivalent conditions of equal carrier amplitude at the input of the amplitude

and frequency modulation receivers, do not take into account the power gain effected by the use of frequency modulation at the transmitter. Since the power in a frequency modulated wave is constant, the radio-frequency amplifier tubes in the transmitter may be operated in the class C condition instead of the class B condition as is required for a low level modulated amplitude modulation system. In changing from the class B to the class C condition, the output voltage of the amplifier may be doubled. Consequently a four-to-one power gain may be realized by the use of frequency modulation when the amplitude modulation transmitter uses low-level modulation. On the other hand, when the amplitude modulation transmitter uses high level modulation—that is, when the final amplifier stage is modulated, the power gain is not so great. However, for the purpose of showing the effect of a transmitter power gain, the amplitude modulation transmitter will be assumed to be modulated at low levels.

As this paper is in the final stages of preparation, systems of amplifying amplitude modulation have been announced wherein plate efficiencies of linear amplifiers have been increased practically to equal the class C efficiencies.[11,12] Since these systems are not in general use as yet, it will suffice to say that such improvements in amplitude modulation transmission will tend to remove the frequency modulation transmitter gain in accordance with these improvements. Hence the overall frequency modulation gain will more nearly approach that due to the receiver[13] alone.

With a four-to-one power gain at the transmitter, a frequency modulation system would deliver twice the carrier voltage to its receiver that an amplitude modulation system would with the same transmitter output stage. Hence (15) and (17), and (23) and (24) become, respectively,

$$\frac{S_f/N_f}{S_a/N_a}\ (\text{peak values, fluctuation noise}) = 2\sqrt{3}F_d/F_a \tag{27}$$

$$\frac{S_f/N_f}{S_a/N_a}\ (\text{peak values, impulse noise}) = 4F_d/F_a \tag{28}$$

[11] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," presented before Eleventh Annual Convention, Cleveland, Ohio, May 13, (1936); published in PROC. I.R.E., vol. 24, pp. 1163–1182; September, (1936).

[12] J. N. A. Hawkins, "A new, high-efficiency linear amplifier," *Radio*, no. 209, pp. 8–14, 74–76; May, (1936).

[13] The receiver and transmitter gain are mentioned rather loosely when they are separated in this way. However, it will be understood that the receiver gain could not be realized without providing a transmitter to match the requirements of the receiver.

$$C_a/N = (C/2N)\sqrt{F_d/F_a} \text{ (fluctuation noise, r-m-s or peak values)} \quad (29)$$

$$C_a/N = (C/2N)F_d/F_a \text{ (impulse noise, peak values).} \quad (30)$$

These equations show that this increase in carrier fed to the frequency modulation receiver not only increases the frequency modulation improvement, but also lowers the carrier-noise ratio received on the amplitude modulation receiver when the improvement threshold exists in the frequency modulation receiver.

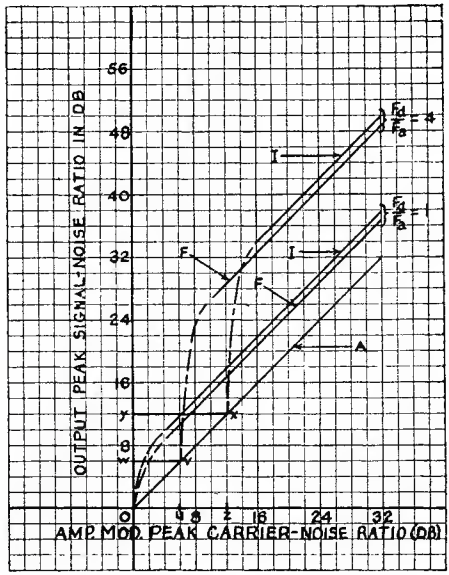


Fig. 5—Theoretical signal-noise ratio characteristics of frequency and amplitude modulation without the transmitter gain taken into account. Curve A = amplitude modulation receiver. The curves marked with I and F show the characteristics of the frequency modulation receivers for impulse and fluctuation noise, respectively. F_d/F_a = deviation ratio.

Theoretical Conclusions

The curves of Fig. 5 and 6 summarize the theoretical conclusions by means of an example in which receivers with deviation ratios of four and one are compared with each other and with an amplitude modulation receiver at various carrier-noise ratios. Fig. 5 shows the receiver gain only, whereas Fig. 6 takes into consideration a transmitter power gain of four to one. The curves are plotted with peak carrier-noise ratio in the amplitude modulation selectivity channel as a standard of comparison. Thus the curve for the amplitude modulation receiver is a straight line with a slope of forty-five degrees. The

curves for the frequency modulation receivers show output signal-noise ratios which are greater or less than those obtained from the amplitude modulation receiver depending upon the carrier-noise ratio.

For Fig. 5, (15) and (17) were used to obtain the strong-carrier frequency modulation improvement factors. Hence the frequency modulation output signal-noise ratios were obtained by multiplying the amplitude modulation signal-noise ratios by the frequency modulation improvement factors. The carrier-noise ratios which exist in the amplitude modulation receiver when the improvement threshold exists

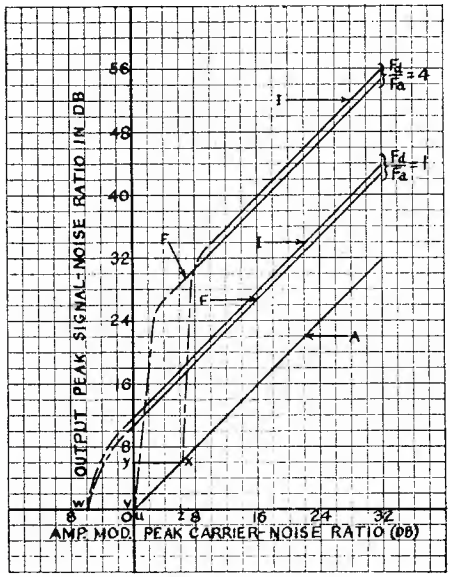


Fig. 6—Theoretical signal-noise ratio characteristics of frequency and amplitude modulation receivers with the transmitter gain taken into account.

in the frequency modulation receiver were determined by substituting a value of unity carrier-noise ratio in (23) and (24). The improvement thresholds are designated in both Figs. 5 and 6 by the points u and z for fluctuation and impulse noise, respectively. Since the theory does not permit actual calculation of signal-noise ratios in the region between high ratios and the improvement threshold, that part of the curves has been sketched in with a dashed line.

The part of the impulse-noise curve, for the deviation ratio of four represented by the line x-y shows the characteristic which would be obtained if the noise and signal were measured in the absence of each other. Because of frequency limiting, the noise is limited to equality

with the signal at the output of the detector and is then reduced in peak voltage by the audio selectivity. The amount of this reduction for impulse noise would be a ratio equal to the deviation ratio or, in this case, twelve decibels. In the case of fluctuation noise, the reduction of the noise, which is present in the absence of modulation, would be equal to the square root of the deviation ratio, or six decibels, and the corresponding curve is shown by the line *v-w*. However, these lines do not portray the actual signal-noise ratio characteristics since the noise depresses the signal when the carrier-noise ratios go below the improvement threshold. In the case of fluctuation noise this signal depression causes the signal to become smothered in the noise as the carrier-noise ratio is lowered below the improvement threshold. On the other hand, with impulse noise such as ignition, where the pulses are short and relatively infrequent, carrier-noise ratios below the improvement threshold will present an output signal which is depressed by the noise impulses, but which is quite usable due to the small percentage of time that the impulse exists.

The curves of Fig. 6, which take into account the frequency modulation transmitter gain, utilize (27) and (28) to obtain the frequency modulation improvements at the high carrier-noise ratios. These curves assume a carrier at the frequency modulation receiver inputs which is twice the strength of that present at the amplitude modulation receiver input. The frequency modulation improvements are therefore increased by six decibels and the improvement thresholds occur at signal-noise ratios in the amplitude receiver which are six decibels below the corresponding ratios for the case where the transmitter gain is not taken into account.

Further conclusions of the theory are as follows: For the high carrier-noise ratios, the application of modulation does not increase the root-mean-square value of the noise above its unmodulated value. Also, in the case of the low deviation ratio receivers, the root-mean-square value of the noise will be slightly reduced as the modulation is applied.

Experiment

In the experimental work it was desired to obtain a set of data from which curves could be plotted showing the frequency modulation characteristics in the same manner as the theoretical curves of Fig. 5. To do this it was necessary to have an amplitude modulation reference system and frequency modulation receivers with deviation ratios of unity and greater than unity. Equal carrier voltages and noise spectra could then be fed to these receivers and the output signal-noise ratios measured while the carrier-noise ratio was varied. Since it was

not convenient to measure the carrier-noise ratio at intermediate frequency, the output signal-noise ratios of the amplitude modulation receiver were measured instead and were plotted as abscissas in place of the carrier-noise ratios. This gives an abscissa scale which is practically the same as that which would be obtained by plotting carrier-noise ratios. The validity of this last statement was checked by measuring the linearity with which the output signal-noise ratio of the amplitude modulation receiver varied from high to low values as the carrier-noise ratio was varied by attenuating the carrier in known amounts in the presence of a constant noise. At the very low root-mean-square ratios the inclusion of the beats between the individual noise frequencies in the spectrum increases the apparent value of the

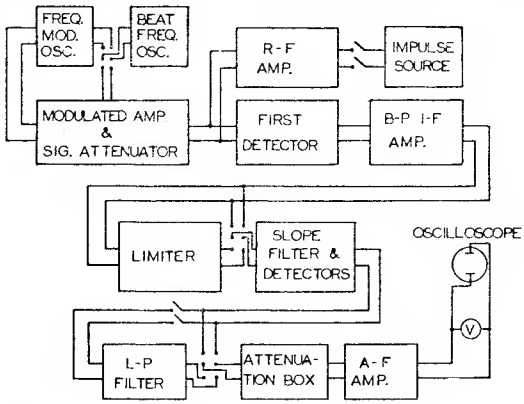


Fig. 7—Block diagram of experimental setup.

root-mean-square resultant of the noise voltages about two or three decibels. Thus, except for this small error at the low root-mean-square carrier-noise ratios, the amplitude modulation signal-noise ratio can be assumed equal to the carrier-noise ratio.

The block diagram of Fig. 7 shows the arrangement of apparatus used in obtaining the experimental data. The frequency modulated oscillator employed a circuit which was similar to that used in the previously mentioned propagation tests.[1] The modulated amplifier consisted of a signal generator which was capable of being amplitude modulated, but whose master oscillator energy was supplied from the frequency modulated oscillator. Thus a signal generator was available which was capable of being either frequency or amplitude modulated. A two-stage radio-frequency amplifier, tuned to the carrier frequency, but with no signal at its input, was used as the source of fluctuation noise. For the impulse noise measurements, the radio-frequency output

of a square-wave multivibrator was fed to the input of this radio-frequency amplifier.

In order to make available frequency modulation receivers with different deviation ratios, a method was devised which made possible the use of a single intermediate-frequency channel and detection system for all receivers. The method consisted in the insertion of a low-pass filter in the audio output of the receiver so as to reduce the width of the audio channel and thereby increase the deviation ratio of the receiver. This procedure is not that which might be normally followed since to increase the deviation ratio, the audio channel would normally be left constant and the intermediate-frequency channel increased.

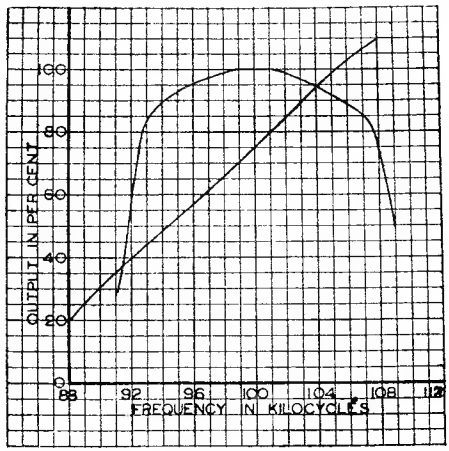


Fig. 8—Band-pass characteristic of receiver intermediate-frequency amplifier, and characteristic of sloping filter.

However, since it is only the *ratio* between the intermediate- and audio-frequency channels which governs the frequency modulation improvement, such an expedient is permissible for the purpose of the experiments.

The band-pass filter of the receiver intermediate-frequency amplifier was adapted from broadcast components and gave an output which was about one decibel down at 6500 cycles off from mid-band frequency. (See Fig. 8.) Hence maximum frequency deviation was limited to 6500 cycles. The audio channel of the receiver cut off at 6500 cycles and the low-pass filter cut off at 1600 cycles. Thus the following four different types of receivers were available: Number one, a frequency modulation receiver with a deviation ratio of unity which would receive a 6500-cycle modulation band. Number two, an amplitude modu-

lation receiver which would receive a 6500-cycle modulation band. Number three, a frequency modulation receiver with a deviation ratio of about four (6500÷1600) which would receive a 1600-cycle modulation band. Number four, an amplitude modulation receiver which would receive a 1600-cycle modulation band.

With these four receivers, a comparison between number two and number one would produce a comparison between amplitude modulation reception and frequency modulation reception with a deviation ratio of unity. A comparison between receivers number four and number three would produce a comparison between amplitude modulation reception and frequency modulation reception with a deviation ratio of four. Thus both frequency modulation receivers had as a standard of comparison an amplitude modulation receiver with an audio channel equal to that of the frequency modulation receiver.

The limiter of the frequency modulation receiver consisted of four stages of intermediate-frequency amplification arranged alternately to amplify and limit. The sloping filter detectors utilized the same circuit as used in the propagation tests[1] except that only one sloping filter was used in conjunction with a flat-top circuit as described in the theoretical section of this paper. Thus a balanced detector type of receiver was available which would also receive amplitude modulation by switching off the frequency modulation detector and receiving the detected output of the flat-top circuit. The characteristic of the sloping filter is shown in Fig. 8.

The output of the detectors was fed to a switching system which connected either to a low-pass filter and attenuator or directly to the attenuator. The output of the attenuator passed to an audio-frequency amplifier having an upper cut-off frequency of 6500 cycles. The indicating instruments were connected to the amplifier output terminals. For the root-mean-square fluctuation noise measurements, a copper-oxide-rectifier type meter was used.[14] A cathode-ray oscilloscope was used for all peak voltage measurements.

In the procedure used to obtain the data, the carrier-noise ratio was varied over a wide range of values and the receiver output signal-noise ratios were measured at each value of carrier-noise ratio. To do this, the output of the noise source was held constant while the carrier was

[14] In the preliminary measurements, a thermocouple meter was connected in parallel with the copper-oxide-rectifier meter in order to be sure that no particular condition of the fluctuation noise wave form would cause the rectifier meter to deviate from its property of reading root-mean-square values on this type of noise. It was found that the rectifier type of instrument could be relied upon to indicate correctly so that the remainder of the measurements of root-mean-square fluctuation noise were made using the more convenient rectifier type of instrument.

varied by means of the signal generator attenuator. The output peak signal-noise ratios were obtained by first measuring the peak voltage of the tone output with the noise source shut off and then measuring the peak voltage of the noise with the tone shut off. The *maximum* peak voltage of the noise was read for its peak voltage. The root-mean-square signal-noise ratios were measured by reading the root-mean-square voltage of the tone in the presence of the noise and then reading the voltage of the noise alone. The signal was then separated from the noise by equating the measured signal-plus-noise voltage to $\sqrt{S^2+N^2}$, substituting the measured noise voltage for N, and solving for the signal, S. In these measurements, a 1000-cycle tone was used to modulate

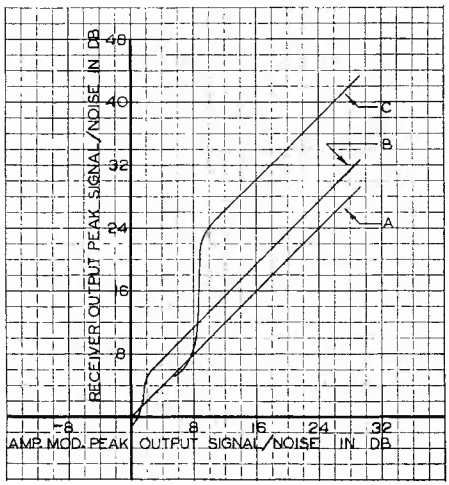


Fig. 9—Measured peak signal-noise ratio characteristics for fluctuation noise. Curve A = amplitude modulation receiver. Curve B = frequency modulation receiver with deviation ratio equal to unity. Curve C = frequency modulation receiver with deviation ratio equal to four.

at fifty per cent the amplitude modulator or to produce one-half frequency deviation (3250 cycles) on the frequency modulator. The output signal-noise ratios were corrected to a 100 per cent, or full modulation, basis by multiplying them by two. The radio frequency used was ten megacycles.

Fluctuation Noise Characteristics

The curves of Fig. 9 show the fluctuation noise characteristics, in which peak signal-noise ratios were measured. These curves check the theoretical curves of Fig. 5 as nearly as such measurements can be expected to check. With the deviation ratio of four (low-pass filter in), the

theoretical strong-carrier improvement should be $4 \times 1.73 = 6.9$ or 16.8 decibels; the measured improvement from Fig. 9 is about 14 decibels. With the deviation ratio of unity (low-pass filter out), the measured improvement was about 3.5 decibels as compared to the 4.76-decibel theoretical figure. The full frequency modulation improvement is seen to be obtained down to carrier-noise ratios about two or three decibels above the improvement threshold (equality of peak carrier and noise). The fact that the frequency modulation improvement threshold occurs at a higher carrier-noise ratio in the case of the receiver with a deviation ratio of four than in the case of the receiver with a deviation ratio of unity, also checks the theoretical predictions. In this case of fluctuation noise, the improvement threshold for the receiver with the deviation ratio of four should occur at a carrier-noise ratio in the amplitude modulation intermediate-frequency channel which was twice the corresponding ratio for the receiver with a deviation ratio of unity. The curves show these two points to be about seven decibels apart or within one decibel of the theoretical figure of six decibels.

The data for the curves of Fig. 9 were obtained by measuring the peak value of the noise alone and signal alone and taking the ratio of these two values as the signal-noise ratio. Hence the signal depressing effect, occurring for carrier-noise ratios below the improvement threshold, does not show up on the curves. In order to obtain an approximate idea as to the order of magnitude of this effect, observations were made in which the carrier-noise ratio was lowered below the improvement threshold while the tone modulation output (100 per cent modulation in the case of the amplitude modulation observation and full frequency deviation in the case of the frequency modulation observation) was being monitored by ear and oscilloscope observation. It was found that the fluctuating nature of the instantaneous peak voltage of the fluctuation noise had considerable bearing upon the effects observed. Due to the fact that the instantaneous value of the peak voltage is sometimes far below the maximum instantaneous value, frequency modulation improvement is obtained to reduce still further the peak voltage of these intervals of noise having instantaneous peak voltages lower than the maximum value. This effect seems to produce a signal at the output of the frequency modulation receiver which sounds "cleaner," but which has the same maximum peak voltage characteristics as the corresponding amplitude modulation receiver. Thus, as far as maximum peak voltage of the noise is concerned, the frequency modulation receiver produces about the same output as the amplitude modulation receiver for carrier-noise ratios below the improvement threshold. The reduction of the peak voltage of the noise during the

intervals of lower instantaneous peak value reduces the energy content of the noise in the output; hence some idea of the magnitude of this effect can be obtained from the root-mean-square characteristics of the noise.

The curves of Fig. 10 are similar to those of Fig. 9 except that the root-mean-square signal-noise ratios are plotted as ordinates. Since the crest factor of the signal is three decibels and that of fluctuation noise is about thirteen decibels (as later curves will show), the root-mean-square signal-noise ratios are ten decibels higher than the corresponding peak ratios. It can be seen that the root-mean-square characteris-

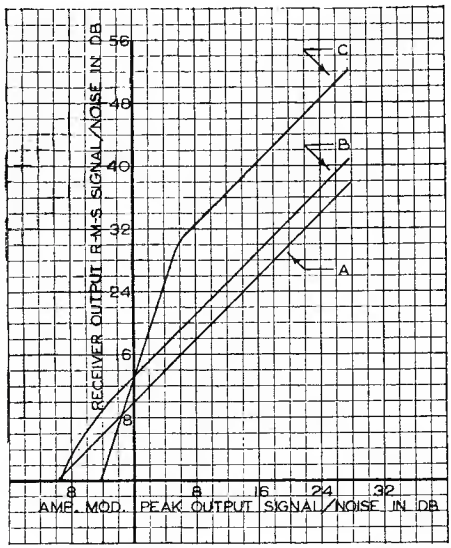


Fig. 10—Measured root-mean-square signal-noise ratio characteristics for fluctuation noise. Curve A = amplitude modulation receiver. Curve B = frequency modulation receiver with deviation ratio equal to unity. Curve C = frequency modulation receiver with deviation ratio equal to four.

tics differ from the peak characteristics in the range of carrier-noise ratios below the improvement threshold; above the improvement threshold, the characteristics are similar.

Since the root-mean-square and peak signal-noise ratios display different characteristics below the improvement threshold, it is quite evident that the crest factor of the noise changes as the carrier-noise ratio is lowered below this point. The crest factor can be obtained from the curves of Fig. 9 and 10 as follows: By adding three decibels to the ordinates of Fig. 10 they will be converted to peak signal to root-mean-square noise ratios. Hence by subtracting from these ratios the corre-

sponding ordinates of Fig. 9, the crest factor of the noise is obtained. The results of such a procedure are shown in Fig. 11.

In the case of the frequency modulation receiver with a deviation ratio of four, Fig. 11 shows that the crest factor increases by about 14.5 decibels at the improvement threshold. Hence the frequency modulation improvement, which is about fourteen decibels by measurement and sixteen by calculation, is counteracted by an increase in crest factor. This same situation exists in the case of the receiver with a deviation ratio of unity. Here the increase in crest factor is about four decibels; the measured frequency modulation improvement is about 3.5 decibels and the calculated value 4.76 decibels.

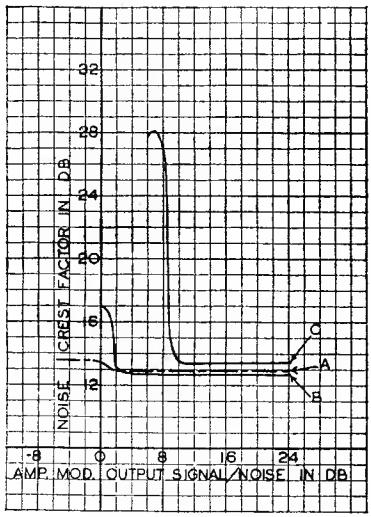


Fig. 11—Crest factor characteristics of frequency and amplitude modulation receivers. Curve A = amplitude modulation receiver with 6500-cycle audio channel. Curve B = frequency modulation receiver with deviation ratio equal to unity. Curve C = frequency modulation receiver with deviation ratio equal to four.

Curve A of Fig. 11 shows the crest factor characteristics of the amplitude modulation receiver. It is seen that this crest factor is about equal to that for frequency modulation above the improvement threshold. The average value of the crest factor for both amplitude and frequency modulation in this region is thirteen decibels or about 4.5 to one. This value checks previous measurements of crest factor where a slide-back vacuum tube voltmeter was used in place of an oscilloscope to measure the peak voltage and a thermocouple was used to measure the root-mean-square values.

The point where the crest factor of the noise increases, which occurs at the frequency modulation improvement threshold, has a rather distinctive sound to the ear. When fluctuation noise is being observed, as this point is approached the quality of the hiss takes on a more intermittent character, somewhat like that of ignition. This point has been termed by the author the "sputter point," and since it coincides with the improvement threshold it is a good indicator for locating the improvement threshold. It is caused by the fact that the fluctuation noise voltage has a highly variable instantaneous peak voltage so that there are certain intervals during which the instantaneous peak voltage of the noise is higher than it is during other intervals. Consequently, as the maximum peak value of the noise approaches the peak value of the signal, the higher instantaneous peaks will have their crest factor increased to a greater degree than the lower instantaneous peaks. Fig. 12 shows oscillograms taken on the fluctuation noise output of the

F *A*

Fig. 12—Wave form of the fluctuation noise output at unity carrier-noise ratio in the frequency modulation receiver. *F* = frequency modulation receiver. *A* = amplitude modulation receiver.

frequency and amplitude modulation receivers with the 1600-cycle low-pass filter in the audio circuit and with the signal-noise ratio adjusted to the sputter point. These oscillograms also tend to show how the frequency modulation signal would sound "cleaner" than the amplitude modulation signal when the carrier-noise ratio is below the improvement threshold.

Data were also taken to show the fluctuation noise characteristics as frequency modulation is applied. These data were taken by inserting low-pass or high-pass filters in the audio system and then applying a modulation frequency to the frequency modulated oscillator which would fall outside the pass band of the filters. The low-pass filter cut off at 1600 cycles so that modulating frequencies higher than 1600 cycles were applied. The output of the filter contained only noise in the range from zero to 1600 cycles and the change of noise versus frequency deviation of the applied modulation could be measured. The high-pass filter also cut off at 1600 cycles so that measurements of the noise in the range from 1600 to 6500 cycles were made while applying

modulation frequencies below 1600 cycles. In the case of the high-pass filter, the harmonics of the modulating frequencies appeared at the filter output in addition to the noise. Consequently, a separate measurement of the harmonics in the absence of the noise was made so that the noise could be separated from the harmonics by the quadrature relations. The results with the low-pass filter are shown in Fig. 13. The results with the high-pass filter are shown in Fig. 14.

The curves of Fig. 13 are representative of a system with a deviation ratio of four. They point out the fact that when the peak carrier-noise ratio in the frequency modulation intermediate-frequency channel

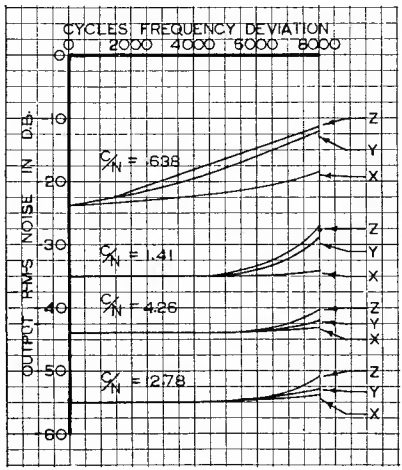


Fig. 13—Variation of frequency modulation receiver output noise as frequency modulation is applied. 1600-cycle low-pass filter in audio output. Modulation frequency: for curve $X = 6000$ cycles, $Y = 3000$ cycles, and $Z = 2000$ cycles. C/N = peak carrier-noise ratio in the output of intermediate-frequency channel.

is greater than unity, the root-mean-square noise is substantially unchanged due to the application of modulation. The one curve for a carrier-noise ratio less than unity shows a gradual increase of the noise, which would effect a decrease of the signal-noise ratio as the modulation is applied; this increase in the noise is displayed to a greater extent on the lower modulation frequency of 2000 cycles than on the higher modulation frequencies of 3000 and 6000 cycles.

Fig. 14 is approximately representative of a receiver with a deviation ratio of unity. This is because the range of noise frequencies from zero to 1600 cycles, which were eliminated by the high-pass filter, were a small part of the total range extending out to 6500 cycles. At the

highest carrier-noise ratio, the noise is decreased as the modulation is applied. This is in accordance with the deductions of the theory in the section *Effect of Application of Modulation.* As the carrier-noise ratio is lowered this tendency is eliminated.

Data similar to that for Fig. 13, with the low-pass filter in the audio circuit, were taken measuring the output *peak* voltage of the noise. The characteristics obtained were identical to those obtained with root-mean-square measurements.

Since the harmonics of the tone present in the output of the high-pass filter could not readily be separated from the noise for the peak

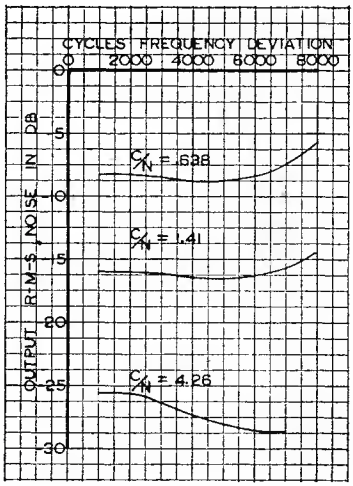


Fig. 14—Variation of frequency modulation receiver output noise as frequency modulation is applied. 1600-cycle high-pass filter in audio output. Modulation frequency = 1000 cycles.

voltage measurements, the high-pass filter data were taken by root-mean-square measurements only.

Measurements were also made to determine how much the audio selectivity reduced the degree of signal depression present at the output of the detector of the frequency modulation receiver. The carrier-noise ratio was set so that the maximum peak voltage of the fluctuation noise was equal to the peak voltage of the carrier. At this carrier-noise ratio the maximum noise peaks depressed the signal down to zero at the output of the detector. At the output of the 1600-cycle low-pass filter, the maximum noise peaks depressed the signal five decibels. Thus, without the audio selectivity, the signal was depressed by an amount equal to its total amplitude; with the audio selectivity, the

signal was depressed to five decibels below full amplitude or down to an amplitude of 56 per cent. Hence the reduction of the depth of the signal depression was from a 100 per cent depression to a depression of $(100-56)=44$ per cent or a reduction of about seven decibels. The theoretical reduction of the fluctuation noise in the absence of the modulation would be equal to the square root of the ratio of band widths or six decibels. Thus the reduction of the signal depression is, for all practical purposes, the same as the reduction in the peak voltage of the noise alone.

Impulse Noise Measurements

The first measurements on impulse noise were made using an automobile ignition system driven by an electric motor. However the output from this generator proved to be unsteady and did not allow a reasonable measurement accuracy. Consequently a square-wave multivibrator was set up. This type of impulse noise generator proved to be even more stable than the fluctuation noise source and allowed accurate data to be obtained. On the other hand, the output of the receiver being fed by this noise generator was not as steady as would be expected. In the absence of the carrier the output was steady, but as the carrier was introduced the output peak voltage started to fluctuate. Apparently the phase relation between the components of the noise spectrum and the carrier varies in such a manner as to form a resultant wave which varies between amplitude modulation and phase or frequency modulation. Hence the output of a receiver which is adjusted to receive either type of modulation alone will fluctuate depending upon the probability considerations of the phase of combination of the carrier and noise voltages.

The preliminary impulse noise measurements were made on an amplitude modulation receiver by comparing the peak voltage ratio between the two available band widths of 6500 and 1600 cycles. The 6500-cycle channel was fed to one set of oscilloscope plates and the 1600-cycle channel to the other. Thus, when the peak voltages at the outputs of the two channels were equal the oscilloscope diagram took a symmetrical shape somewhat like a plus sign. The two channel levels were equalized by means of a tone. Hence, when the noise voltage was substituted for the tone, the amount of attenuation that had to be inserted in the wider band to produce a symmetrical diagram on the oscilloscope was taken as the ratio of the peak voltages of the two band widths. In this manner a series of readings was taken which definitely proved that the peak voltage ratio of the two band widths was proportional to the band width ratio. These readings were taken on both the

ignition system noise generator and the multivibrator generator. As a check, readings on fluctuation noise were also taken which showed that the peak voltage of fluctuation noise varies as the square root of the band width.

The final measurements on impulse noise were made using the same procedure followed for the fluctuation noise measurements of Fig. 9. Only peak voltage measurements were made on this type of noise. The curves are shown in Fig. 15. It can be seen that the peak voltage characteristics of impulse noise are similar to those of fluctuation noise except for the location of the improvement threshold. For the receiver

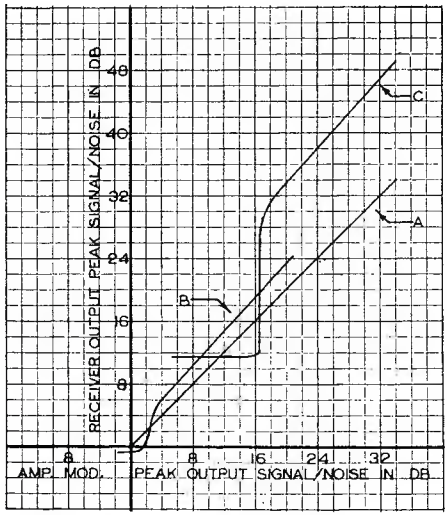
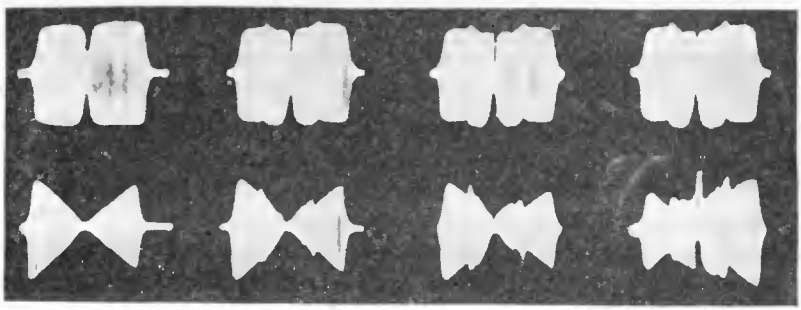


Fig. 15—Measured peak signal-noise ratio characteristics of impulse noise. Curve A = amplitude modulation receiver. Curve B = frequency modulation receiver with deviation ratio of unity. Curve C = frequency modulation receiver with deviation ratio of four.

with a deviation ratio of four, the improvement threshold occurs at a carrier-noise ratio slightly above sixteen decibels as compared with slightly above eight decibels for fluctuation noise. The difference between the improvement thresholds for the two frequency modulation receivers is about fourteen decibels; the corresponding theoretical figure, which is equal to the ratio of the two deviation ratios, is twelve decibels. The theoretical difference between the strong-carrier frequency modulation improvements for impulse and fluctuation noises, as indicated by the difference between the factors two and the square root of three respectively, is too small to be measurable with such variable quantities as these noise voltages.

Since the signal-noise ratios for the curves of Fig. 15 were obtained by measuring the noise and signal in the absence of each other, the signal-depressing effect of the noise does not show up. However, in the case of impulse noise, these curves are more representative of the actual situation existing, because the noise depresses the signal for only a small percentage of the time. In the listening and oscilloscope observations conducted with carrier-noise ratios below the improvement threshold, it was observed that at unity carrier-noise ratio the noise peaks depressed the amplitude of the signal to zero at the output of the detector. When the low-pass filter was inserted in the audio circuit, the impulse noise peaks depressed the signal about 2.5 decibels or reduced the amplitude from 100 per cent to 75 per cent. The effective signal-noise ratio is then increased from unity to $100/(100-75)=4$ or 12



$C/n=20$ $C/n=2$ $C/n=1.26$ $C/n=1$

Fig. 16—Over-all transmission oscillograms of the frequency and amplitude modulation receivers. 1600-cycle low-pass filter in audio output. Top row = amplitude modulation, bottom row = frequency modulation. C/n = ratio between the carrier and the variable-frequency heterodyning voltages.

decibels. This is equal to the theoretical reduction in peak voltage of impulse noise which would be effected by this four-to-one band width ratio. It is then evident that the reduction of the depth of the signal depression caused by the impulse noise is of the same magnitude as the reduction of the peak voltage of the noise alone.

Over-all Transmissions

The oscillograms of Fig. 16 show the over-all transmissions of the amplitude and frequency modulation receivers at various carrier-noise ratios. These oscillograms were taken by tuning the receiver to a carrier, and then, to simulate the noise, manually tuning a heterodyning signal across the intermediate-frequency channel. The audio beat output of the receiver was applied through the low-pass filter to the verti-

cal plates of the oscilloscope. A bias proportional to the frequency change of the heterodyning voltage was applied to the other set of oscilloscope plates. Consequently the spectra obtained are those which would be produced by the combination of a single noise component of variable frequency and the carrier. At the higher carrier-noise ratios, the spectrum is rectangular for amplitude modulation and triangular for frequency modulation. The dip in the middle of the amplitude modulation spectrum is where the audio output is near zero beat. As the carrier-noise ratio is decreased, the frequency modulation spectrum deviates from its triangular shape and the wave form of the receiver output has increased harmonic content at the lower audio frequencies where the audio selectivity does not eliminate the harmonics.

The amplitude modulation spectra of Fig. 16 also show the presence of added harmonic distortion on the lower modulation frequencies and lower carrier-noise ratios. However, the effect is so small that it is of little consequence.

The spectra of Fig. 16 allow a better understanding of the situation which is theoretically portrayed by (7) of the theory.

Experimental Conclusions

It can be concluded that the experimental data in general confirm the theory and point out the following additional information:

The improvement threshold starts at a carrier-noise ratio about three or four decibels above equality of peak carrier and noise in the frequency modulation intermediate-frequency channel. Hence the full frequency modulation improvement may be obtained down to a peak carrier-noise ratio in the frequency modulation receiver of three or four decibels.

The root-mean-square fluctuation noise characteristics differ from the peak fluctuation noise characteristics for carrier-noise ratios below the improvement threshold. The improvement threshold starts at about the same peak carrier-noise ratio, but the improvement does not fall off as sharply as it does for peak signal-noise ratios. Thus, for carrier-noise ratios below the improvement threshold the energy content of the frequency modulation noise is reduced, but the peak characteristics are approximately the same as those of the amplitude modulation receiver. The characteristics are not exactly the same due to the frequency limiting which allows the noise peaks to depress the signal, but does not allow them to rise above the signal.

The crest factor of the fluctuation noise at the outputs of the frequency and amplitude modulation receivers is about thirteen decibels or 4.5 to one for the strong-carrier condition. The crest factor of ampli-

tude modulation fluctuation noise remains fairly constant regardless of the carrier-noise ratio. At equality of peak carrier and peak noise in the frequency modulation intermediate-frequency channel, the crest factor of the noise in the output of the frequency modulation receiver rises to a value which counteracts the peak signal-noise ratio improvement over amplitude modulation; the improvement threshold manifests itself in this manner.

At the improvement threshold, the application of the audio selectivity reduces the signal depression due to a noise peak by the same ratio that it reduces the noise in the absence of the signal. Thus the depth of a noise depression in the signal is reduced by a ratio equal to the square root of the deviation ratio in the case of fluctuation noise, and equal to the deviation ratio in the case of impulse noise.

General Conclusions

The theory and experimental data point out the following conclusions:

A frequency modulation system offers a signal-noise ratio improvement over an equivalent amplitude modulation system when the carrier-noise ratio is high enough. For fluctuation noise this improvement is equal to the square root of three times the deviation ratio for both peak and root-mean-square values. For impulse noise the corresponding peak signal-noise ratio improvement is equal to twice the deviation ratio. When the carrier-noise ratio is about three or four decibels above equality of peak carrier and peak noise in the frequency modulation intermediate-frequency channel, the peak improvement for either type of noise starts to decrease and becomes zero at a carrier-noise ratio about equal to unity. Below this "improvement threshold," the peak characteristics of the frequency modulation receiver are approximately the same as those of the equivalent amplitude modulation receiver. The root-mean-square characteristics of the frequency modulation noise show a reduction of the energy content of the noise for carrier-noise ratios below the improvement threshold; this is evidenced by the fact that the improvement threshold is not as sharp for root-mean-square values as for peak values.

At the lower carrier-noise ratios, frequency modulation systems with lower deviation ratios have an advantage over systems with higher deviation ratios. Since the high deviation ratio system has a wider intermediate-frequency channel, more noise is accepted by that channel so that the improvement threshold occurs at a higher carrier level in the high deviation ratio system than in the low. Hence the

low deviation ratio systems retain their frequency modulation improvement down to lower carrier levels.

The peak voltage of fluctuation noise varies with band width in the same manner as the root-mean-square voltage, namely, as the square root of the band width. The peak voltage of impulse noise varies directly as the band width. In frequency modulation systems with a deviation ratio greater than unity, this difference in the variation with band width makes the improvement threshold occur at a higher carrier level with impulse noise than with fluctuation noise. Hence frequency modulation systems with higher deviation ratios are more susceptible to impulse noise interference.

Because of a phenomenon called "frequency limiting" the peak frequency deviations of the noise or the noise-plus-signal are limited so that the peak value cannot rise above the maximum peak value of the signal at the output of the detector. The application of audio selectivity reduces this maximum value of the noise so that fluctuation noise cannot rise to a value higher than the maximum value of the signal divided by the square root of the deviation ratio; the corresponding value of impulse noise cannot rise to a value higher than the maximum peak voltage of the signal divided by the deviation ratio. Inherent with this limiting effect is a signal-depressing effect which causes the fluctuation noise gradually to smother the signal as the carrier-noise ratio is lowered below the improvement threshold. However in the case of impulse noise, the signal depression is not as troublesome, and a noise-suppression effect is created which is similar to that effected in the recent circuits for suppressing impulse noise which is stronger than the carrier in an amplitude modulation system. When the deviation ratio is greater than unity, this frequency limiting is more effective than the corresponding amplitude modulation noise-suppression circuits; this is caused by the audio selectivity reducing the maximum peak value of the noise so that it is less than the peak value of the signal.

For carrier-noise ratios greater than unity, the application of frequency modulation to the carrier does not increase the noise above its value in the absence of applied frequency modulation.

At the transmitter, a four-to-one power gain is obtained by the use of class C radio-frequency amplification for frequency modulation instead of the customary class B amplification as is used for low level amplitude modulation. Therefore, for the same transmitter power input, a frequency modulation system will produce at its receiver a carrier which is twice as strong as that produced at the receiver of an amplitude modulation system. This results in two effects: first, the

frequency modulation improvement is doubled for carrier-noise ratios above the improvement threshold; second, when the improvement threshold occurs in the frequency modulation receiver, the carrier-noise ratio existing in the amplitude modulation receiver is one half of what it would have been without the transmitter power gain.

ACKNOWLEDGMENT

The author thanks Mr. H. H. Beverage and Mr. H. O. Peterson for the careful guidance and helpful suggestions received from them during the course of this work. The assistance of Mr. R. E. Schock in the experimental work is also appreciated.

FREQUENCY MODULATION*†

By

S. W. Seeley

RCA License Laboratory, New York

Summary—This paper, which contains generalized theory of the principal characteristics of frequency-modulation systems, is purposely made broad in its attack in order to achieve simplicity of presentation of fundamentals. Detailed treatment of specific problems associated with this type of modulation is given in the articles listed in the appended bibliography.

I. Frequency and Phase Modulation

COMBINATION of discussions of frequency modulation and phase modulation into one treatment may seem improper, at first glance, when the subject to be treated is frequency modulation alone. It has been done in this section of the paper because the two are so related that an understanding of both is necessary for a true picture of either.

The simplest conception of a frequency-modulated wave is of one with unvarying amplitude whose frequency is altered cyclically above and below its mean unmodulated value. However, if we consider phase modulation in its simplest form, it can be seen that frequency modulation also results therefrom. Consider the case of Figure 1, in which a carrier wave of unvarying amplitude A has its phase advanced and retarded sinusoidally between values of $+\phi$ and $-\phi$, with respect to its unmodulated phase. In this case the peak-to-peak phase modulation is, of course, 2ϕ. At the end of its excursions at either $+\phi$ or $-\phi$ the frequency of the wave is identical to that of its unmodulated value. However, as it progresses from $-\phi$ toward $+\phi$ and back again, its frequency is alternately increased and decreased, with frequency maxima and minima occurring at the point P. Thus it can be seen that the instantaneous periods of maximum phase displacement and maximum frequency deviation are misplaced from each other by 90° at the modulating frequency. In other words, if the phase modulation is assumed to be a cosine function, the frequency modulation is inherently the rate of change of phase, which is the first derivative of that cosine function and thus is a sine function. This assumed example is of the simplest case where the modulating component is a pure sinusoidal frequency. If the modulating frequency is raised, but the amount of

* Decimal Classification: R148.2.
† Reprinted from *RCA REVIEW*, April, 1941.
‡ Now Manager, Industry Service Laboratory, RCA Laboratories Division, New York, N. Y.

phase modulation maintained constant (deviation between $+\phi$ and $-\phi$ maintained) the rate of rotation of the vector A as it passes point P will be increased, which means that the maximum and minimum frequencies occurring when the vector is at point P are also increased. Thus constant phase modulation at all modulating frequencies results in more frequency modulation when higher modulating frequencies are used than for lower ones.

Figure 2 shows another simple relation between frequency and phase modulation. In the upper plot the frequency is shown to vary as a square wave, first above and then below its mean value. As long as the frequency remains at its increased value, the phase must be advancing, as shown in the plot immediately beneath. Then when the frequency is reduced on the other part of the square wave, the phase

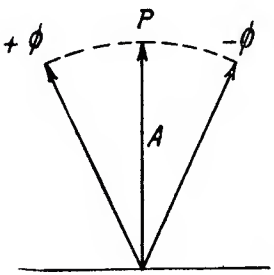


Fig. 1

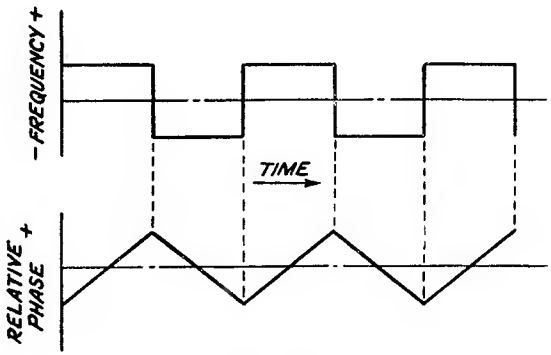


Fig. 2

is progressively retarded. The square wave will be instantly recognized as the first derivative of the triangular wave below. This is another simplified means of illustrating the fact that frequency modulation is a function of the first derivative or rate of change of any existing phase modulation.

From these considerations it is evident that if an end product is to consist of a frequency-modulated wave whose frequency deviation is proportional to the instantaneous amplitudes of a complex modulating wave, and if that frequency modulation is to be derived from a phase-modulating process, the complex modulating wave must first be integrated. In other words the applied modulating components must have amplitudes inversely proportional to their frequency, since a small amount of phase modulation at a high modulating frequency will result in the same amount of frequency modulation as a greater phase displacement at a lower modulating frequency.

II. Deviation and Band Width

The simple conception of a frequency-modulated wave, in which the frequency of a carrier wave of unvarying amplitude is altered cyclically about its mean value, is sufficiently adequate if the maximum frequency excursions of the carrier-wave frequency are large compared to the modulating frequency, for example, in the case of a wave being varied between limits separated by, say, 150 kc with the variation occurring sinusoidally at a 100-cycle per second rate.

If, however, the maximum frequency excursion is of the same order of magnitude, or equal to, or less than the frequency of change, the case is not so simple. It has often been proposed in all good faith that a communication channel could be narrowed by the use of frequency modulation in which the deviation in cycles per second is less than the modulating frequency. However, it is axiomatic that any side-band energy (and consequently the necessary intelligence component) in any type of modulation, whether frequency, phase, or amplitude, will not approach the carrier closer than a frequency interval equal to the modulating frequency. Therefore, if a narrow-band receiver is used to intercept a small deviation frequency-modulated wave whose modulating frequency is greater than one-half of the band width of the receiver, none of the modulation will be received. If we assume the case of a frequency-modulated wave with a modulating frequency of 10,000 cycles per second, which causes a deviation of + and — 10,000 cycles in the frequency of the carrier, a receiver with theoretically perfect band pass in its r-f and i-f circuits 19.9 kc wide with perfectly perpendicular sides, tuned with the carrier at the center of the band to receive that frequency-modulated wave, would not give any evidence of what modulating frequency was being used. A pure unvarying carrier with neither frequency, phase, nor amplitude modulation would be applied to its demodulating detector. It is true that in the example cited the received carrier would be somewhat decreased in value (actually to about 76 per cent) from its unmodulated amplitude, but this need not concern us for the moment. The example is cited simply to emphasize the fact that in frequency and phase modulation, as well as in amplitude modulation, the intelligence components are carried in side-band energy, and that that side-band energy must be received in order to interpret or reproduce the original modulation.

In the example cited above, side bands will exist not only at frequencies of + and — 10 kc with respect to the carrier, but also at + and — 20 kc, as well as at 30 kc, etc.[1] If the theoretically perfect

[1] See publications listed in the bibliography for more complete explanation of this characteristic.

band pass of our receiver were just sufficiently wide to pass only those side components at + and − 10 kc, we would have the condition shown in Figure 3A. In this figure the first three side bands are shown with their relative amplitudes when the peak frequency excursion divided by the modulating frequency is equal to unity. The theoretically perfect pass band is shown just including the carrier and the first two side bands. The resultant wave is as represented vectorially in Figure 3B, with the upper and lower side bands revolving about the point *P*, first causing a phase displacement to point *P'* and then to *P"*.

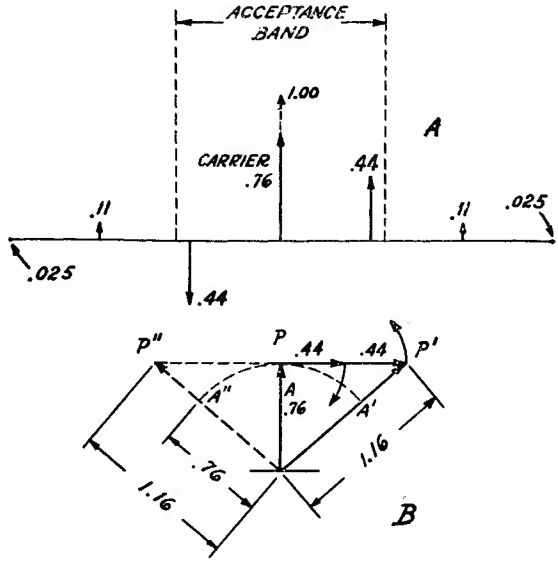


Fig. 3

Under these conditions it can be seen that the resultant instantaneous scalar magnitude of the carrier will vary between .76 and 1.16. In other words, amplitude modulation at twice the modulating frequency has resulted from deletion of all but the first side bands. This amplitude modulation contains no components at the 10,000-cycle frequency, is chiefly second harmonic, and contains small amounts of other even harmonics.

If now this wave shown emerging from our perfect 20-kc band pass filter is passed through a limiter stage which denudes the carrier of any amplitude modulation, the vector *A* will oscillate between the point *A'* and *A"*, which means that additional side bands have been recreated to cancel the amplitude modulation. It so happens that all of the even harmonic side bands are reinserted by a limiter stage, but that the odd harmonic side bands are not so reinserted, and thus a

small amount of distortion (principally third harmonic) in a demodulated wave which had been subjected to the process above described would result.

III. Noise Reducing Properties of Frequency Modulation

If a single, sharp, interfering impulse is applied to a band pass receiver, the resultant wave train which ensues is a function of the upper and lower frequencies of the pass band, and, for purposes of this discussion, can be assumed to be at one frequency only, namely that at the center of the pass band. An idealized mathematical expression for the wave train takes the form of

$$\frac{A \cos (w_1 - w_2)t}{t} \cos \frac{(w_1 + w_2)t}{2}$$

where A is a coefficient and w_1 and w_2 correspond to the upper and lower limits of the pass band. The expression of the form $\frac{\cos x}{x}$ (similar to the first part of the above expression) indicates a falling amplitude of wave envelope for positive and negative values of "t" as shown in Figure 4. Most of the energy of this wave train is, of course, under the large middle portion of the envelope, but the significant point is that for purposes of analysis the frequency can be assumed to be constant and equal to the mean frequency of the pass band, as shown by the second portion of the expression. Thus if this peak of interference, emerging from the pass band filters, is applied to a perfectly balanced back-to-back frequency-modulation detector or discriminator, its effect on both halves of that circuit is identical, and the net result is zero output.

If, however, a carrier wave at the mean frequency of the pass band is also present, the instantaneous phase of the interference energy and of the steady state carrier may take on any random value. Thus if the phase of the wave train due to the interfering impulse is, say 90° ahead of the carrier-wave phase and the interfering impulse is much stronger than the carrier wave, the instantaneous phase of the resultant of the two is altered considerably during the transmission of the interfering pulse. Therefore, even though the two waves together are passed through a limiter stage, the output of which is unvarying in amplitude, a certain amount of phase modulation (and consequently frequency modulation as well) has been developed, and the frequency modulation detector interprets it in the form of an output voltage or current. The same phenomenon may occur if recurrent interfering pulses overlap.

IV. Interference Between Carriers

If two transmitted carriers are received at one time and the beat between the two is at audible frequency, the sum of the two produces both amplitude and phase (or frequency) modulation. If one is twice the amplitude of the other, the resultant phase modulation amounts to ± 26.5 degrees regardless of the frequency of separation.

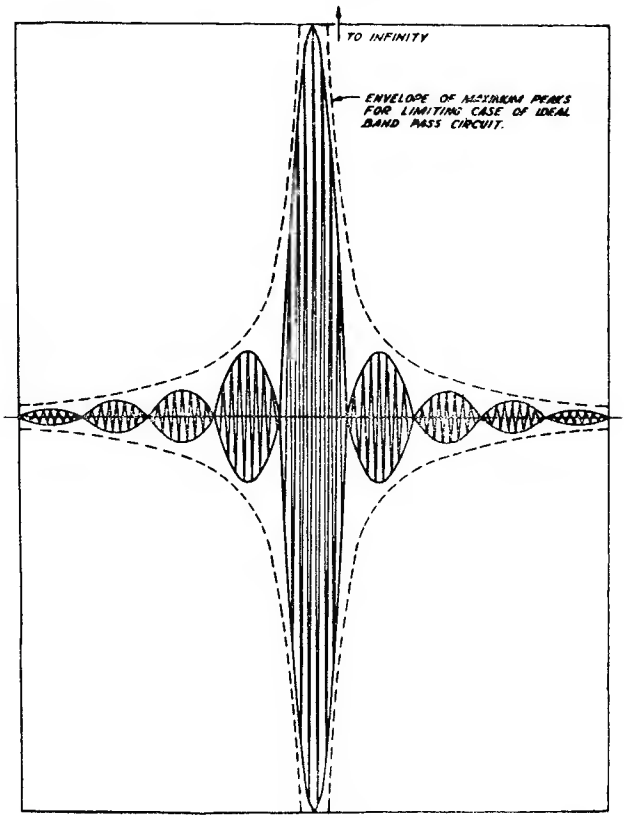


Fig. 4

This is shown in Figure 5. However, as pointed out in the first part of this discussion, the amount of frequency modulation resulting from a given amount of phase modulation is a function of the modulating frequency, which in this case can be assumed to be the frequency of separation of the two waves. Thus if the two frequencies are separated by 15,000 cycles, the equivalent frequency deviation of the frequency modulation amounts to approximately 7000 cycles per second. However, if the frequency of separation is only 100 cycles, the same

amount of phase modulation produces a frequency excursion or deviation of only 45 cycles per second.[2] Since the output of the frequency-modulation detector of a frequency-modulated wave receiver is dependent only upon the deviation, or excursion, from the mean frequency, it can be seen that the output will be 150 times greater when the carriers are separated by 15,000 cycles than when the separation is only 100 cycles. It is for this reason that the beat note between two frequency-modulated carrier wave stations becomes inaudible as the frequencies are brought closer and closer together.

V. Frequency-Modulation Methods

A frequency-modulated wave transmitter should emit a wave of

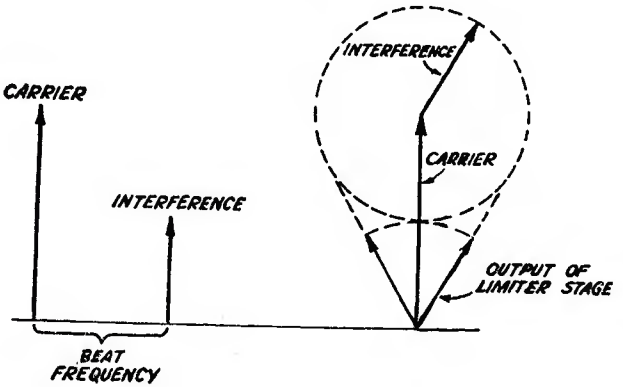


Fig. 5

unvarying amplitude with a frequency deviation about its mean (carrier) frequency in perfect conformity with the speech or program modulation. This neglects the matter of high-frequency pre-emphasis which will be discussed shortly.

There are two basic methods of producing this desired result. In either case, however, and unlike amplitude-modulated wave transmitters, the modulation takes place at the source of the transmitter frequency-determining circuits. Thereafter, the original oscillator-frequency energy may be amplified, multiplied in frequency, or heterodyned to any other frequency without altering the modulating components except to the extent that any multiplication of the source

[2] Actually when a single side component "phase modulates" an existing carrier, as in this case, the frequency modulation is not symmetrical. However, with a two to one ratio of original amplitude the "peak to peak" frequency shift is nearly equal to the values which the above figures would indicate.

frequency increases the deviation by an equal factor. Thus a doubler stage not only doubles the carrier frequency but also doubles the deviation width.

VI. Modulation of Phase to Produce Frequency Modulation

One means of producing frequency modulation is first to "integrate" the program material and then phase modulate some low-frequency source of energy; subsequent frequency multiplication can increase the resulting frequency deviation to any desired amount. After the desired band width has been obtained, the carrier with its modulating components can be heterodyned either directly to the transmitter frequency or to some sub-multiple thereof. In the latter case, additional deviation and band width are obtained by the subsequent multiplication to the transmitter frequency.

The integrating process usually consists of obtaining the potentials across a pure capacitive reactance in the plate circuit of a high impedance vacuum tube whose grid is supplied with the normal audio components. This causes a phase delay of 90° at all modulating frequencies and gives the program material an amplitude characteristic which is inversely proportional to frequency. If high-frequency pre-emphasis is to be used, the plate circuit may have a resistance in series with the capacitance with the output potentials taken across both. This allows more of the high-frequency components to enter the modulating circuit than would otherwise be the case. Present-day practice in pre-emphasis of both amplitude and frequency-modulated wave transmitters is to use a pre-emphasizing circuit equivalent to a series L/R network having a time constant of about 100 microseconds. Therefore, if the program integrating circuits for a phase-modulating frequency-modulated wave transmitter are to perform the pre-emphasis function also, the CR series combination must have a time constant of that value. Of course, the integration and pre-emphasis functions may be performed in separate stages, but it seems desirable to combine the two and thus prevent the extremely wide discrepancies between the levels of the low-frequency and high-frequency components of the program material at any point in the preparatory networks.

If the original frequency which is to be phase modulated by the prepared (integrated) program is produced by a crystal oscillator or other stable source of oscillations, the usual method of procedure is to amplitude modulate some of the energy at that frequency in a balanced modulator and then to re-insert carrier energy with an exactly 90°-phase rotation. The result is the same vectorially as is shown in Figure 6. The output contains both phase and amplitude modulation.

The amplitude modulation has little or no significance since it is always "limited" off in subsequent stages, but it is interesting to note that at least with sine-wave modulation, all amplitude modulation is at even harmonics of the modulation frequency without any of the fundamental being present. This is shown by the fact that the vector representing the carrier maximizes at each end of its phase excursion or twice each modulating cycle.

The developed phase modulation can be seen to be a function of the amplitudes of the horizontal vectors in Figure 6, which have been derived from the balanced modulator. As long as the angle, whose tangent is equal to the sum of the modulating vectors divided by the carrier vector, does not exceed 30° (the range through which the

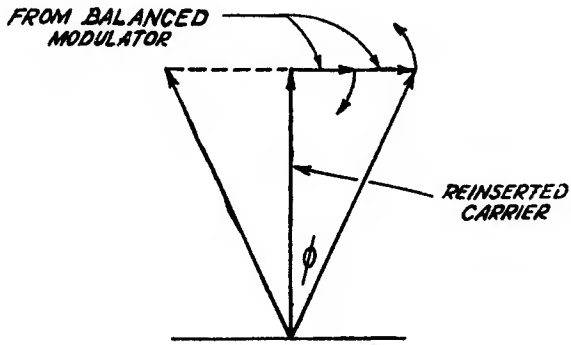


Fig. 6

tangent is approximately equal to the angle) the phase modulation can be said to vary linearly with the modulating components.[3]

The peak-frequency deviation of a frequency-modulated wave is related to the peak-phase deviation by the expression

$$\phi = \frac{360}{2\pi} \times \frac{\text{Frequency Deviation}}{\text{Modulating Frequency}}$$

in which ϕ is the peak-phase deviation in degrees.

If it is desired to produce a frequency deviation of ± 75,000 cycles with a 50-cycle modulating frequency, the phase displacement would amount to ± 86,000 degrees. If the assumption of 30 degrees for the

[3] Actually 30° phase modulation by this process results in nearly 10 per cent harmonic distortion, but since the resulting frequency modulation is a function of the product of the phase deviation and the modulating frequency the maximum phase displacement is required only for full modulation at the lowest modulating frequency.

maximum permissible-source modulation is divided into 86,000, the required carrier-frequency multiplication is seen to approach a value of 3000 times. If the original phase-modulated frequency is, say, 200 kc, eight "doubler" stages, providing a multiplication of 256 times, would give an output frequency of 51.2 megacycles. This frequency might then be heterodyned down to one-twelfth of the desired final-transmitter frequency. Two additional doubler stages and a tripler stage would then provide the correct carrier-output frequency and an overall multiplication of 3072. Obviously the heterodyning process has no effect on the absolute frequency deviation.

A sytem for phase modulation capable of producing phase deviations of several hundred degrees was described by Mr. R. E. Shelby of the National Broadcasting Company in a paper presented at the annual meeting of the Institute of Radio Enginers in New York in July, 1939. This system utilizes a cathode-ray tube of special design having a spiral target, which is scanned by a circularly deflected beam rotating at the carrier frequency or some sub-multiple thereof. The diameter of the circular pattern is varied in accordance with the desired modulation, and thereby alters the phase of the cycle at which the beam strikes the spiral anode. Output circuits connected to the spiral electrode can then extract the phase-modulated energy. This energy requires less frequency multiplication to produce a desired frequency deviation.

VII. Direct Frequency Modulation

A more direct system for producing frequency modulation utilizes a "reactance tube", similar to those used for automatic-frequency control, directly across the tank circuit of an oscillating vacuum tube. With this arrangement the transmitter-output frequency may be modulated directly without the need for frequency multiplication. However, it is usually convenient and desirable to use some doubler or tripler stages similar to those employed in a high-frequency amplitude-modulated wave transmitter. Of course, the source frequency cannot be crystal controlled directly, but carrier-frequency control can be maintained by heterodyning the carrier wave with a crystal-controlled oscillator to some low value, which is caused to operated a sensitive discriminator circuit. The output of the discriminator in turn affects the d-c bias of the original reactance control tube in the proper manner to correct for mistuning or frequency drift of the source oscillator. Carrier-wave stability within the limits of a few hundred cycles, even for a 40-Mc transmitter, can be maintained in this manner.

VIII. Generalized Frequency-Modulated-Wave Receiver Considerations

The radio frequency, first detector and intermediate frequency circuits of a frequency-modulated-wave receiver are identical basically with those in an amplitude-modulated-wave receiver. The only important difference is in the width of the pass band, which is generally several times greater in the frequency-modulated-wave receiver.

A detector of amplitude modulation, of and by itself, is immune to variations in the frequency of the applied wave. There can be no exact counterpart in the form of a frequency-modulation detector immune to amplitude variations. Zero amplitude is bound to produce zero output from any detector, so the output of all detectors must be some function of amplitude.

For this reason the second detector of a frequency-modulated-wave receiver is usually preceded by a limiter stage whose output is constant for wide variations in the amplitude of the applied signal. However, this is necessary only because of the inherent limitations of the detector and thus it might properly be considered as an integral part of the detection circuit.

Another characteristic of the usual frequency-modulation detector still further reduces its output response to amplitude variations. This is occasioned by the common practice of using so-called back-to-back or "push-pull" detectors. These are so arranged that a frequency deviation in a given direction causes an increase in the current or voltage output of one, while simultaneously causing a decrease in the output of the other. Both detectors, however, respond similarly to changes in amplitude of the applied wave. Thus with the output circuits of the two connected in series opposition, amplitude variations can act *only* to increase or decrease the "sensitivity" of the combination to frequency modulation, and cannot produce output in the absence of frequency deviations.

Frequency-Modulation Bibliography

Notes on the Theory of Modulation—J. R. Carson. *Proc. I.R.E.*, February 1922, p. 57.

The Reduction of Atmospheric Disturbances—J. R. Carson. *Proc. I.R.E.*, July 1928, p. 967.

Uber Frequenzmodulation—H. Roder. *Telefunken-Zeitung*, No. 53, 1929, p. 48.

Frequency Modulation—J. Harmon. *Wireless World*, January 22, 1930, p. 89.

A Study of the Frequency-Modulation Problem—A. Heilman. *E.N.T. (in German)*, June 1930, p. 217.

Frequency Modulation—B. Van der Pol. *Proc. I.R.E.*, July 1930, p. 1194.

Frequency Modulation and Distortion—T. L. Eckersley. *Experimental Wireless & Wireless Engineer*, Sept. 1930, p. 482.

Note on Relationships Existing Between Radio Waves Modulated in Frequency and in Amplitude—C. H. Smith. *Experimental Wireless & Wireless Engineer*, Nov. 1930, p. 609.

Amplitude, Phase and Frequency Modulation—Hans Roder. *Proc. I.R.E.*, Dec. 1931, p. 2145.

The Reception of Frequency-Modulated Radio Signals—V. J. Andrew. *Proc. I. R.E.*, May 1932, p. 835.

Phase Shift in Radio Transmitters—W. A. Fitch, *Proc. I.R.E.*, May 1932, p. 863.

A New Electrical Method of Frequency Analysis and Its Application to Frequency Modulation—W. S. Barrow. *Proc. I.R.E.*, October 1932, p. 1626.

Frequency Modulation and the Effects of a Periodic Capacitance Variation in a Non-dissipative Oscillatory Circuit—W. L. Barrow. *Proc. I.R.E.*, Aug. 1933, p. 1182.

High-Frequency Measurements—August Hund. *(A book) McGraw-Hill Co.*, 1933.

Transmission Lines as Frequency Modulators—A. V. Eastman and E. D. Scott. *Proc. I.R.E.*, July 1934, p. 878.

Theoretical and Experimental Investigation of Frequency and Phase-Modulated Oscillators—F. Lautenschlager. *E.N.T.*, October 1934, p. 357.

The Detection of Frequency-Modulated Waves—J. G. Chaffee. *Proc. I.R.E.*, May 1935, p. 517.

Phase-Frequency Modulation—D. G. Fink. *Electronics*, November 1935, p. 431.

Frequency Modulation on Ultra-Short Waves—D. Pollack. *Radio News*, February 1936, p. 458.

A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation—E. H. Armstrong. *Proc. I.R.E.*, May 1936, p. 689.

Frequency-Modulation Propagation Characteristics—M. G. Crosby. *Proc. I.R.E.*, June 1936, p. 898.

A Study of the Characteristics of Noise—V. D. Landon. *Proc. I.R.E.*, Nov. 1936, p. 1514.

Frequency-Modulated Generators—A. W. Barber. *Radio Engineering*, Nov. 1936, p. 14.

Frequency-Modulation Noise Characteristics—M. G. Crosby. *Proc. I.R.E.*, April 1937, p. 472.

Communication Engineering—W. L. Everitt. *(A book) McGraw-Hill Co.*, 1937, p. 408 (2nd edition).

Radio Engineering—F. E. Terman. *(A book) McGraw-Hill Co.*, 1937, p. 380.

Noise in Frequency Modulation—H. Roder. *Electronics*, May 1937, p. 22.

Application of the Autosynchronized Oscillator to Frequency Demodulation—J. R. Woodyard. *Proc. I.R.E.*, May 1937, p. 612.

Variable-Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation—Carson and Fry. *Bell System Technical Journal*, October 1937, p. 513.

Effects of Tuned Circuits on a Frequency-Modulated Signal—Hans Roder. *Proc. I.R.E.*, December 1937, p. 1617.

Armstrong's Frequency Modulator—D. L. Jaffe. *Proc. I.R.E.*, April 1938, p. 475.

Carrier and Side-Frequency Relations with Multi-Tone Frequency for Phase Modulation—M. G. Crosby. *RCA Review*, July 1938, p. 103.

Reduction of Interference by Frequency Modulation—E. H. Plump. *Hochfrequenztechnik und Electroakustic*, Sept. 1938, p. 73.

Communication by Phase Modulation—M. G. Crosby. *Proc. I.R.E.*, Vol. 27, No. 2, Feb. 1939, p. 126.

The Application of Negative Feedback to Frequency-Modulation Systems—J. G. Chaffee. *Bell System Technical Journal*, Oct. 1937, p. 404. *Proc. I.R.E.*, May 1939, p. 317.

A Receiver for Frequency Modulation—J. R. Day. *Electronics*, June 1939, p. 32.

A Noise-Free Radio Receiver for the Reception of Frequency-Modulated Ultra-Short Waves—G. W. Fyler and J. A. Worcester. *Gen. Electric Review*, July 1939, p. 307.

Frequency Modulation—C. H. Yocum. *Communications*, Nov. 1939, p. 5.

Frequency-Modulated Transmitters. *Electronics*, Nov. 1939, p. 20.

BAND WIDTH AND READABILITY IN FREQUENCY MODULATION*†

By

Murray G. Crosby

RCA Communications, Inc., Riverhead, L. I., New York

Summary—*The signal-to-noise ratio characteristics of frequency modulation are considered with reference to the effect of band width on readability in voice communication. Results of listening tests are described which indicate the optimum deviation ratio to be unity for maximum readability and distance of transmission. Other tests show the superiority of frequency modulation over amplitude modulation with respect to readability.*

ONE of the first questions which arises when frequency modulation is chosen for a communication system is that of the band width or amount of frequency deviation to be used. Among the many factors which affect the choice of band width is the subject of quality or readability of signal required. It is the purpose of this article to consider that subject and to attempt to show the relationship between band width and readability.

When the term readability is mentioned, consideration is limited to types of service in which the primary object of the system is the bare transfer of intelligence. This confines the consideration to services where the desire is to transmit voice the maximum distance with full readability for, while a high signal-to-noise ratio is desirable in the reception of voice, increasing the ratio above a certain value does not improve the readability. Consequently if the high signal-to-noise ratio is obtained at the expense of the ability to receive weaker signals, which is the case when the band width is made too wide in a frequency modulation system, the system is not working at its best efficiency. As will be shown here, when the primary object is the transmission of intelligence, the maximum distance will be covered when the band width is made the minimum possible.

---

* Decimal Classification: R148.2.

† Reprinted from *RCA REVIEW*, January, 1941.

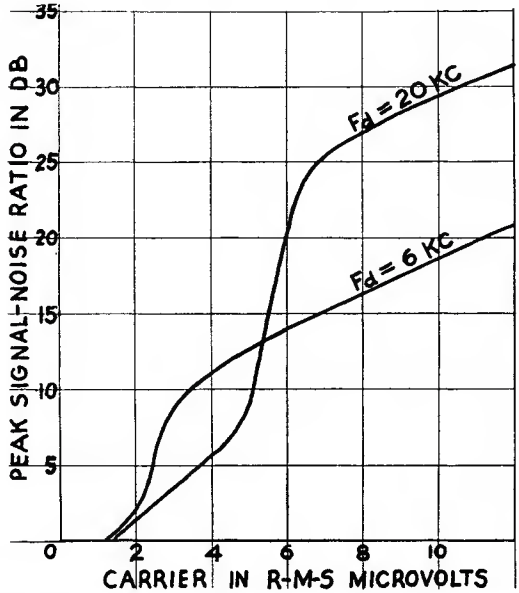


Fig. 1—Peak signal-to-noise ratio characteristics for frequency-modulation systems using maximum frequency deviations, F_d, of 20 and 6 kilocycles, respectively. The audio band width was 5 kilocycles. The noise was internal receiver noise and the signal-to-noise ratios were measured on an oscilloscope.

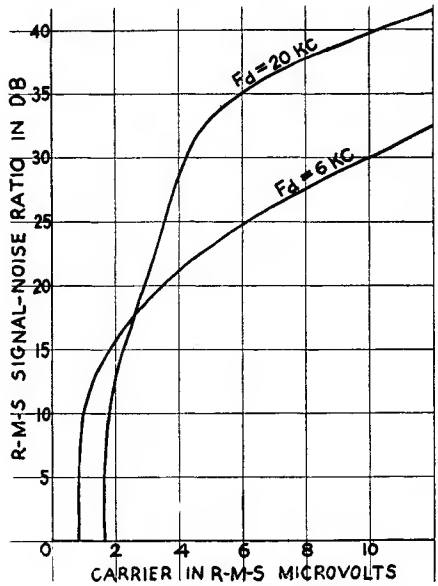


Fig. 2—Root-mean-square signal-to-noise ratio characteristics for the same systems used for Fig. 1. The signal-to-noise ratios were measured on a rectifier-type meter.

The reason for this superiority of a system using a low frequency deviation over one using a high frequency deviation, when maximum distance is the consideration, can best be shown by a study of the curves of Figures 1 and 2. These curves compare the peak and root-mean-square signal-to-noise ratio characteristics of two frequency modulation systems having maximum frequency deviations* of 6 and 20 kilocycles, respectively. The curves were taken by varying the carrier strength of the frequency-modulated signal generator and measuring the signal-to-noise ratio at the output of each receiver for the full frequency deviation that each receiver was capable of. The radio-frequency input circuit of the two receivers was common so that both receivers were on an equal basis as far as carrier strength and input noise were concerned. The noise consisted of the thermal agitation and tube hiss within the receiver. For the curves of Figure 1, the signal and noise in the receiver output were measured by means of an oscilloscope which indicates peak voltages. For those of Figure 2, an ordinary rectifier-type meter was used so that root-mean-square signal-to-noise ratios were obtained.

From these curves, it can be seen that, for carriers below a certain value, the low-deviation system produces a greater signal-to-noise ratio and is therefore more capable of "reaching down" into the noise to receive a signal. It will be noted that both systems give a signal-to-noise ratio which is approximately proportional to the carrier strength down to a certain carrier strength. Below that strength there is a rather sudden drop-off in signal-to-noise ratio. This drop-off is due to a phenomena which is peculiar to a frequency modulation system[1] and which is called the "improvement threshold" effect. Any frequency modulation system has an improvement threshold above which the frequency modulation gain or improvement is realized and below which the signal becomes submerged in the noise. The threshold occurs when the peak voltage of the carrier is equal to the peak voltage of the noise in the intermediate-frequency channel of the receiver. The full frequency modulation improvement is not realized until the carrier is about twice as strong as the noise.

It will be noted that the improvement threshold for the wider systems of Figures 1 and 2 occurs at a stronger carrier strength than that for the narrower system. Hence, other things being equal, the wider system requires more transmitter power to produce a signal which will be above the threshold. The reason for this will be apparent when it is realized that the wider receiver must have a wider

* "Deviation" in this paper refers to the amount of frequency shift to *one side* of the carrier.

[1] Murray G. Crosby, "Frequency Modulation Noise Characteristics", *Proc. of I.R.E.*, Vol. 25, No. 4, April 1937.

intermediate-frequency channel which inherently accepts a wider spectrum of the noise. The wider spectrum of noise has a larger peak voltage so that the improvement threshold distance for the wide system occurs at a higher carrier strength. Thus the threshold for the wider system of Figure 1 occurs at a carrier strength which is about twice the corresponding strength for the narrower system.

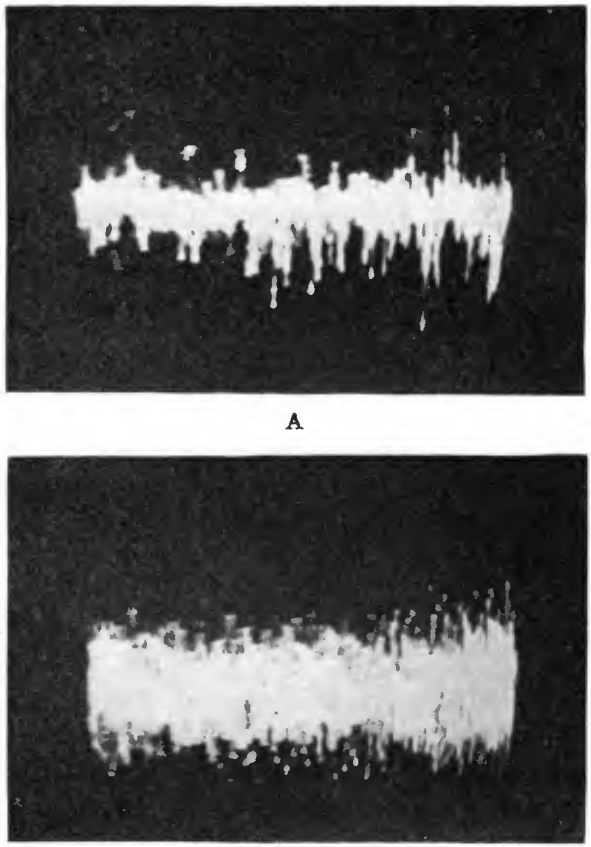
The difference between the shapes of the curves of Figures 1 and 2 is due to differences in crest factor (the ratio between the peak and root-mean-square voltage) of the noise. For carrier strengths above the improvement threshold, the crest factor is constant at a ratio of about 4.5. When the carrier is on the improvement threshold, the crest factor is higher by an amount which depends upon the deviation ratio (the ratio between the maximum frequency deviation and the maximum audio frequency of the system) of the frequency modulation receiver. This increase in crest factor is caused by the fact that the higher peaks of the noise approach equality with the carrier so that these peaks are at the threshold while the lower peaks produce a carrier-to-noise ratio which is above the threshold. When the peak voltages of the carrier and noise approach equality, the effective frequency variation of the resultant wave rises to very high values and the stronger of the two voltages assumes control of the receiver. Thus, if the noise is stronger than the signal the noise assumes control and depresses the signal. Hence as the carrier is lowered towards the threshold, the effective frequency deviation of the noise rises until the higher peaks begin punching holes in the signal. As the carrier is weakened still further, the weaker noise peaks also punch holes in the signal so that it is submerged in the noise. The point at which the highest peaks of the noise just begin to reach equality with the carrier peak voltage, produces a "sputtering" type of noise which changes the character of tube hiss or thermal agitation so that the improvement threshold is easily recognized. Figures 3A and B show the wave form of the noise at the threshold (which has been called the "sputter point") and above the threshold respectively.

When the noise is ignition or similar man-made noise the situation is similar to that shown by the peak signal-noise ratio curves of Figure 1. In addition, the frequency modulation receiver has inherent to it a noise-silencing action which is at least as effective if not better than the best amplitude-modulation noise silencer. This noise-silencing action is self-adjusting and is automatically adjusted for best operation as soon as the signal is tuned in. More detailed description of this action is considered elsewhere.[1,2]

[2] Murray G. Crosby, "The Service Range of Frequency Modulation", *RCA Review*, Vol. 4, No. 3, January 1940.

Listening Tests

With the signal-to-noise ratio characteristics as described above, it can be seen that the answer to the question regarding the relative readability obtainable with systems using different maximum frequency deviations depends upon the magnitude of signal-to-noise ratio required for a given readability. For instance, taking the curves of Figure 1,



A

B

Fig. 3—Oscillograms of internal receiver noise appearing at the frequency modulation receiver output. A is with carrier-to-noise ratio at the threshold or "sputter point". B is with carrier-to-noise ratio above the threshold.

if readability is obtainable on the narrower system for the signal strengths which are below the threshold of the wider system, the narrower system will have a range of superiority in readability of weaker signals. As will be seen from the following curves, practically full readability is obtained right down to the improvement threshold of the narrower system. With the wider system, full readability is

not obtainable until the signal is strong enough to reach the threshold of the wider system.

In order to determine the actual difference in signal readability for systems using different frequency deviations, a listening test was conducted on the two frequency-modulation systems which were used for Figures 1 and 2. A 5-kilocycle low-pass filter was inserted in the audio output of both receivers so that one had a deviation ratio of 1.2 and the other 4. The noise consisted of the thermal agitation and tube hiss originating in the radio-frequency circuits.

The curves of Figure 4 show the results of the listening tests. Readability numbers of the amateur RST system are plotted against

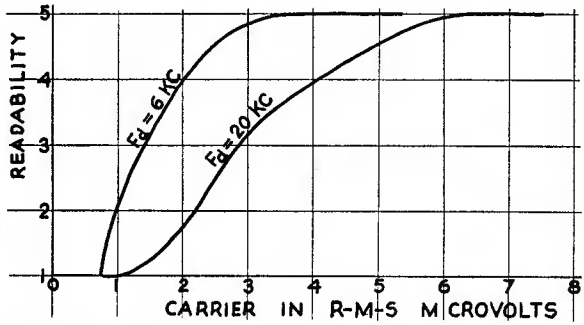


Fig. 4—Readability comparison of frequency modulation systems using frequency deviations of 20 and 6 kilocycles, respectively. The a.f. of both receivers cut off at 5 kilocycles. The ordinates are plotted in the readability scale of the RST signal-reporting system which is as follows:

1—Unreadable.
2—Barely readable, occasional words distinguishable.
3—Readable with considerable difficulty.
4—Readable with practically no difficulty.
5—Perfectly readable.

the microvolts output of the frequency-modulated signal generator. The points for the curves were taken from the averaged readings of three separate observers, namely A. M. Braaten, R. E. Schock, and the writer.

It is obvious from these tests that the narrower system is capable of "dipping down" deeper in the noise to receive a signal. For equal readability on the two systems, it appears that the system with a deviation ratio of 1.2 will receive a signal about one-half as strong as that possible with the system having a deviation ratio of 4. Hence the minimum readable signal strength decreases approximately proportional to the square-root of the ratio of the two deviation ratios. In the particular case of the systems used for Figure 4, changing the deviation ratio of the system from 4 to 1.2 is equivalent to an increase in power of about 4 times, as far as readability is concerned.

The curves of Figure 5 were taken in the same manner as those of Figure 4, but the two systems compared were the narrower frequency-modulation system of Figures 1 and 2 and its equivalent amplitude-modulation system. The maximum frequency deviation of the frequency-modulation system was 6 kilocycles and the audio band was 5 kilocycles. It can be seen that the narrower frequency-modulation system gives a readability which is always greater than that obtained on the amplitude-modulation system.

It is apparent from these tests that the optimum frequency deviation for a frequency-modulation system designed to obtain maximum distance for full readability is that which corresponds to a deviation ratio of one. Such a system produces very nearly the same intermediate-frequency and audio channel widths as the corresponding amplitude-

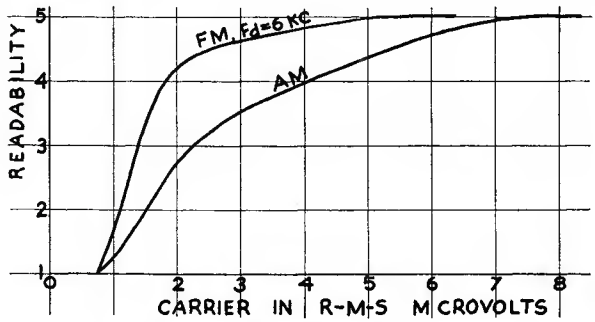


Fig. 5—Readability comparison of a frequency-modulation system and its equivalent amplitude-modulation system. Maximum frequency deviation = 6 kilocycles. Audio band = 5 kilocycles.

modulation system. To use a deviation ratio less than one would of course impair the efficiency without a compensating benefit in reduction of band width since it would be equivalent to an amplitude-modulation system which employed a modulation percentage less than 100.

The results of other listening tests[2] conducted by the writer have indicated that the use of pre-emphasis and de-emphasis, as is used in the present broadcast frequency-modulation systems, is of doubtful value when the object is the mere transmission of intelligibility instead of the enjoyment of a high-fidelity program. The peculiar triangular nature of the noise spectrum in the output of a frequency-modulation receiver is apparently more tolerable than the same noise spectrum after it has been made practically flat by the de-emphasis circuit. The results of the averaged observations of three observers showed that about 8 decibels more noise could be tolerated with the triangular frequency-modulation noise than with the rectangular amplitude-modu-

lation noise. That is to say that when the intelligibility of voice is being received through fluctuation noise like tube hiss or thermal agitation, the same intelligibility may be received with 8 decibels more noise of the triangular characteristic obtained from the output of a frequency-modulation receiver than with the flat or rectangular noise characteristic obtained from the output of an amplitude-modulation receiver. This advantage is lost when pre-emphasis and de-emphasis are used because the triangular noise spectrum is converted to a rectangular spectrum.

VARIATION OF BANDWIDTH WITH MODULATION INDEX IN FREQUENCY MODULATION*†

By

MURLAN S. CORRINGTON

Home Instrument Department, RCA Victor Division, Camden, N. J.

Summary—*Equations are derived for the carrier and side-frequency amplitudes which are obtained when a carrier wave is frequency-modulated by a complex audio signal. The bandwidth occupied by such a frequency-modulated wave is defined as the distance between the two frequencies beyond which none of the side frequencies is greater than 1 per cent of the carrier amplitude obtained when the modulation is removed.*

Curves are given to show the amount this bandwidth exceeds the extremes of deviation for a range of modulation indexes from 0.1 to 10,000, for sinusoidal, square, rectangular, and triangular modulation. For more precise definitions of bandwidth, curves are also given for side-frequency amplitude limits of 0.1 per cent and 0.01 per cent of the carrier-wave amplitude. For complex modulation the total bandwidth can be estimated by computing the bandwidth that would be required by each audio-frequency component, if it were on separately, and adding the results.

INTRODUCTION

IF A CARRIER wave is frequency-modulated with a sinusoidal audio voltage, an infinite number of side frequencies is produced. The carrier amplitude is reduced when the modulating voltage is applied, and may even become zero. If the deviation is increased, additional side frequencies are produced in both sidebands, and the distribution of the intensities of the previous ones is changed. For a single audio tone, the distance between side frequencies is always equal to the audio frequency. When two or more modulating tones are used simultaneously, side frequencies are produced which are separated from the carrier by all possible combination frequencies which can be obtained from sums and differences of harmonics of the modulating frequencies.[1,2] Thus, if there are two audio tones of frequencies μ_1 and μ_2, the side frequencies are separated from the

* Decimal classification: R148.2.

† Reprinted from *Proc. I.R.E.*, October, 1947.

[1] Murray G. Crosby, "Carrier and Side-Frequency Relations with Multitone Frequency or Phase Modulation," *RCA REVIEW*, vol. 3, pp. 103-106; July, 1938.

[2] M. Kulp, "Spektra und Klirrfaktoren Frequenz- und Amplituden-Modulierter Schwingungen," Part I. *Elek. Nach.-Tech.*, vol. 19, pp. 72-84; May, 1942.

carrier by $\pm r\mu_1 \pm s\mu_2$ where r and s are positive integers or zero.

Although, theoretically, an infinite number of side frequencies is produced, in practice the ones separated from the carrier by a frequency greater than the deviation decrease rapidly toward zero, so the bandwidth always exceeds the total frequency excursion, but nevertheless is limited. For large modulation indexes and a sinusoidal modulating voltage, the bandwidth approaches, and is only slightly greater than, the total frequency excursion.

To show how the bandwidth changes with modulation index, exact mathematical expressions for the spectrum will now be obtained.

The Spectrum of a Carrier Wave which is Frequency-Modulated with a Sinusoidal Signal

When a carrier wave is frequency-modulated with a single audio tone, the equation for the voltage is

$$e = E \sin \left(\omega t + \frac{D}{\mu} \sin 2\pi\mu t \right) \tag{1}$$

where

E = amplitude of the wave

ω = angular frequency of the carrier, radians per second

D = deviation, cycles per second

μ = audio frequency, cycles per second

t = time in seconds

D/μ = modulation index.

This expression can be expanded in a spectrum consisting of a carrier and side frequencies, in accord with the result[3]

$$e = E \sum_{n=-\infty}^{\infty} J_n (D/\mu) \sin (\omega t + 2\pi n\mu t) \tag{2}$$

where $J_n(D/\mu)$ is a Bessel function of the first kind of order n and argument D/μ.

Graphs of the Bessel Functions

To plot the spectrum of a frequency-modulated wave for a given

[3] A. Bloch, "Modulation Theory," *Jour. I.E.E.*, (London), Part III, vol. 91, pp. 31-42; March, 1944.

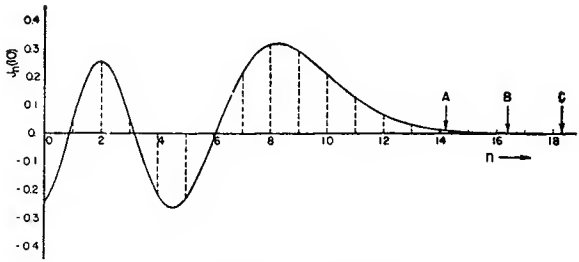


Fig. 1—Graph of $J_n(10)$.

modulation index D/μ, use a table of Bessel functions[4,5] to obtain the amplitudes of the carrier and the side frequencies in the upper sideband. The odd-order side frequencies in the lower sideband will have signs opposite to those in the upper sideband, and the even-order side frequencies will have the same sign. Figure 1 is a graph of $J_n(10)$. If the ordinates are drawn for each integer, as shown by the dotted lines, the side frequencies in the upper sideband will be proportional to these ordinates and the carrier will be proportional to the ordinate at $n = 0$.

If the modulation index is increased to 1000, the part of the curve for n nearly equal to the modulation index is similar, but is reduced in amplitude.[6] Figure 2 shows the variation of the side frequencies near the upper edge of the band. The curve oscillates with gradually increasing amplitude and slowly increasing period all the way from the origin to the last maximum, which is also the absolute maximum, and then decreases rapidly toward zero. The maximum energy occurs at a point

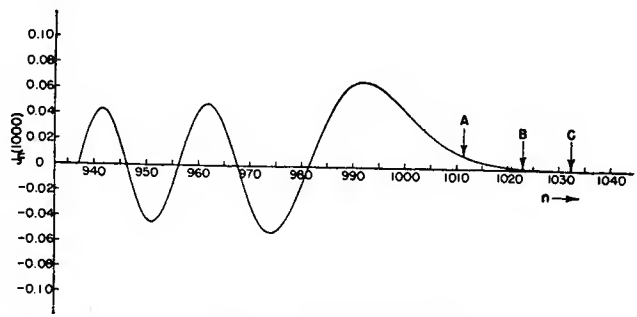


Fig. 2—Graph of $J_n(1000)$.

[4] E. Jahnke and F. Emde, TABLES OF FUNCTIONS WITH FORMULAE AND CURVES, Dover Publications, New York, N. Y., 1943, p. 171.

[5] August Hund, FREQUENCY MODULATION, McGraw-Hill Book Co., New York, N. Y., 1942; Table VI, p. 352.

[6] Murlan S. Corrington, "Tables of Bessel Functions $J_n(1000)$," *Jour. Math. Phys.* (M.I.T.), vol. 24, pp. 144-147; November, 1945.

in the band just inside the frequencies which correspond to the ends of the swings. When the deviation increases and the modulating frequency remains constant, the total energy of the spectrum is spread over a greater bandwidth, and the average amplitude of the side frequencies must decrease uniformly to maintain constant total energy in the modulated carrier wave.

The absolute maximum value of the Bessel function, for positive values of m, is shown by Figure 3. For a given modulation index, this maximum occurs for a value of the order n slightly less than the modulation index m. For example, for a modulation index of 1000 the maximum occurs at $n = 991.91$ and equals 0.06756. If the modulation index is 10, the maximum occurs at $n = 8.23$ and equals 0.3210. The

Fig. 3—Absolute maximum value of Bessel function $J_n(m)$.

curve of Figure 3 shows this maximum value for a range of modulation indexes from 10 to 10,000. It was computed from the formulas of Meissel[7] which state that the Bessel function $J_n(k)$ reaches its absolute maximum

$$J_n(k) = \frac{0.6748\ 8509\ 6430}{\sqrt[3]{k}} + \frac{0.0727\ 6309\ 8182}{k} + \frac{0.0199\ 5975\ 0328}{\sqrt[3]{k^5}} + \cdots \qquad (3)$$

for the value

[7] E. Meissel, "Beitrag zur Theorie der Bessel'schen Functionen," *Astronom. Nach.*, vol. 128, cols. 435-438; 1891.

$$n = k - 0.8086\ 1651\ 7466\sqrt[3]{k} - \frac{0.0606\ 4998\ 7910}{\sqrt[3]{k}} - \frac{0.0316\ 7351\ 0263}{k} - \cdots . \qquad (4)$$

A family of curves for modulation indexes from one to twenty is shown by Figure 4. The vertical scale represents the amplitude of the given side frequency for each modulation index. The curve of Figure 1 can be obtained by cutting a section through the surface for a modulation index of 10. Contour line A is for the constant value of the Bessel function $J_n(D/\mu) = 0.01$. Similarly, the contour B corresponds

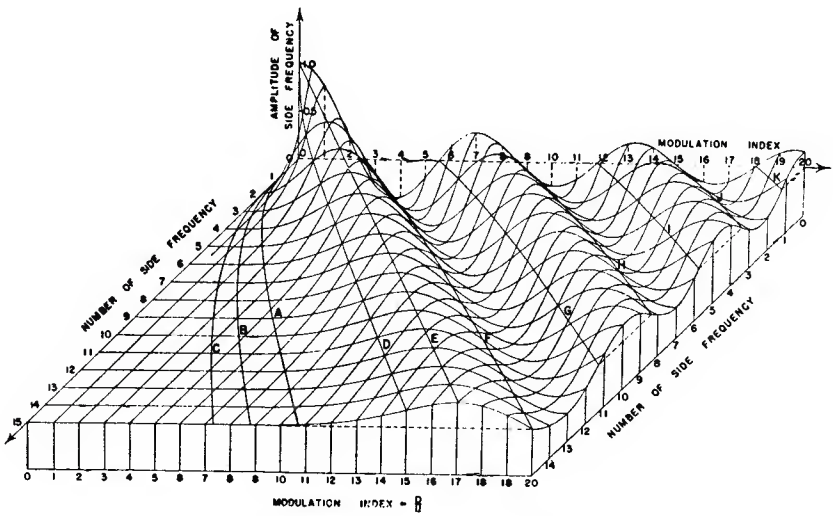


Fig. 4—Side-frequency amplitudes.

to $J_n(D/\mu) = 0.001$, and contour C is drawn for $J_n(D/\mu) = 0.0001$.

Curve D is shown for the order of the Bessel function equal to the argument. If the bandwidth of a frequency-modulated carrier wave were just equal to twice the deviation, the side frequencies would not extend beyond curve D. It is evident that for a given modulation index the bandwidth extends beyond curve D (say to curve A), but that the intensities of the side-frequencies beyond curve D are decreasing rapidly.

Curve E is drawn along the top of the first crest and gives the absolute maximum value of the envelope of the side frequencies for each modulation index. This curve is also given by Figure 3. The curves F,

G, H, I, J, and K show where the surface goes through zero, i.e., the zeroes of the Bessel functions.

Definition of Bandwidth

Theoretically, there is an infinite number of side frequencies in the spectrum of a frequency-modulated carrier wave, but the amplitudes decrease very rapidly beyond the last maximum. Point A on Figure 1 corresponds to the value of $J_n(10) = 0.01$ and will be defined as the edge of the band. Point B is shown for $J_n(10) = 0.001$, and point C for $J_n(10) = 0.0001$. These latter two points can be used for a more precise definition of bandwidth, but point A is to be taken as the usual limit for practical purposes.

The curves of Figure 5 show the variation of the bandwidth as the modulation index is changed.[8]

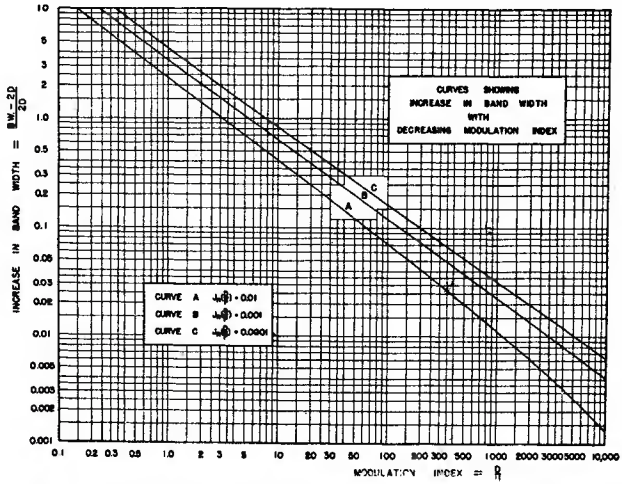


Fig. 5—Variation of bandwidth with modulation index.

[8]A rather simple method for computing the argument for $J_n(x) = 0.01$, 0.001, 0.0001 is to use the approximate formula

$$J_v(v \operatorname{sech} \alpha) = \frac{\tanh \alpha}{\pi\sqrt{3}} \exp\{v(\tanh \alpha + \tfrac{1}{3}\tanh^3 \alpha - \alpha)\} K_{1/3}\left(\frac{v}{3}\tanh^3 \alpha\right)$$

given in G. N. Watson, "A treatise on the theory of Bessel functions," The Macmillan Company, New York, N. Y., Second Edition, 1944, p. 250, where $K_{1/3}(x)$ is a modified Bessel function of the second kind of order $\frac{1}{3}$ and argument x. A series of values of α was chosen and the corresponding Bessel functions computed. These values were plotted on semilog paper and the arguments corresponding to the ordinates 0.01, 0.001, and 0.0001 were read off. The curves of Figure 5 were obtained directly from these readings.

Example: Let the deviation be ± 50 kilocycles and the audio frequency be 5 kilocycles. Find the bandwidth.

The modulation index is $50/5 = 10$. From curve A of Figure 5, the increase in bandwidth is approximately 0.42 or 42 per cent, so the bandwidth is approximately 2 (50) (1 + 0.42) = 142 kilocycles.

Bandwidth Required for Complex Modulation

If several modulating tones are present simultaneously, the carrier wave can be expressed as

$$e = E \sin \left\{ \omega t + \sum_{s=1}^{S} \frac{D_s}{\mu_s} \sin (2\pi\mu_s t + \varepsilon_s) \right\} \tag{5}$$

where E is the amplitude of the wave, ω is the carrier angular frequency, D_s is the deviation corresponding to the audio frequency μ_s, t is the time, and ε_s is the phase angle corresponding to μ_s. This modulated carrier wave can be represented by a spectrum[3, 9–11]

$$e = E \sum_{k_s=-\infty}^{\infty} \left\{ \prod_{s=1}^{S} J_{k_s} (m_s) \right\} \sin \left(\omega t + \sum_{s=1}^{S} k_s \theta_s \right) \tag{6}$$

where

$$m_s = D_s/\mu_s \quad \text{and} \quad \theta_s = 2\pi\mu_s t + \varepsilon_s.$$

In the case of two-tone modulation this becomes[1]

$$E \sin \{\omega t + D_1/\mu_1 \sin 2\pi\mu_1 t + D_2/\mu_2 \sin 2\pi\mu_2 t\}$$

$$= E \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} J_m (D_1/\mu_1) J_n (D_2/\mu_2) \sin (\omega + 2\pi m\mu_1 + 2\pi n\mu_2)\, t. \tag{7}$$

This result shows that the spectrum is now much more complicated than for a single modulating tone, and that side frequencies will be produced at spacings from the carrier given by all the possible combinations $\pm m\mu_1 \pm n\mu_2$. The amplitude of each side frequency will be proportional to the product of the two Bessel functions. Just as the

[9] E. C. Cherry and R. S. Rivlin, "Non-linear Distortion, with Particular Reference to the Theory of Frequency Modulated Waves, Part I," *Phil. Mag.*, vol. 32, pp. 265-281; October, 1941.

[10] A. S. Gladwin, "Energy Distribution in the Spectrum of a Frequency Modulated Wave, Part I," *Phil. Mag.*, vol. 35, pp. 787-802; December, 1944.

[11] K. R. Sturley, "Frequency Modulation," *Jour. I.E.E.* (London), vol. 92, Part III, pp. 197-218, September, 1945.

maximum deviation occurs when D_1 and D_2 are in phase, the maximum bandwidth is given approximately by the sum of the two bandwidths that would be obtained with the two modulating tones used one at a time.

The graph of Figure 6 shows the spectrum obtained when two tones are present simultaneously, in accord with (7). The side frequencies are no longer symmetrical about the carrier and, when they are separated from the carrier by a frequency greater than $D_1 + D_2$, decrease rapidly toward zero. The upper sideband contains 57.9 per cent of the power, the lower sideband 42.0 per cent, and the carrier 0.1 per cent.

General Method for Computing Side-Frequency Amplitudes

If the modulating signal is given, the variations of the phase angle

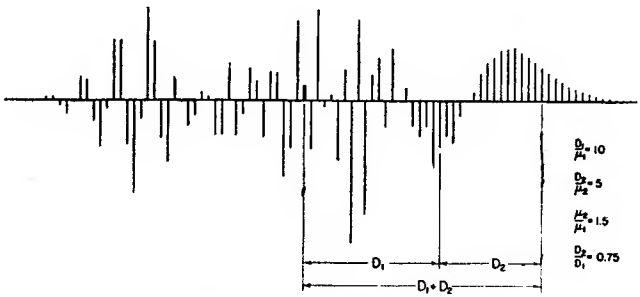


Fig. 6—Spectrum for complex modulation.

will be proportional to the integral of the signal. This integration can be done directly, or by numerical integration, and the constant of integration should be chosen so the average value of the phase angle, over a complete cycle, is zero. If the phase angle is $S(t)$ the frequency-modulated carrier wave can be expressed as

$$\begin{aligned} e &= E \sin \{\omega t + S(t)\} \\ &= E \sin \omega t \cos S(t) + E \cos \omega t \sin S(t). \end{aligned} \tag{8}$$

Expand in the Fourier series

$$\cos S(t) = \sum_{n=0}^{\infty} \{a_n \sin n\theta + b_n \cos n\theta\} \tag{9}$$

$$\sin S(t) = \sum_{n=0}^{\infty} \{c_n \sin n\theta + d_n \cos n\theta\} \tag{10}$$

where $\theta = 2\pi\mu t$ and μ is the repetition rate of the signal in cycles per second. This expansion can be done by direct integration of the integrals for the Fourier coefficients or by one of the numerical methods for harmonic analysis.[12, 13]

Then

$$e = E \sin \omega t \sum_{n=0}^{\infty} \{a_n \sin n\theta + b_n \cos n\theta\}$$

$$+ E \cos \omega t \sum_{n=0}^{\infty} \{c_n \sin n\theta + d_n \cos n\theta\}$$

$$= E \sum_{n=0}^{\infty} \{\tfrac{1}{2}(a_n + d_n) \cos(\omega - 2\pi n\mu)t - \tfrac{1}{2}(a_n - d_n) \cos(\omega + 2\pi n\mu)t$$

$$+ \tfrac{1}{2}(b_n - c_n) \sin(\omega - 2\pi n\mu)t + \tfrac{1}{2}(b_n + c_n) \sin(\omega + 2\pi n\mu)t\} \quad (11)$$

which gives the side-frequency amplitudes directly.

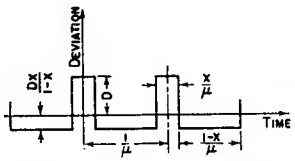


Fig. 7—Modulating signal.

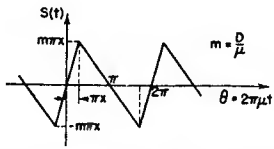


Fig. 8—Variation of phase angle.

The results of numerical computation can be checked by taking the sum of the squares of the carrier and each of the side-frequency amplitudes; they should add up to E^2.

The Spectrum of a Carrier Wave which is Frequency-Modulated with a Rectangular Signal

When the signal is a rectangular or square wave, as in frequency-modulated telegraphy, or in television video and synchronizing signals, the carrier wave can be analyzed into a spectrum in a similar manner. If the modulating signal is as shown by Figure 7, the phase angle $S(t)$ will be 2π times the integral of this curve, as shown by Figure 8,

[12] C. Runge and H. König, VORLESUNGEN ÜBER NUMERISCHES RECHNEN, Julius Springer, Berlin, 1924, pp. 211-231.

[13] R. P. G. Denman, "36 and 72 Ordinate Schedules for General Harmonic Analysis," *Electronics*, vol. 15, pp. 44-47, September, 1942. In addition to the corrections listed in vol. 16, pp. 214-215, April, 1943, change the correction in column one, p. 215, from "Column for B_{11} and B_{22}, line for $\alpha = 20°$, for $\triangle_2$, read $\triangle 2$," to read "... for $\triangle_2$, read $-\triangle_2$."

where $m = D/\mu$. The equation for the frequency-modulated wave becomes

$$e = E \sin \{\omega t + S(t)\} = E \sin \omega t \cos S(t) + E \cos \omega t \sin S(t). \tag{12}$$

Since $\cos S(t)$ is symmetrical about the origin, it can be expanded in a cosine series. Similarly, $\sin S(t)$ is skew symmetric about the origin and can be expanded in a sine series. From Figure 8,

$$\begin{aligned} \cos S(t) &= \cos m\theta && 0 \leqq \theta \leqq \pi x \\ &= \cos \frac{mx\ (\pi - \theta)}{1 - x} && \pi x \leqq \theta \leqq \pi \\ &= \sum_{n=0}^{\infty} b_n \cos 2\pi n \mu t \end{aligned} \tag{13}$$

where

$$b_n = \frac{2}{\pi} \int_0^{\pi x} \cos m\theta \cos n\theta d\theta + \frac{2}{\pi} \int_{\pi x}^{\pi} \cos \frac{mx}{1 - x} (\pi - \theta) \cos n\theta d\theta \tag{14}$$

$$b_0 = \frac{1}{\pi m x} \sin \pi m x. \tag{15}$$

Similarly,

$$\begin{aligned} \sin S(t) &= \sin m\theta && 0 \leqq \theta \leqq \pi x \\ &= \sin \frac{mx\ (\pi - \theta)}{1 - x} && \pi x \leqq \theta \leqq \pi \\ &= \sum_{n=0}^{\infty} c_n \sin 2\pi n \mu t \end{aligned} \tag{16}$$

where

$$c_n = \frac{2}{\pi} \int_0^{\pi x} \sin m\theta \sin n\theta d\theta + \frac{2}{\pi} \int_{\pi x}^{\pi} \sin \frac{mx}{1 - x} (\pi - \theta) \sin n\theta d\theta. \tag{17}$$

When these results are substituted into (11), the spectrum is given by

$$e = \sum_{n=-\infty}^{\infty} \frac{mE}{\pi(m-n)(mx-nx+n)} \sin \pi x(m-n) \cdot \sin(\omega + 2\pi n\mu)t \qquad (18)$$

where

E = amplitude of the wave

m = modulation index $= D/\mu$

D = maximum frequency deviation, cycles per second

μ = repetition rate, cycles per second

x = fraction of the time the frequency is at the extreme deviation D

ω = angular frequency of the carrier, radians per second

t = time in seconds.

When $n = 0$, the carrier amplitude is given by

$$\text{Carrier} = \frac{E}{\pi mx} \sin m\pi x \sin \omega t. \qquad (19)$$

The side frequencies adjacent to the carrier are given by $n = \pm 1$ and are separated from the carrier by an amount equal to the audio frequency. They are given by:
First upper side frequency

$$= \frac{mE}{\pi(m-1)(mx-x+1)} \sin \pi x(m-1) \sin(\omega + 2\pi\mu)t. \qquad (20)$$

First lower side frequency

$$= \frac{mE}{\pi(m+1)(mx+x-1)} \sin \pi x(m+1) \sin(\omega - 2\pi\mu)t. \qquad (21)$$

The other side frequencies can be determined by assigning appropriate values to n in (18).

The indeterminate cases must be evaluated separately:

Case I $\qquad m = n, \quad \frac{1}{2}(b_n + c_n) = x \qquad (22)$

Case II $\qquad \frac{mx}{1-x} = -n, \ \frac{1}{2}(b_n + c_n) = \frac{1}{\pi(m-n)} \sin \pi x(m-n) + (-)^n(1-x) \qquad (23)$

Case III $\quad m = -n, \ \frac{1}{2}(b_n - c_n) = x \qquad (24)$

Case IV $\quad \dfrac{mx}{1-x} = n, \quad \frac{1}{2}(b_n - c_n) = \dfrac{1}{\pi(m+n)} \sin \pi x(m+n) + (-)^n(1-x). \qquad (25)$

The case of square-wave modulation is obtained by setting $x = \frac{1}{2}$. This gives the result

$$e = \sum_{n=-\infty}^{\infty} \frac{2mE}{\pi(m^2-n^2)} \sin (m-n) \frac{\pi}{2} \sin (\omega + 2\pi n\mu)t$$

$$= \frac{2E}{\pi m} \sin m \frac{\pi}{2} \sin \omega t$$

$$+ \frac{2mE}{\pi(m^2-1^2)} \cos \frac{m\pi}{2} \{\sin (\omega - 2\pi\mu)t - \sin (\omega + 2\pi\mu)t\}$$

$$- \frac{2mE}{\pi(m^2-2^2)} \sin \frac{m\pi}{2} \{\sin (\omega - 4\pi\mu)t + \sin (\omega + 4\pi\mu)t\}$$

$$- \frac{2mE}{\pi(m^2-3^2)} \cos \frac{m\pi}{2} \{\sin (\omega - 6\pi\mu)t - \sin (\omega + 6\pi\mu)t\}$$

$$+ \cdots . \qquad (26)$$

This result, for $x = \frac{1}{2}$, agrees with that previously obtained by van der Pol.[14]

The limits for the amplitudes of the side frequencies can be determined from the coefficients of (18). Thus, if $m = D/\mu = 5$, and if $x = \frac{1}{4}$, the limit of the amplitudes becomes

$$\text{Amplitiude limit} = \frac{m}{\pi(m-n)(mx - nx + n)} = \frac{20}{\pi(5-n)(5+3n)} . \qquad (27)$$

This curve is shown by Figure 9. Actually, most of the side-frequency amplitudes will be less than this because of the first sinusoidal term of (18). As shown by Figure 10, the amplitudes oscillate within the limits of the curve of Figure 9. It may be easily seen that

[14] Balth. van der Pol, "Frequency Modulation," *Proc. I.R.E.*, vol. 18, pp. 1194-1205; July, 1930.

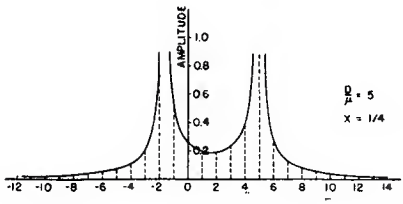


Fig. 9—Limits for side-frequency amplitudes.

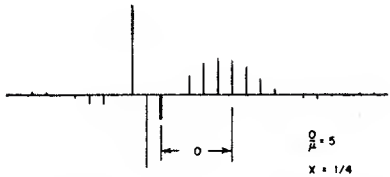


Fig. 10—Spectrum for rectangular modulation.

most of the energy of the spectrum is concentrated about the frequencies that correspond to the two limits of the deviation.

Bandwidth Required for Rectangular Modulation

Equation 18 shows that there is an infinite number of side frequencies in the spectrum of a frequency-modulated wave with rectangular modulation. As shown by Figures 9 and 10, the amplitudes of these side frequencies decrease uniformly beyond the limits of the deviations. If the edges of the band are defined as the points corresponding to a limiting amplitude of $0.01E$, the bandwidth can be computed directly from (18). For the case of square-wave modulation, $x = 0.5$, and the increase in bandwidth with decreasing modulation index will be as shown by Figure 11. If a more strict definition of bandwidth is required, curve B shows the width for the limiting amplitude $0.001E$. Curve A is an accurate enough limit for most practical cases.

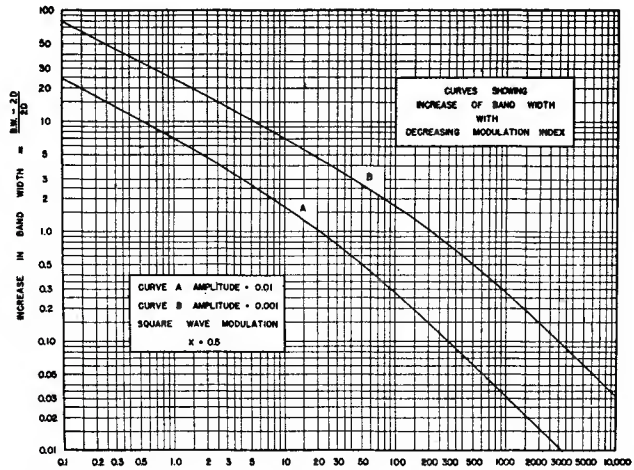


Fig. 11—Variation of bandwidth with modulation index.

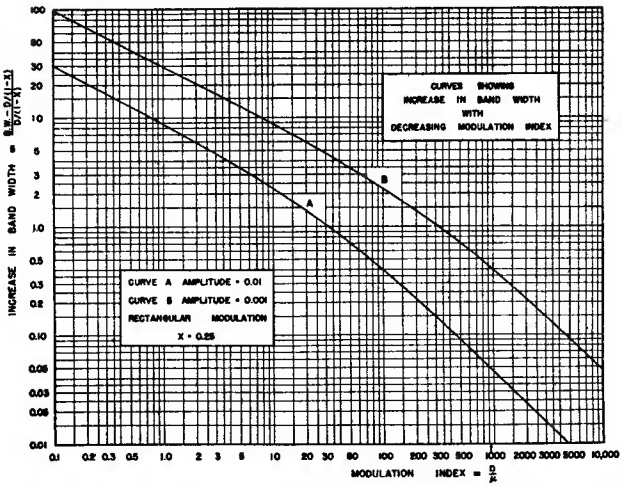


Fig. 12—Variation of bandwidth with modulation index.

If the maximum deviation is for one-fourth the time, $x = 0.25$, the curves of Figure 12 show the corresponding limits of the bandwidth. Other sets of curves, for other values of x, can be computed from (18).

It will be noted that the band does not end as abruptly with rectangular modulation as it did with sinusoidal modulation. The curves of Figures 11 and 12 are much farther apart than the corresponding curves of Figure 5.

The Spectrum of a Carrier Wave which is Frequency-Modulated with a Triangular Signal

When a uniformly spaced series of parallel bars, each one unit wide, is scanned at a uniform rate with a rectangular aperture of unit width, as shown by Figure 13, the resulting signal is proportional to the area of the bar covered by the aperture. The signal will have a triangular wave form, as shown by Figure 14. During the time the aperture is between the bars, the output will be constant. As the aperture starts to cover a bar, the output increases linearly until

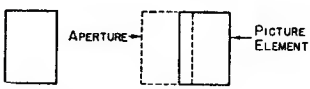


Fig. 13—Scanning of picture element.

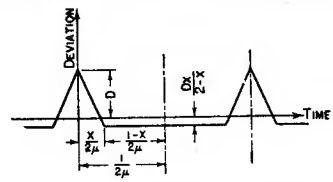


Fig. 14—Modulating signal.

the aperture covers the entire bar. As the aperture moves on, the signal decreases linearly until it reaches the previous constant value, and remains constant until the next bar is reached. If this wave form is used to modulate the frequency of a carrier wave, the variation of the phase angle will be 2π times the integral of this curve of Figure 14, as shown by Figure 15. The equation for this frequency-modulated wave becomes

$$e = E \sin \{\omega t + S(t)\} = E \sin \omega t \cos S(t) + E \cos \omega t \sin S(t) \qquad (28)$$

$$S(t) = \frac{D}{\mu}\left\{\frac{\pi x(2-x)\ \theta - \theta^2}{\pi x\ (2-x)}\right\} \qquad 0 \leqq \theta \leqq \pi x \qquad (29)$$

$$= \frac{D}{\mu}\left\{\frac{x\ (\pi - \theta)}{2-x}\right\} \qquad \pi x \leqq \theta \leqq \pi. \qquad (30)$$

When $S(t)$ is expanded in a Fourier series[15] and (11) is used, the amplitude of the nth side frequency

$$= \frac{1}{\sqrt{2\pi\alpha}} \cos \frac{\beta^2}{4\alpha}\left[\operatorname{sgn} \gamma\, C\left\{\frac{\gamma^2}{4\alpha}\right\} - \operatorname{sgn} \beta\, C\left\{\frac{\beta^2}{4\alpha}\right\}\right] + \frac{1}{\sqrt{2\pi\alpha}} \sin \frac{\beta^2}{4\alpha}$$

[15] The integrals can be evaluated by the following process:

$$\int_0^{\pi x} \cos\ (\alpha\theta^2 + \beta\theta)\, d\theta = \int_0^{\pi x} \cos \{\alpha(\theta + \beta/2\alpha)^2 - \beta^2/4\alpha\} d\theta$$

$$= \cos \frac{\beta^2}{4\alpha} \int_0^{\pi x} \cos \{\alpha(\theta + \beta/2\alpha)^2\} d\theta + \sin \frac{\beta^2}{4\alpha} \int_0^{\pi x} \sin \{\alpha(\theta + \beta/2\alpha)^2\} d\theta.$$

Let

$$\sqrt{\alpha}(\theta + \beta/2\alpha) = \pm \sqrt{v} \text{ and } \sqrt{\alpha}\, d\theta = \pm \frac{dv}{2\sqrt{v}}.$$

Then

$$\int_0^{\pi x} \cos \{\alpha(\theta + \beta/2\alpha)^2\} d\theta = \sqrt{\frac{\pi}{2\alpha}} \int_{v_1}^{v_2} \frac{\cos v\, dv}{\sqrt{2\pi v}}$$

where

$$v_1 = \operatorname{sgn} \beta \frac{\beta^2}{4\alpha}; \qquad v_2 = \operatorname{sgn} \gamma \frac{\gamma^2}{4\alpha}.$$

The same transformation can be used on the second integral.

$$\left[\operatorname{sgn} \gamma \, S\left\{ \frac{\gamma^2}{4\alpha} \right\} - \operatorname{sgn} \beta \, S \left\{ \frac{\beta^2}{4\alpha} \right. \right. \qquad \left. \left. \frac{1}{\pi\gamma} \sin (\pi\gamma x + \epsilon) \right. \right. \tag{31}$$

and the amplitude of carrier

$$= \frac{2}{\sqrt{2\pi\alpha}} \cos \frac{\beta^2}{4\alpha} \left[C \left\{ \frac{\alpha\pi^2 x^4}{4} \right\} + C \left\{ \frac{\beta^2}{4\alpha} \right\} \right] + \frac{2}{\sqrt{2\pi\alpha}} \sin \frac{\beta^2}{4\alpha}$$

$$\left[S \left\{ \frac{\alpha\pi^2 x^4}{4} \right\} + S \left\{ \frac{\beta^2}{4\alpha} \right\} \right] + \frac{2}{\epsilon} \sin (\pi\gamma x + \epsilon) \tag{32}$$

where

$$\alpha = \frac{m}{\pi x (2 - x)} \qquad \gamma = \frac{mx}{2 - x} + n$$

$$\beta = n - m \qquad \epsilon = \frac{-\pi m x}{2 - x}.$$

sgn β means the algebraic sign of β, and the C and S functions are the Fresnel integrals

$$C(z) = \frac{1}{\sqrt{2\pi}} \int_0^z \frac{\cos t \, dt}{\sqrt{t}} = \frac{1}{2} \int_0^z J_{-1/2}(t) \, dt \tag{33}$$

$$S(z) = \frac{1}{\sqrt{2\pi}} \int_0^z \frac{\sin t \, dt}{\sqrt{t}} = \frac{1}{2} \int_0^z J_{1/2}(t) \, dt. \tag{34}$$

These integrals are tabulated over a considerable range.[16, 17]

The vertical lines of Figure 16 show the spectrum for triangular modulation with a modulation index of 10. The dotted line is the Bessel function $J_n(10)$; it gives the amplitudes of the side frequencies for the corresponding sine-wave signal. During triangular modulation,

[16] See Table V, pp. 744-745, of footnote reference 8. Tables of $C(x)$ and $S(x)$, $x = 0.02(0.02)1.00$; 7*D*, and $x = 0.5(0.5)50.0$; 6*D*. For list of errors, see J. W. Wrench, Jr., "Mathematical tables—Errata," MATHEMATICAL TABLES AND OTHER AIDS TO COMPUTATION, VOL. 1, pp. 366-367; January, 1945.

[17] J. R. Airey, Sec'y, FRESNEL'S INTEGRALS, $S(x)$ AND $C(x)$, British Association for the Advancement of Science, Report of the Ninety-fourth Meeting, 1926, pp. 273-275. Tables of $C(x)$ and $S(x)$, $x = 0.0(0.1)20.0$; 6*D*.

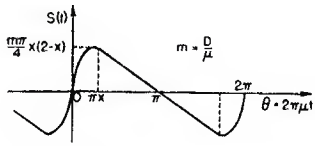


Fig. 15—Variation of phase angle.

$x = 1$, the frequency varies linearly from one extreme of the frequency excursion to the other, while for sinusoidal modulation the frequency is near the extremes of frequency a greater portion of the time. As might be expected, more of the energy in the spectrum is near the ends of the swing for sine-wave modulation than for triangular modulation.

Bandwidth Required for Triangular Modulation

If the bandwidth is defined as the extremes of frequency beyond which none of the side-frequency amplitudes are greater than 1 per cent of the carrier amplitude that would be obtained if the modulation were removed, the variation of bandwidth with modulation index can be computed from the equations for the side-frequency amplitudes. Curve A of Figure 17 shows how the bandwidth increases as the repetition rate is increased. For a more precise definition of bandwidth, either curve B or curve C can be used.

If x is reduced to 0.1, the signal becomes a series of triangular pulses with blank spaces between. Most of the sideband energy will occur near the frequency which the carrier wave has between pulses, but the pulses will cause energy to be distributed on both sides of this frequency. Figure 18 shows the spectrum for a modulation index of 10. The amplitudes decrease much more slowly than in the case of triangular or sinusoidal modulation.

Conclusions

When a carrier wave is modulated in frequency, an infinite number of side frequencies is produced. As the modulation index is changed,

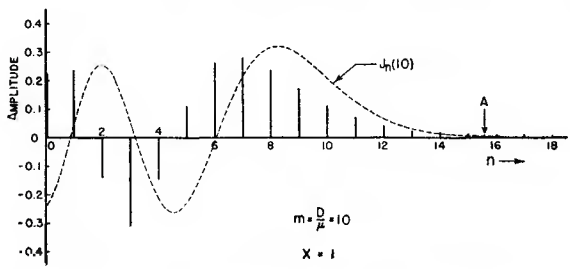


Fig. 16—Spectrum for triangular modulation.

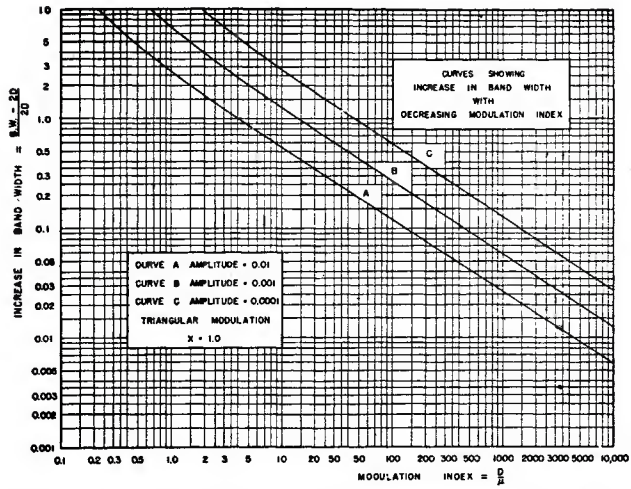


Fig. 17—Variation of bandwidth with modulation index.

the amplitudes of the side frequencies change and the carrier is likewise reduced and may even become zero. Although the bandwidth is theoretically infinite, in practice the side frequencies gradually decrease in amplitude for frequencies beyond the extremes of the total frequency excursions. The bandwidth can be defined as the extremes of frequency beyond which none of the side-frequency amplitudes are greater than 1 per cent of the carrier voltage obtained when the modulation is removed.

The bandwidth so defined always exceeds the total frequency excursion, but is nevertheless limited. For large modulation indexes, i.e., the deviation much greater than the repetition rate, the bandwidth approaches the actual variation in frequency and is only slightly greater. For small modulation indexes, the bandwidth may be several

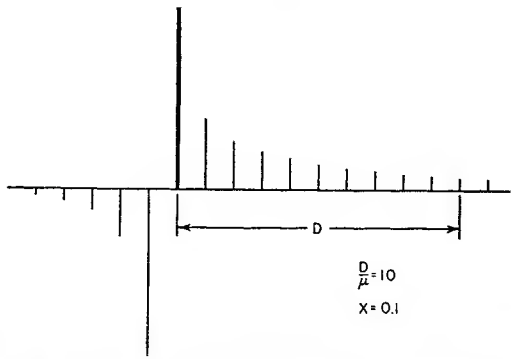


Fig. 18—Spectrum for triangular pulse modulation.

times the actual frequency excursion. Curves are given to show the bandwidth for modulation indexes from 0.1 to 10,000 for sinusoidal, square, rectangular, and triangular modulation. For a more precise definition of bandwidth, curves are also given for amplitude limits of $0.001E$ and $0.0001E$.

When several modulating tones are on simultaneously, the side frequencies are produced at frequencies separated from the carrier by all combination frequencies that can be obtained by taking sums and differences of all the harmonics of the tone frequencies. The same curves can be used to determine the bandwidth when several audio tones are used simultaneously, since the bandwidth will be equal approximately to the sum of the bandwidths for each tone separately.

GENERALIZED THEORY OF MULTITONE AMPLITUDE AND FREQUENCY MODULATION*†

By

L. J. GIACOLETTO

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

The frequeney spectrum produeed by single-tone, two-tone, and multitone modulating signals in the ease of amplitude modulation, frequeney modulation, and eombined amplitude and frequeney modulation are studied in turn. Amplitude, symmetry, and energy of sidebands for the different cases are eonsidered. Extensive eomputations are made of frequeney speetrums, and these are eompared with aetual frequency spectrums observed by means of a spectrum analyzer. The agreement between eomputed and measured frequency speetrums is very elose.

(14 pages; 21 figures)

---

* Decimal Classification: R148.1 × R148.2.

† *Proe. I.R.E.*, July, 1947.

FREQUENCY MODULATION PROPAGATION CHARACTERISTICS*†

By

Murray G. Crosby

RCA Communications, Inc., Riverhead, N. Y.

Summary—*Early work on frequency modulation is described wherein the propagation characteristics of frequency modulation were determined for frequencies between 9000 and 18,000 kilocycles. Oscilloscopic wave form and aural program observations, taken on a circuit between California and New York, showed that frequency modulation is much more distorted by the effects of multipath transmission than is amplitude modulation. The distortion is greatest at the lower modulation frequencies and higher depths of modulation where the side frequencies are most numerous.*

Oscilloscopic observations of the Lissajou figures formed by placing the outputs of receivers connected to spaced antennas on opposite oscilloscope plates showed that the diversity characteristics of frequency modulation are similar to those of amplitude modulation. That is, the detected outputs of the receivers tend to remain in phase for the lower modulation frequencies and become more phase random as the modulation frequency is increased. However, this tendency is almost obliterated on the lower modulation frequencies of frequency modulation by the presence of unequal harmonic distortion in the two receiver outputs.

Theory is given analyzing the distortion encountered in a two-path transmission medium under various path amplitude and phase relation conditions. The theory explains phenomena observed in the tests and points out the extreme distortion that can be encountered.

Introduction

IN THE annals of radio, most of the work done on frequency modulation has been of a theoretical nature[1,2,3] and very little has been offered as a result of any experimental development.

Considerable theory has also been presented in the consideration of this type of modulation as a defect of amplitude modulation.[3,4,5,6,7] A

---

* Decimal classification: R630.11.

† Reprinted from Proc. I.R.E., June, (1936).

[1] John R. Carson, "Notes on the theory of modulation," Proc. I.R.E., vol. 10, pp. 57-64; February, (1922).

[2] Balth van der Pol, "Frequency modulation," Proc. I.R.E., vol. 18, pp. 1194-1205; July, (1930).

[3] Hans Roder, "Amplitude, phase, and frequency modulation," Proc. I.R.E., vol. 19, pp. 2145-2176; December, (1931).

[4] R. Bown, D. K. Martin, and R. K. Potter, "Some studies in radio broadcast transmission," Proc. I.R.E., vol. 14, pp. 57-131; February, (1926).

[5] R. K. Potter, "Transmission characteristics of a short-wave telephone circuit," Proc. I.R.E., vol. 18, pp. 633-648; April, (1930).

[6] J. C. Schelleng, "Some problems in short-wave telephone transmission," Proc. I.R.E., vol. 18, pp. 933-937; June, (1930).

[7] T. L. Eckersley, "Frequency modulation and distortion," *Experimental Wireless and The Wireless Engineer*, vol. 7, pp. 482-484; September, (1930).

more recent contribution[8] includes an experimental investigation checking theory on the detection products produced in a frequency modulation receiver. It is the purpose of this paper to report on development work undertaken by the engineering department of R.C.A. Communications, Inc., wherein the propagation characteristics of frequency modulation were determined for a circuit between California and New York.

Early work done on frequency modulation within the Radio Corporation was that of H. O. Peterson[9] in which a laboratory frequency modulation circuit was set up. Later, in an attempt to reduce fading by frequency diversity, one of the telegraph transmitters on the Argentina and Brazil circuits was frequency modulated. Finally a definite development program was undertaken consisting of the development of a frequency modulation transmitter at the Rocky Point transmitting research and design laboratories and the development of a suitable receiver at the Riverhead receiving research and design laboratories.

After the transmitter and receiver were developed to a sufficient stage to permit a long-distance test to determine the propagation characteristics of frequency modulation, the transmitter was shipped to Bolinas, California, where it was operated by J. W. Conklin who had assisted in its development. The transmissions were observed at the Riverhead station on the receivers developed by the author. The tests were carried out in 1931 during the months from March to June, inclusive.

Propagation Tests

In the tests carried on between Bolinas and Riverhead, two major problems presented themselves for solution. First, how would frequency modulation withstand the ravages of fading. Second, could the detected outputs of two frequency-modulation receivers, fed by spaced antennas, be added directly[10] in a diversity[11] receiving system; that is, would the detected audio outputs remain in phase so that they could be combined directly, or would some kind of diversity antenna choosing device, as is in present use, be necessary.

It is interesting to note that in this article Eckersley makes the following prediction from his theory which agrees with the results obtained in the work described in this paper: "It will easily be seen that this results in most appalling distortion, and renders futile any attempt to use even pure frequency modulation in any transmission where appreciable echo delays (of the order of two or three milli-seconds) are present."

[8] J. G. Chaffee, "The detection of frequency modulated waves," Proc. I.R.E., vol. 23, pp. 517-540; May, (1935).

[9] H. O. Peterson, U. S. Patent No. 1,789,371.

[10] C. W. Hansell, U. S. Patent No. 1,803,504.

[11] H. O. Peterson, H. H. Beverage, and J. B. Moore, "Diversity telephone receiving system of R.C.A. Communications, Inc.," Proc. I.R.E., vol. 19, pp. 562-584; April, (1931).

In order to solve the problem of the effect of fading on frequency modulation, oscilloscopic wave form and aural program observations were made on the tone and program modulation after it had traversed the radio circuit from California to New York. In this manner the instantaneous harmonic distortion could be observed.

The diversity problem was answered by applying the outputs of two receivers, fed by spaced antennas, to opposite oscilloscope plates. Hence, by proper interpretation of the Lissajous figures obtained, the relative phases and amplitudes of the two outputs could be determined.

The Transmitter

To obtain the frequency modulation, a circuit of the type shown in Fig. 1 was employed. This type of modulator was used as the master

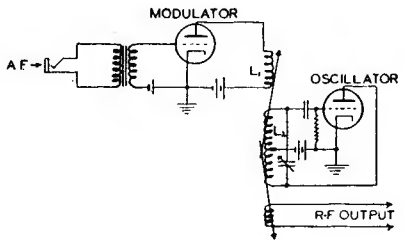


Fig. 1—Circuit diagram of frequency modulated master oscillator.

oscillator to replace the crystal oscillator normally used in a standard R.C.A. Communications telegraph transmitter. Special precautions were taken to insure constant element voltages for mean frequency stability. For the same reason, a one-hour warm-up period was observed before each transmission to allow the tube temperatures to settle and prevent drifting. By modulating at master oscillator frequency a frequency multiplication of from eight to sixteen was employed depending upon the radiated frequency. Hence, the frequency deviation, with modulation, was multiplied by a factor of from eight to sixteen and the frequency deviation applied to the oscillator tube was therefore, one eighth to one sixteenth of that radiated.

The operation of the modulator of Fig. 1 depends upon the variation of the effective inductance of the oscillator tuned circuit inductance, L_2, by virtue of the variation of the modulator tube plate resistance connected across coil L_1 which is coupled to the oscillator inductance, L_2. The modulator tube is biased at such a point that as the modulation is applied at its grid, its plate resistance will vary in accordance with the modulation. By a proper choice of the modulator bias and the inductance of L_1, together with proper adjustment of the

coupling between L_1 and L_2, a linear frequency modulation may be obtained.

During the frequency modulation transmission the transmitter amplifiers and frequency doublers were operated class C throughout so as to limit off any amplitude modulation introduced in the modulator or otherwise.

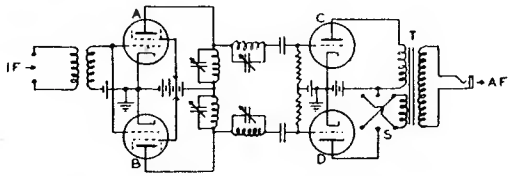


Fig. 2—Circuit diagram of frequency modulation receiver showing the conversion filters and detectors.

An amplitude modulating stage was included in the transmitter so that the transmitter could be adjusted for amplitude modulation and direct comparison could be made between frequency modulation and amplitude modulation.

The Receiver

A schematic diagram of the frequency modulation receiver is shown in Fig. 2. Energy at the superheterodyne intermediate frequency is fed to the two coupling tubes, *A* and *B*, having the conversion filters for converting the frequency modulation into amplitude modulation in

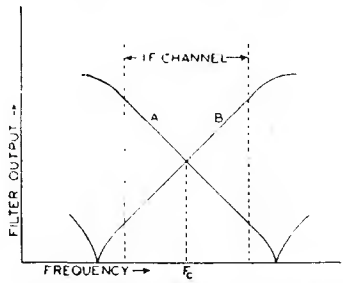


Fig. 3—Characteristics of the conversion filters used to convert frequency modulation into amplitude modulation for detection.

their plate circuits. The characteristics of these two filters are given by the curves *A* and *B* in Fig. 3. The carrier frequency F_c, is tuned to the middle of the linear portion of the sloping characteristic formed by the two tuned circuits connected as shown. By adjusting the tuned circuits so that the points of maximum and minimum output fall outside of the intermediate-frequency channel, only the linear portion of the characteristics are utilized. The outputs of these filters are fed to the detectors

C and *D* of Fig. 2 whose audio outputs are combined by the transformer *T* in push-pull or parallel combination, depending upon the switch *S*. With the switch in the push-pull position, frequency modulation may be received and amplitude modulation balanced out. The balance obtainable on amplitude modulation, however, is complete only for the condition of no frequency modulation present. As soon as the carrier is frequency deviated from its unmodulated position, the carriers fed to the detectors are no longer balanced due to the fact that the sloping filters have opposite slopes. Hence the balance is thrown off by an amount proportional to the frequency deviation from the unmodulated carrier position, and the resultant balance is an average between the balanced and unbalanced conditions. With the switch in the push-pull position, second harmonic square-law detector distortion is completely

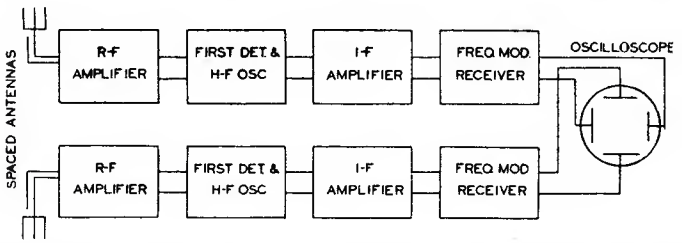


Fig. 4—Block diagram of receiver arrangement used in determining the phase relation between the detected outputs from receivers connected to spaced antennas.

balanced out. With the switch in the parallel position amplitude modulation may be received and frequency modulation balanced out.

The intermediate-frequency amplifier of the receiver was adapted from broadcast components and had a twelve-kilocycle pass band. Consequently the maximum frequency deviation was limited to one half of this intermediate frequency channel or six kilocycles. An experimental harmonic analysis of the receiver showed that it was capable of this amount of deviation without serious distortion.

An amplitude limiter was arranged so that it could be switched in and out of the intermediate-frequency circuits. This limiter consisted of a separate multistage intermediate-frequency amplifier feeding a tube with lowered element voltages so that the tube would be heavily overloaded and its output would be constant regardless of the changes of the input. Thus amplitude modulation could be removed. This method of removing amplitude modulation has an advantage over the balance obtained with the opposite sloping conversion filters in that the amplitude modulation is completely removed in the presence of frequency modulation as well as in the unmodulated condition.

For the diversity experiments, an arrangement as shown in Fig. 4 was used. Two complete frequency modulation receivers were fed by two harmonic wire antennas spaced 450 feet apart. The detected outputs of the tone modulation were fed to the opposite oscilloscope plates. Hence, by noting whether the oscilloscope diagram was a line in the second and fourth quadrants, a circle, or a line in the first and third quadrants, the relative phase relation of the two detected outputs could be determined as being in phase, ninety degrees out of phase, or 180 degrees out of phase, respectively. (See Fig. 5a, b, c, d, and e.)

For the wave form observations, the output of a single receiver was applied to one set of oscilloscope plates and a timing voltage to the other. Quality distortion was also checked by the aural observation of music which was transmitted by the frequency and amplitude modulation transmitters.

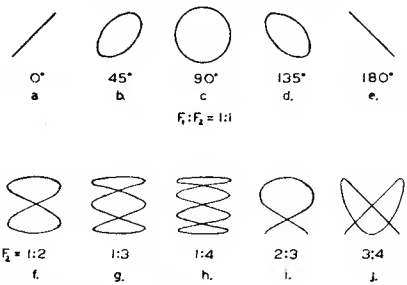


Fig. 5—Lissajou figures obtained with various phase relations and frequency ratios between voltages applied to opposite oscilloscope plates.

Wave Form Data

Day and night observations were made using music and tone modulation on frequencies of 18,020, 13,690, and 9010 kilocycles. The general conclusions derived were that frequency modulation is far more susceptible to the effects of selective fading than amplitude modulation. On music, the bass and lower notes were distorted and jumbled together while the violin and the other higher solo instruments came through fairly well. The intelligibility of speech was very low; in some cases a signal which gave fair intelligibility on amplitude modulation was practically unintelligible on frequency modulation. The distortion seemed to be proportional to the depth of modulation, since a sort of blasting effect was observed on the modulation peaks, and transmissions with lowered deviation came through better than those with higher deviation. The distortion also seemed to be greatest during fading minimums at which time "bursts" of distortion would appear; this condition was also present to a lesser extent on amplitude modulation.

The effect due to limiting seemed to be the indication that the distortion was caused by a change in the instantaneous frequency deviation of the wave and not due to an added amplitude modulation. This was evidenced by the fact that the limiter effectively removed all amplitude modulation introduced by fading, but the extreme distortion remained with little or no perceptible difference attributable to the limiter.

The general result of the use of different radiation frequencies over the circuit was to indicate that the quality of modulation was most impaired on the frequencies where the selective fading was most marked. Of the three frequencies observed, 18,020 kilocycles showed the least amount of selective fading, 9010 the greatest, and 13,690 a mean between the other two. Consequently, the quality of frequency modulation on 18,020 kilocycles was only occasionally inferior to that on amplitude modulation. On 9010 kilocycles there was scarcely an interval of good quality on the frequency modulation, and the amplitude modulation was also considerably distorted. 13,690 kilocycles was considered a representative frequency for the observations, since it presented periods of extremely poor quality along with periods of fairly good quality. Consequently, the oscillograms were taken on the 13,680-kilocycle radiations.

Fig. 6 gives the oscillograms showing the effects of distortion fading on wave form. From a comparison of the 300- and 1000-cycle oscillograms taken on frequency modulation, it is apparent that the amount of distortion on 300 cycles is much greater than that on 1000 cycles. At the lower modulation frequencies on frequency modulation, the harmonics were comparable to and stronger than the fundamental for a large percentage of the time. As the modulation frequency was increased, the amount of distortion decreased until at modulation frequencies above about 3000 cycles, little or no harmonic distortion was encountered. One of the reasons for this absence of harmonic distortion on the higher frequencies is probably the fact that the pass band of the receiver intermediate-frequency amplifier was twelve kilocycles and that of the audio amplifier six kilocycles. Hence, for modulation frequencies above 3000, the side bands above the first were eliminated by the intermediate-frequency filter and the harmonics were eliminated by the audio system. The reason for the excessive distortion at the lower modulation frequencies of frequency modulation is no doubt due to the effects of multipath transmission in disturbing the phase relations between the many side frequencies produced by these lower modulation frequencies.

Fig. 6J shows a wave form taken on amplitude modulation when

aural observations indicated that the carrier and not the side bands had faded. The presence of second harmonic is very evident. It has been conjectured that frequency modulation might be less susceptible to the effects of carrier fading because of the fact that the modulated energy is distributed among the side bands to a greater extent; consequently, the detected output is not as dependent for its value upon a single frequency in the transmission medium as is amplitude modulation. The observations did indicate that with frequency modulation,

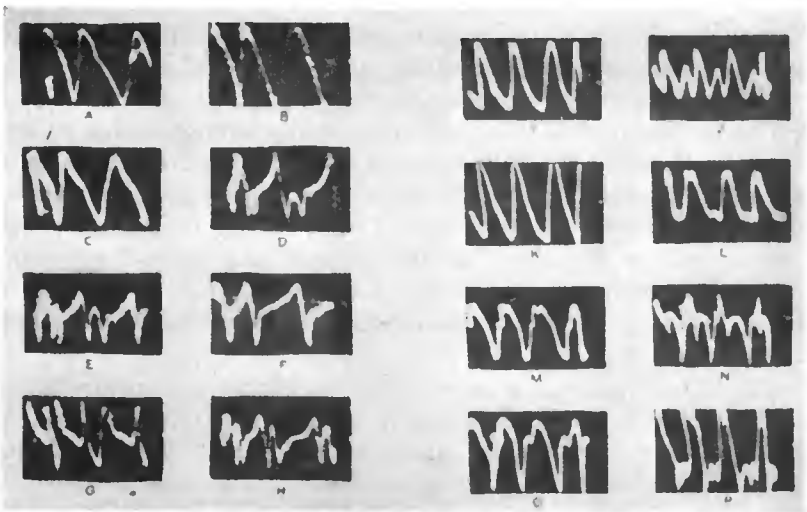


Fig. 6—Frequency and amplitude modulation wave form oscillograms showing the distortion due to fading.

A, B, and C. Amplitude modulation, 300 cycles.
D, E, F, G, and H. Frequency modulation, 300 cycles.
I, J, and K. Amplitude modulation, 1000 cycles.
L, M, N, O, and P. Frequency modulation, 1000 cycles.

the output was more constant; the quality of this output, however, prevented its constancy from being of any value for the purpose for which it was intended. This observation tends to indicate the advantage of systems wherein telegraph transmitters are frequency modulated to effect a frequency diversity to reduce fading.

Diversity Data

In general the diversity experiments indicated that on both frequency and amplitude modulation the lower modulation frequencies produced outputs from the two receivers which remained in phase. As the modulation frequency was increased, the tendency toward phase

differences became very marked until at a 5000-cycle modulation frequency the phase relation was usually random, varying occasionally a full 360 degrees.

The effect of radiation frequency on the phase relations between the two receiver outputs proved to be similar to the effect on wave form. That is, the phase distortion became less as the radiated frequency was increased. Thus on the 18,020-kilocycle radiations there were frequent periods when no phase distortion could be observed in

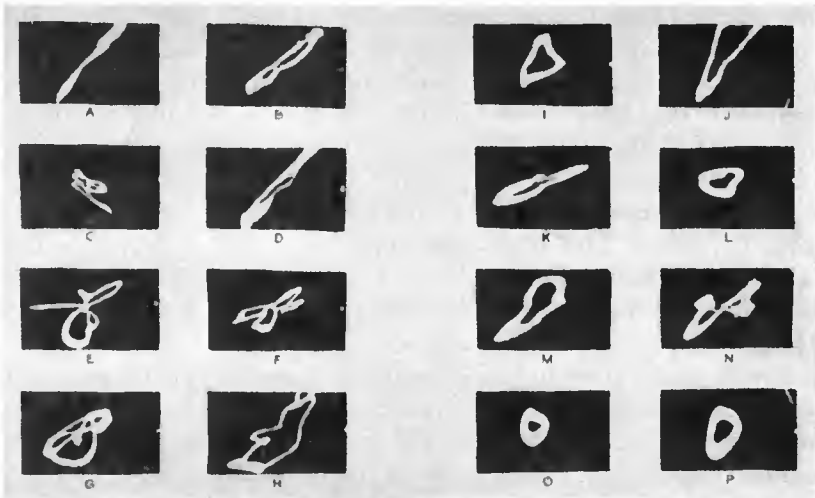


Fig. 7—Oscillograms showing the phase characteristics obtained on frequency and amplitude modulation using the arrangement of Fig. 4.

A, B, and D. Amplitude modulation, 300 cycles.
C, E, F, G, and H. Frequency modulation, 300 cycles.
I and J. Amplitude modulation, 1000 cycles.
K, L, M, and N. Frequency modulation, 1000 cycles.
O and P. Frequency modulation, 5000 cycles.

the audio range. On the 13,690-kilocycle radiations the distortion was usually less than ninety degrees for modulation frequencies below 1000 cycles, while for modulation frequencies from 1000 to 5000 cycles marked distortions were apparent. On the 9010-kilocycle radiations, the figures showed random phase variations on all modulation frequencies most of the time. The oscillograms were taken on 13,690 kilocycles with a frequency deviation of about 2000 cycles.

Fig. 7 shows the Lissajou figures obtained in the diversity setup. The diagrams for amplitude modulation given in Fig. 7A, B, and D for a modulation frequency of 300 cycles show how the two receiver outputs remain in phase on a low-frequency tone. The diagrams ob-

tained on frequeney modulation with a 300-cycle modulation frequency, given in Fig. 7C, E, F, G, and H, obtain their grotesque distortions from the presence of different harmonic distortions in the two receiver outputs. Thus one receiver might be receiving a signal producing a pure fundamental tone only, whereas the other reeeiver might be receiving a signal rich in harmonics; consequently, the diagram is distorted aceording to the difference in the amounts of harmonic content present at the instant of exposure. The effect produced by this unequal harmonic content on the two sets of oscilloscope plates can be seen from a study of the Lissajou figures given in Fig. 5 f, g, h, i, and j. Thus the diagram of Fig. 5 f might be formed by the eondition of fundamental tone from one receiver and second harmonic from the other, or the diagram of Fig. 5i by second harmonic from one receiver and third harmonic from the other. The rapid shift from one harmonic relation to another together with the presence of complex instead of simple wave forms on the two sets of plates makes possible erratic and grotesque patterns.[12] This effect was apparent to such a degree on frequency modulation that difficulty was experienced in determining the phase tendency of the fundamental tones. The effeet was only oeeasionally noticeable on amplitude modulation.

The diagrams given in Fig. 7 I and J for a modulation frequeney of 1000 cycles, show how the higher modulation frequencies on amplitude modulation are marked by phase differences which are not present on the lower tones. The higher tones on frequency modulation given in Fig. 7 K, L, M, and N for a modulation frequency of 1000 cycles, and in Fig. 7 O and P for a modulation frequency of 5000 cycles, also indicate marked phase differences and only a slight indication of different harmonic distortions. Amplitude and frequency modulation showed about the same amount of phase distortion on the higher modulation frequencies.

Theory

It is well known that the cause of fading distortion, taking place in the range of frequencies used in these tests, is due to multipath transmission in which the signal is conveyed to the receiving antenna by means of more than one path. These paths consist of refractions or reflections from the conducting layers of the ionosphere. Since these conducting layers vary in height and since the signal may follow a course consisting of various numbers of ricochets between the ionosphere and the earth, the signal traveling over the longer path is given

[12] A more complete set of these Lissajou figures may be found in the book, "High-Frequency Measurements," by A. Hund, pp. 71–75; McGraw-Hill, (1933).

a time lag with respect to that traveling over the shorter path. Hence the frequency modulated signal traveling over an earlier path may be expressed by

$$e = E_0 \sin\left(\omega t + \frac{F_d}{F_m}\cos pt\right) \tag{1}$$

where $\omega = 2\pi \times F_c$, F_c = carrier frequency, F_d = peak frequency deviation due to modulation, F_m = modulation frequency, and $p = 2\pi \times F_m$. The signal traveling over a later path may be expressed by

$$e = E_1 \sin\left\{\omega t + \beta + \frac{F_d}{F_m}\cos(pt + \alpha)\right\} \tag{2}$$

where β and α take into account the time delay imparted to the carrier wave and the modulation, respectively.

Considering only a two-path case first, the two waves given by (1) and (2) may be combined into a single resultant wave by means of the cosine law resulting in

$$e = \sqrt{E_0^2 + E_1^2 + 2E_0E_1\cos\left\{\frac{F_d}{F_m}\cos(pt+\alpha) + \beta - \frac{F_d}{F_m}\cos pt\right\}}$$

$$\sin\left[\omega t + \frac{F_d}{F_m}\cos pt + \tan^{-1}\frac{\sin\left\{\frac{F_d}{F_m}\cos(pt+\alpha)+\beta-\frac{F_d}{F_m}\cos pt\right\}}{\frac{E_0}{E_1}+\cos\left\{\frac{F_d}{F_m}\cos(pt+\alpha)+\beta-\frac{F_d}{F_m}\cos pt\right\}}\right] \tag{3}$$

Applying the sum and difference formula for the cosine, calling $E_0/E_1 = R$, and $2F_d/F_m \sin \alpha/2 = Z$, gives

$$e = \sqrt{E_0^2 + E_1^2 + 2E_0E_1\cos\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}$$

$$\sin\left[\omega t + \frac{F_d}{F_m}\cos pt - \tan^{-1}\frac{\sin\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}{R+\cos\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}\right] \tag{4}$$

Since the receiver either limits off or balances out the amplitude modulation on the signal, the amplitude term of (4) is reduced to a constant voltage and the only part requiring investigation is the angle or phase of the wave. To obtain the effective frequency deviation, the instantaneous frequency, or the rate of change of phase of (4) must be determined. This is given by

$$\frac{d\left[\omega t+\frac{F_d}{F_m}\cos pt-\tan^{-1}\frac{\sin\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}{R+\cos\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}\right]}{\frac{2\pi}{dt}}=f(\text{cycles})$$

$$=F_c+F_d\left[\cos pt-\frac{2\sin\frac{\alpha}{2}\cos\left(pt+\frac{\alpha}{2}\right)}{\frac{R+\cos\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}{\frac{1}{R}+\cos\left\{z\sin\left(pt+\frac{\alpha}{2}\right)-\beta\right\}}+1}\right] \tag{5}$$

When $R=1$, (5) reduces to:

$$f = F_c + F_d\left[=\cos pt - \sin\frac{\alpha}{2}\cos\left(pt+\frac{\alpha}{2}\right)\right] \tag{6}$$

which, by application of the cosine law gives

$$f = F_c + F_d\sqrt{1+\sin^2\frac{\alpha}{2}-\sin\alpha}\cos\left[pt-\tan^{-1}\frac{\sin^2\frac{\alpha}{2}}{1-\frac{\sin\alpha}{2}}\right] \tag{7}$$

However when $R=1$, the amplitude term of (4) must be considered since there is a possibility of the wave being 100 per cent amplitude modulated so that the limiter would be unable to hold its output constant and a signal modulated by noise would result. For the case where the amplitude does not go to zero, that is, for conditions in which the quantity $(2\ F_d/F_m \sin \alpha/2-\beta)$ does not pass through the points π, 3π, 5π, etc., the instantaneous frequeney of the wave applied to the frequency modulation sloping filters is given by (7). From this equation it can be seen that the resultant distortion under these conditions is merely a change in the effective frequency deviation of the modulation by a factor depending on α, and the addition of a phase angle depending upon α to the modulation frequeney. Thus, provided the amplitude does not go to zero, no harmonic distortion is encountered.

When R is large compared to one, (5) becomes:

$$f = F_c + F_d \cos pt - \frac{2 \sin \frac{\alpha}{2} F_d}{R} \cos\left(pt + \frac{\alpha}{2}\right) \cos\left\{z \sin\left(pt + \frac{\alpha}{2}\right) - \beta\right\}. \quad (8)$$

By an application of the addition formulas for the sine and cosine, together with the Bessel function expansions for cos (x sin ϕ) and sin (x sin ϕ) and the recurrence formulas for $J_n(x)$, (8) may be transformed into the following:

$$f = F_c + F_d \cos pt - \frac{2F_m \cos \frac{\alpha}{2}}{R}\left[\cos\beta\left\{\sum_{n=0}^{\infty}(2n+1)J_{2n+1}(z)\cos(2n+1)\left(pt+\frac{\alpha}{2}\right)\right\} + \sin\beta\left\{\sum_{n=1}^{\infty} 2nJ_{2n}\sin 2n\left(pt+\frac{\alpha}{2}\right)\right\}\right] \quad (9)$$

Thus when β is zero, a distortion consisting of fundamental and odd harmonics is added to the frequency deviation originally present. These harmonics are proportional to the ratio between the weaker and the stronger of the signals from the two paths, the modulation frequency, and $nJ_n(2F_d/F_m \sin \alpha/2)$ where n is the order of the harmonic. When β is ninety degrees, the distortion consists of even harmonics proportional to the same factors.

By far the most destructive distortion occurs when R is not large compared to unity, or is not equal to unity, but is slightly greater than unity. Fig. 8 shows wave forms calculated from the part of (5) in the square brackets. These wave forms are for a two-path case in which the ratio between the amplitudes arriving over the two paths is 1.2:1 Figs. 8(A), (B), and (C) show the effect of various phase differences between the two radio-frequency carriers with a constant phase difference of ninety degrees between the modulation frequencies modulating the two carriers and a constant ratio of F_d/F_m of ten. It is quite apparent that the distortion is greatest when the radio-frequency carriers are 180 degrees out of phase. This checks the observations that the distortion seemed to be greatest during the fading minimums. With the two radio-frequency carriers out of phase, the resultant unmodulated amplitude would be at a minimum, producing a fade in the signal. Figs. 8(C) and (D) show the effect of the value of F_d/F_m on the

wave form. The increase of the distortion with increase in F_d/F_m agrees with the observations that the lower modulation frequencies were the most distorted and that the distortion seemed to be proportional to the depth of modulation. Thus for modulation frequencies greater than one half the maximum modulation frequency where F_d/F_m can not be greater than two, a wave form similar to Fig. 8(D) would be

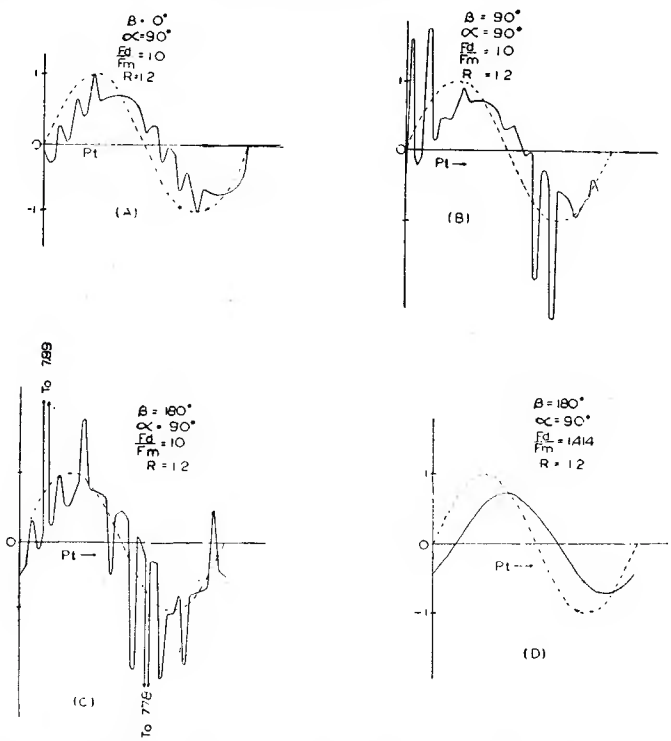


Fig. 8—Calculated wave forms for frequency modulation transmitted over a two-path medium plotted from (6). Ratio between path amplitudes, $R = 1.2$. F_d = frequency deviation. F_m = modulation frequency. α = phase difference between modulation frequencies. β = phase difference between carrier frequencies. The dotted line curve is a pure sine wave (representing no distortion) for comparison purposes.

obtained. For the lower modulation frequencies, for instance, 200 cycles, the wave form of Fig. 8(D) would exist for a deviation of 282.4 cycles and that of Fig. 8(C) for a deviation of 2000 cycles; hence the distortion increases with increase in depth of modulation.

From a study of (5), it can be seen that the effect of a change in the value of the phase difference, α, between the modulation frequencies on the two paths is approximately the same as a change in the value of

F_d/F_m since sin $\alpha/2$ enters in the quantity Z in the same manner as does the quantity F_d/F_m. However it can also be seen that the distortion is zero when α is equal to 0, 2π, 4π, etc. Thus the most severe distortion occurs when $\alpha=\pi$ radians or 180 degrees. Since $\alpha=2\pi DF_m$, where D = time delay between the two paths, for a given time delay sin $\alpha/2$ will go through maximums and minimums as the modulation frequency is varied. Hence bands of maximum and minimum distortion would be expected throughout the modulation frequency band.

By an application of the same theoretical method, the distortion due to three or more paths could be determined. The addition of a wave arriving over the third path to the resultant given by (4) would add another arc-tangent term to the phase of the resultant in which R would be equal to the ratio between the resultant amplitude of the first two paths and the amplitude of the third. This would add distortion which would result in a wave form undoubtedly more distorted than the two-path case. However, the degree of distortion would be proportional to the same parameters as the two-path case since the distortion term added by each path is similar to that produced by the first two paths.

The distortion encountered by amplitude modulation in multipath transmission has been theoretically considered in previous literature.[13] The general results of this theory are that the distortion consists of a change of the amplitude of the fundamental modulation frequency together with the introduction of second harmonic distortion. However, the higher order harmonic distortion so prone to appear in frequency modulation, and depending upon the ratio F_d/F_m, is not present. It is this extreme high order harmonic distortion which makes the reception of frequency modulation over a multipath medium far more distorted than the reception of amplitude modulation over the same medium.

Conclusions

The general conclusion derived from the tests and theory is that on circuits where multipath transmission is encountered, frequency modulation is impracticable. This, of course, precluded the use of frequency modulation on frequencies which are transmitted by refraction from the Kennelly-Heaviside layer. For the frequencies in the immediate vicinity of eighteen megacycles, where transmission is largely by a

[13] Charles B. Aiken, "Theory of the detection of two modulated waves by a linear rectifier," Proc. I.R.E., vol. 21, pp. 601–629; April, (1933). The "Case of Identical Modulating Frequencies" considered on page 616 covers exactly the same situation as the two-path multipath transmission case.

single path, and where echo is small when it is present,[14] fair success would be possible with a frequency modulation circuit where the modulation frequencies were not higher than about 5000 cycles. However, the logical place for frequency modulation is on the ultra-high frequencies where the only multipath transmission that exists[15] is that due to the direct ray and the ray reflected from the ground and other near-by objects. On the other hand, even this multipath transmission on ultra-short waves will introduce appreciable distortion to modulation frequencies of the order used in television. This distortion would be particularly severe in the case where the transmitter and receiver are both at rather high elevations, but would have the possibility of elimination by the use of directivity to discriminate against one of the rays.

The specific conclusions of the tests may be summarized as follows: The distortion is most severe on the lower modulation frequencies and on the lower radiation frequencies. The distortion was equally severe with or without limiting. The diversity tests showed a tendency for the two receivers to stay in phase at the lower modulation frequencies, but this tendency was almost obliterated by unequal harmonic distortion on the two receivers. The higher modulation frequencies and the lower radiation frequencies showed the most random phase characteristics.

The specific conclusions of the theory concerning the two-path case may be summarized as follows: When the two paths are exactly equal in amplitude, provided conditions are such that the resultant amplitude does not go to zero, the effect is a change in the effective frequency deviation with no harmonic distortion. When the strength of one path is great compared to the other, odd harmonics are introduced when the phase difference between the carriers is zero and even harmonics are introduced when the phase difference is ninety degrees. The most destructive distortion occurs when the amplitudes of the paths are not quite equal. Under this condition the distortion is greatest for the higher values of F_d/F_m and for phase relations such that the phase differences between the modulation frequencies and between the carrier frequencies are both 180 degrees.

Acknowledgment

The author is indebted to Messrs. H. H. Beverage and H. O. Peterson under whose guidance the work of this paper was carried out.

[14] T. L. Eckersley, "Multiple signals in short-wave transmission," Proc. I.R.E., vol. 18, pp. 106–122; January, (1930).

[15] Bertram Trevor and P. S. Carter, "Notes on propagation of waves below ten meters in length," Proc. I.R.E., vol. 21, pp. 387–426; March, (1933).

A CATHODE-RAY FREQUENCY MODULATION GENERATOR*†

By

Robert E. Shelby‡

Video Operations Engineer,
National Broadcasting Company, Inc.

Summary—*A device similar to a cathode ray tube is described which makes possible conversion from amplitude to phase shift modulation with a maximum phase shift of many times 360°. The modulating signal alters the rotating electron stream in such a way as to produce phase changes in the multi-piece anode current. A suitable network in the audio input effects the conversion from phase to frequency modulation.*

THE device herein described is a means of frequency- or phase-modulating a carrier signal obtained from an oscillator of constant frequency and phase, such as a crystal-controlled oscillator. The majority of the known methods for generating frequency or phase modulation involve variation of the frequency or phase of the master oscillator. There is one well-known method of producing phase modulation of a carrier obtained from a constant source which involves the addition of two voltages of the same constant frequency having a constant phase difference of 90°, one voltage having constant amplitude and the other varying in amplitude with the modulation signal.

This method is limited to maximum phase shifts of the order of 30°, which means that in order to obtain a large shift at the final carrier frequency, it is necessary to employ frequency multiplication of large ratio. The method proposed here does not have this limitation. Theoretically it should give distortionless phase shift of many times 360°.

Structurally the device consists of an electron gun, two sets of electrostatic deflection plates (or magnetic deflection coils), and a target anode of special design, enclosed in an evacuated container of suitable size and shape. A typical arrangement is shown schematically in Figure 1. It will be seen that the device illustrated is similar to a conventional cathode ray oscillograph tube. The only fundamental

* Decimal Classification: R630.1.
† Reprinted from *Electronics*, February, 1940.
‡ Now Director of Television Engineering Operations, National Broadcasting Co., Inc., New York, N. Y.

difference is the novel design of the target anode. Figure 2 illustrates one form which the target anode may have. It consists of two (or more) metallic plates with curved edges, upon which the electrons from the electron gun impinge. In order to obtain phase modulation in which the angular shift is a linear function of the modulation amplitude the edges of the plates have a curvature given by the polar equation

$$r = a\,\theta \tag{1}$$

which defines an Archimedian spiral.

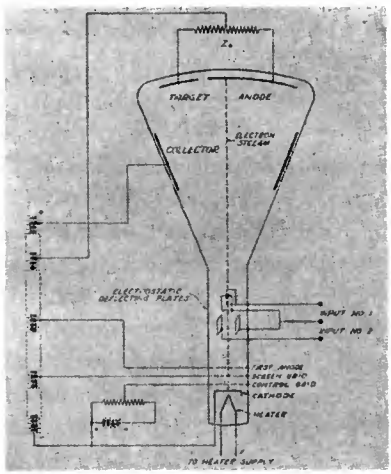


Fig. 1—Arrangement of generator. The structure is very similar to that of an electrically-deflected cathode-ray oscilloscope tube, except that a target anode is used in place of the fluorescent screen.

In operation, the electron stream is deflected in such a way that it traces out a circle on the target anode, the diameter of the circle being a variable which is directly proportional to the instantaneous value of the modulation signal. This involves no new concepts other than those associated with the familiar cathode-ray oscillograph. When controlled in this way the electron stream produces a phase (or frequency) modulated wave upon striking an anode having the special design just described.

Details of Operation

The electron gun is controlled and focused by adjusting the d-c potentials applied to the cathode, control grid, screen grid, and first

anode, just as in the case of oscillograph tubes and picture tubes. The electron stream is sharply focused on the target anode and when there is no voltage applied to the electrostatic deflecting plates it is adjusted to strike the exact geometrical center of the target anode. The inputs, which consist of two carrier waves of equal frequency and almost equal amplitude but differing in phase by 90° and having the same amplitude modulation, are applied to the two sets of electrostatic deflecting plates. Figure 3 illustrates one means of supplying the inputs. The output appears in the target anode circuit, across the impedance Z_A. If Z_A is a pure resistance the output voltage appearing across it will be a flat-topped wave. By using a target anode composed of a larger number of curved sections the fundamental frequency of the flat-topped output

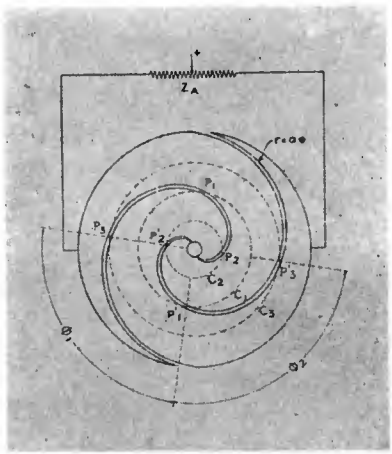


Fig. 2—Typical target anode, consisting of two metal plates divided by a pair of Archimedian spirals. The cathode ray beam scans the target in a circle of variable radius, producing square waves in the output.

wave may be made any desired multiple of the input frequency. Thus it is seen that, if desired, frequency multiplication may be obtained during the process of converting amplitude modulation into frequency or phase modulation.

For a more detailed explanation of the way in which amplitude modulation is translated into phase modulation, we refer now to Figure 3. First, with no modulation, the phase shifting network and amplitude controls are adjusted so that the electron beam describes a circle on the target anode. If the master amplitude control is adjusted so that this circle is of the size designated by c_1 (Figure 2) then the voltage appearing across Z_A will be as shown in Figure 4A. Note that the

electron stream passes from one segment of the target to the other at points p_1 and p_1'. If now the two deflecting voltages are decreased 50 per cent by changing the master amplitude control, all other controls being left the same, the locus of the end of the electron stream will be c_2 and the voltage across Z_A will be as shown in Figure 4B. The electron stream now passes from one segment of the target to the other at points p_2 and p_2'. Likewise if the master amplitude control is adjusted to give voltages 50 per cent greater than those which gave the locus c_1, then c_3 will be the new locus, the electron stream will pass from one target segment to the other at points p_3 and p_3' and the voltage appearing across Z_A will be as shown in Figure 4C. Now, if the master amplitude control is reset so that the electron circle falls on c_1 and a

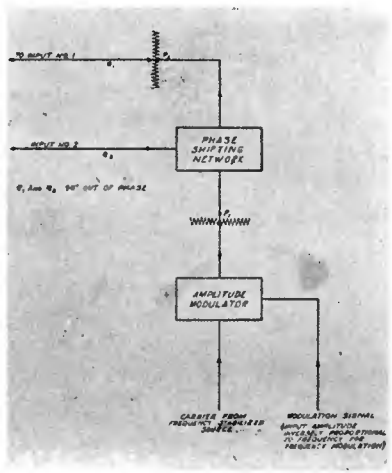


Fig. 3—Circuits for supplying the inputs to the generator. The carrier and modulation signal are combined in an amplitude modulator whose output is divided into two phases and applied to the deflecting plates.

fifty per cent modulation is then applied to the carrier in the manner indicated by Figure 3 the locus of the end-point of the electron stream will expand and contract between the limits c_2 and c_3 and the output wave will shift in phase between the limits indicated by Figures 4B and 4C.

The amount of maximum phase shift in the device described above will be determined by the curvature of the target anode boundary — that is, it will depend upon the value of a in the equation $r = a\,\theta$. For the target illustrated in Figure 2 the phase shift is plus and minus approximately 90° when the input is amplitude modulated 50 per cent. Amplitude modulation of 75 per cent on the input will give shift of plus and minus 135°, etc.

It should be noted that this device is fundamentally a generator of phase modulation — not frequency modulation. However, any phase modulator may be made to produce the equivalent of frequency modulation by means of a network in the audio input having a characteristic that is inversely proportional to frequency. It will be understood that whenever the cathode ray modulator is referred to as a generator of frequency modulation, the use of such a network is implied.

Variation of the anode voltages will cause the sensitivity of the electron stream to vary so that for constant voltages on the deflecting plates the size of the circle described on the final anode by the electron stream will vary as the anode voltages vary, thereby producing a form

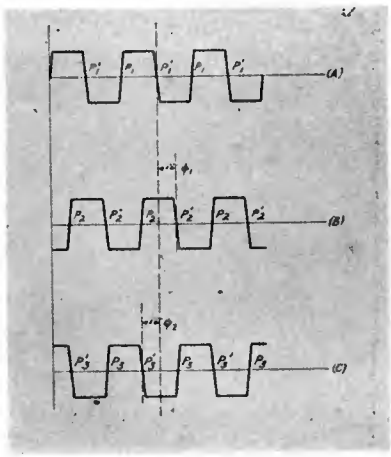


Fig. 4—Square wave outputs of the generator, lettered to correspond with the scanning circles shown in Fig. 2, opposite. The phase shift results in frequency modulated waves when the signal is passed through a frequency multiplier.

of phase modulation. This means that the d-c voltages supplied to the electron gun must be well-filtered and free from fluctuations.

Auxiliary electrodes may be added to the modulator tube for control or monitoring purposes. For example a fluorescent screen may be provided beyond the target anode, so that the electron stream will produce a pattern upon it when it passes between segments of the anode or beyond the outer edges of the anode plates. Additional electrodes, located in the same plane as the final anode, but electrically separate from it, may be used for adjusting the modulator and also for indicating overmodulation. Such an electrode, of small area, located at the geometrical center of the anode (where the anode plates are cut away) is useful in centering the electron stream. A narrow annular ring

around the outside of the main target facilitates adjustment of the phase shifting network to obtain circular deflection of the electron stream. Many other auxiliaries are possible.

The description so far has related to the production of modulation in which the phase shift is directly proportional to the modulating voltage, but the system is quite flexible in this respect. By use of a properly shaped target, the phase shift may be made any reasonable function of the modulating voltage, this being determined by the curvature of the edges of the target plates and the nature of the path traced on the target by the electron stream.

The amplitude of the flat-topped output voltage wave may be varied independently of the frequency (or phase) modulation by varying the d-c potential applied to the control grid. If it is desired to amplitude modulate the output wave in addition to or instead of phase modulating

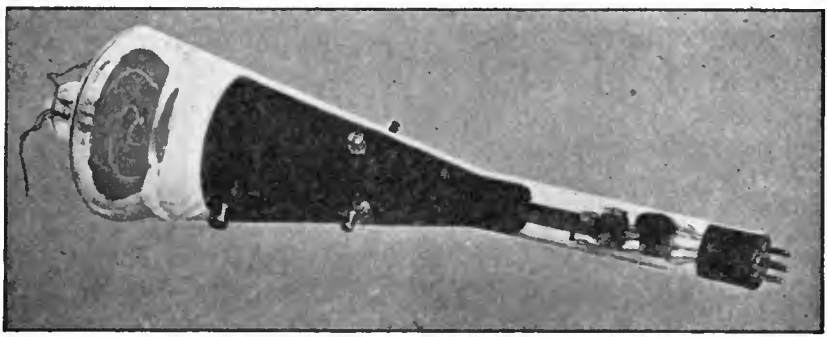


Fig. 5—Early form of the cathode-ray frequency modulation generator, mounted in a standard oscilloscope tube envelope. The target anode is composed of spirals which make one complete turn.

it in the manner described, this may be done readily by applying the amplitude modulation signal to the control grid, provided the electron gun is so operated that the rate of electron emission from the gun is a linear function of control grid voltage over the operating range.

The photograph of Figure 5 shows the general appearance of an early model of this device constructed by the Radiotron Division of the RCA Manufacturing Co. It consists of a standard type cathode ray oscilloscope tube with a spiral target of the type specified above sealed into the large end in place of the usual fluorescent screen. The target was constructed by applying a coating of platinum to a sheet of mica of the kind used for Iconoscope mosaics and scribing the separating line in the platinum surface to give the proper shape to the two electrodes. The target is mounted by wire supports which are sealed into the large

end of the tube. Electrical connections to the two electrodes of the target are brought out through this same seal.

The spiral of this target was designed to give a maximum over-all phase shift of 360°—that is to say, for this particular target the value of a in the equation $r = a\,\theta$ is $\frac{1}{\pi}$ inches per radian (the diameter of the target is approximately four inches). In order to facilitate adjustment and to judge operating performance four concentric circles and sixteen radial lines were marked on the surface of the target with willemite. The entire surface of the target was also covered with a very thin coating of willemite. This results in the electron beam tracing out a visible path on the target, which is helpful in testing the tube.

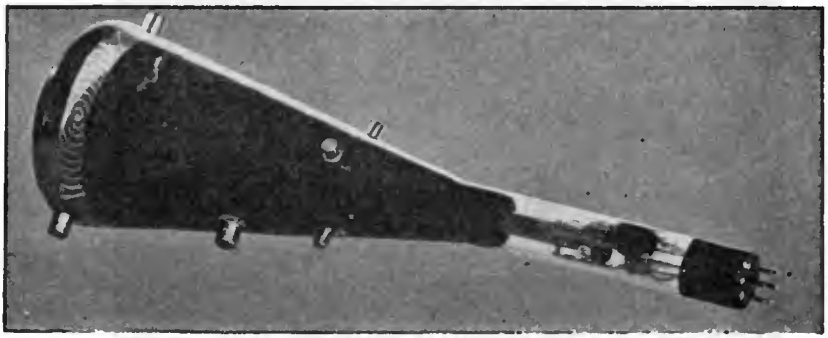


Fig. 6—Later form of the generator. The spiral anode makes five complete turns, and hence the degree of phase shift, for a given change in scanning radius, is proportionately increased.

In constructing this tube the Radiotron engineers made provision for utilizing secondary emission from the target anode to provide additional output. This was done by providing a separate collector for secondary electrons in the form of a coating on the inner wall of the tube.

Figure 6 shows a photograph of a later model. The general design is the same except that the target anode consists of a conducting coating deposited directly upon the inner surface of the large end of the tube and the spiral in this case contains five complete cycles. The configuration of this target is shown by Figure 8.

Figure 7 shows a schematic of the transmitter set-up employed to test the frequency modulator tube. That portion of the circuit enclosed by the upper dotted rectangle represents a conventional r-f transmitter. It was of the crystal-controlled type with buffer amplifiers and fre-

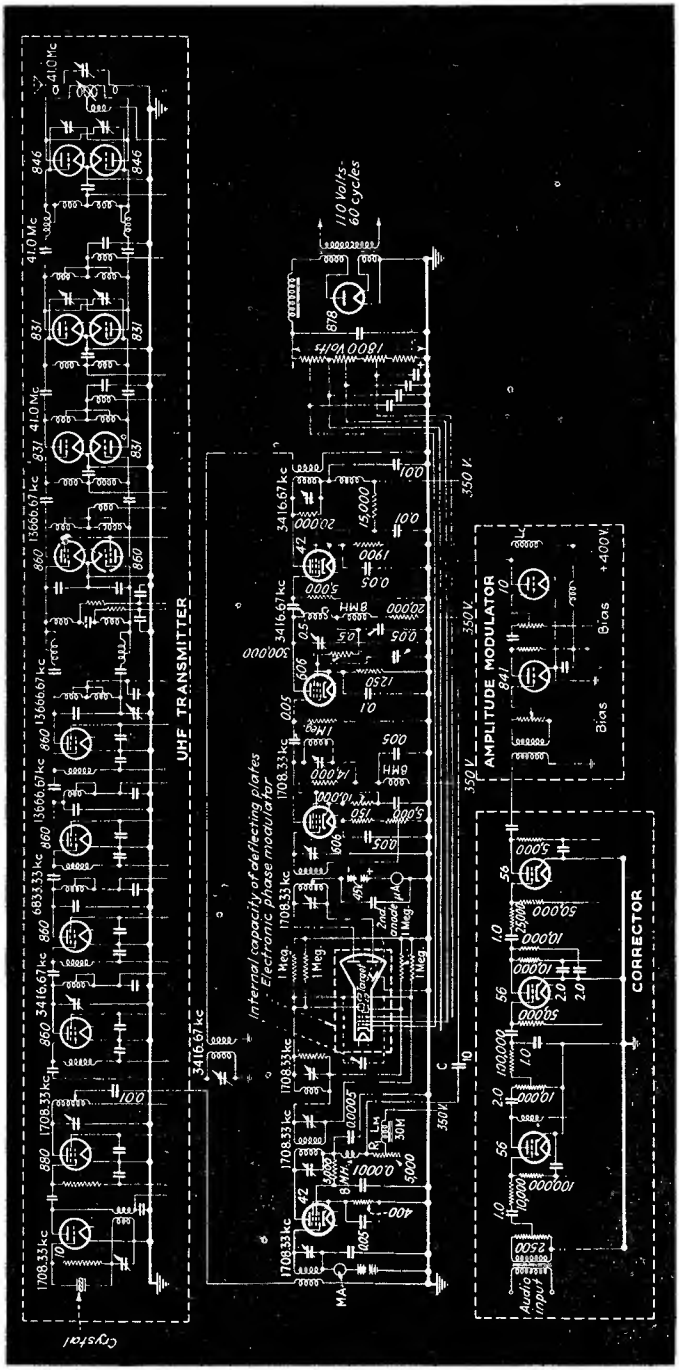


Fig. 7—Complete circuit diagram of transmitter used to test practical operation of the f-m generator.

quency multiplier stages driving a final amplifier delivering a carrier output of approximately two kilowatts.

The crystal-controlled oscillator in the transmitter was used as the primary source of stabilized r-f voltage, the amplifier chain in the transmitter being broken between the buffer amplifier and the first doubler stages for insertion of the cathode ray phase modulator and associated apparatus, as shown.

To obtain the 90° phase relationship between the voltages on the two sets of deflecting plates of the cathode ray tube, a phase-splitting network consisting of a capacitive reactance in series with a pure resistance was employed, one set of deflective plates being connected across the capacitive reactance and the other set across the resistance.

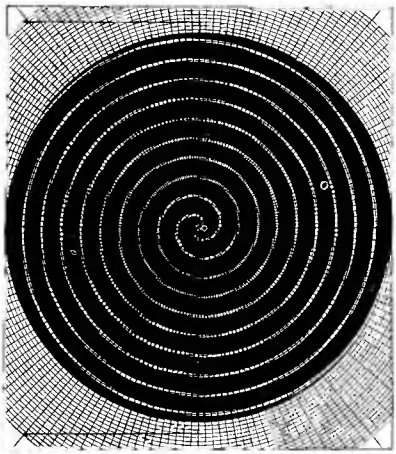


Fig. 8—Spiral anode of five turns, similar to that shown in the tube in Figure 6. The spirals were formed by scribing through a platinum film on a mica support, thus separating the surface into two insulated segments.

In this set-up the internal capacity between one pair of the deflecting plates was utilized as the capacitive reactance. A parallel tuned circuit was bridged across the other set of deflecting plates to permit tuning out the reactance of the variable resistor and also that due to the capacity between these deflecting plates, thus obtaining a purely resistive impedance. The variable resistor in this circuit was used to adjust the relative amplitudes of the voltages on the two sets of deflecting plates, and the amplitudes of the two deflecting voltages were adjusted simultaneously by means of the variable resistor R_1 in series with the d-c plate supply to the modulated amplifier.

The output circuit connected to the two halves of the target anode was tuned to the fundamental crystal frequency and coupled inductively

to the grid of an amplifier which in turn fed a frequency doubler. The succeeding stage fed a shielded r-f transmission line, the other end of which was coupled to the 860 stage in the transmitter which ordinarily operated as a doubler but which was used as a straight amplifier in this case. From that point on the various stages of the transmitter were operated in their usual manner.

The transmitter was located in the Empire State Building and reception tests were made with a frequency modulation receiver located in the Development Laboratory in Radio City. These tests showed that the cathode ray frequency modulator performed as predicted. When properly shielded against stray magnetic fields and provided with well-filtered d-c potentials, it introduced no measurable distortion and was perfectly stable in its operation.

In conclusion, acknowledgment is made of the cooperation of the RCA Radiotron Laboratories in constructing and supplying the tubes used in these tests and of the helpful suggestions offered by Radiotron and NBC engineers. The author also wishes to acknowledge with deep appreciation the encouragement and cooperation extended by Professor E. H. Armstrong of Columbia University during the early tests on this device.

NBC FREQUENCY-MODULATION FIELD TEST*†

By

Raymond F. Guy and Robert M. Morris

National Broadcasting Company, Inc., New York

***Summary**—The full theoretical advantages of frequency modulation may be obtained in practice if the transmitting and receiving apparatus are properly designed.*

For primary service, amplitude modulation on the ultra high frequencies offers some advantages over standard broadcasting. Frequency modulation offers advantages over amplitude modulation on the ultra high frequencies. The advantages to the listener of frequency modulation on the ultra high frequencies consist of freedom from the 10-kc beat note and side-band interference which result from the frequency allocation of standard broadcasting, and also the reduction of locally generated noise, atmospherics, and interference from distant stations operating on the same channel. Standard broadcasting has the advantage of providing clear channel night-time service to vast areas which would not be served by frequency modulation on the ultra high frequencies.

TWENTY years ago all frequencies above 1500 kc were generally considered to be of such little value that even the amateurs had objections to being confined to them. There is no need to state here what has since occurred on these "useless" frequencies nor to dwell on the fact that the surface has but been scratched. One service after another has wholly or partially transferred the bulk of its activities to them, and a multitude of new and invaluable services have been made possible by their use. So-called "Standard Broadcasting" had a most humble beginning on 830 kc, which was then in the middle of the marine band of 500 kc to 1000 kc. Broadcasting quickly crowded the original occupants out of most of this band. It is not one of the services which have since moved into the high-frequency spectrum. It remains on the former marine frequencies where it started. But there is a possibility that a shift may be approaching.

The use of the ultra-high frequencies for sound broadcasting offers some technical advantages. These advantages consist of escaping the present 10-kc channel limitation, getting away from static and eliminating all except spasmodic long-distance interference. We have known this for many years, have for a decade experimentally operated low power u-h-f stations and at times have had the experience of receiving good service from our low-powered u-h-f transmitters when static ruined reception from our high-powered standard broadcasting plants. Five years ago the FCC had applications for, or had licensed,

* Decimal Classification: R630 × R270.
† Reprinted from *RCA REVIEW*, October 1940.

over 100 u-h-f transmitting stations and it seemed that a trend was developing toward u-h-f broadcasting, but this trend was not sustained. Interest has been revived in recent months through wide-spread discussion of the advantages of frequency modulation on the ultra high frequencies.

Amplitude-modulated u-h-f stations can provide greater coverage than standard broadcasting shared-channel stations limited by night-time interference from distant stations operating on the same channel. This interference usually causes such a station's useful service area to shrink to a small fraction of its daytime area and many of these stations are, in addition, required to reduce power at night to minimize similar interference to other stations. Few such stations are free from night-time interference within their 2000 microvolt contour and many are limited to 8000 microvolts. The ultra high frequencies offer an escape from such limitations by virtue of practical freedom from static and shared-channel interference. Frequency modulation can provide a much greater degree of improvement than can amplitude modulation on these ultra high frequencies.

Purpose of F-M Field Test

The National Broadcasting Company has been one of the groups which viewed realistically the possibilities of the ultra high frequencies and has pioneered in conducting experimental operations in that field for many years. It has been NBC's belief that at some time in the future these ultra high frequencies would come into much wider use. It is the policy of NBC to investigate every technical development affecting its field and apply it where possible to provide better public service. Toward that end 18 months ago NBC formulated plans for an exacting field test of frequency modulation similar to the field test of television which preceded the dedication of that service on April 30, 1939. The purposes of the field tests were to quantitatively evaluate the advantages of frequency modulation over amplitude modulation on the ultra high frequencies, not by confined laboratory measurements which had already been made by others, nor merely by operating a frequency-modulated station, but *by painstakingly measuring and comparing the two systems under all kinds of actual service conditions in the field* and determining how much of the theoretical advantage of FM could be obtained in practice.

The impression has been gained by some, that only by the use of FM may the public now enjoy high fidelity sound reproduction in the home. With ultra high frequencies, the same fidelity can be provided with either amplitude or frequency modulation. Improved fidelity is

made possible by increasing the transmitting channel width beyond the 10-kc channel allocations of the standard broadcasting band and not by using a particular type of modulation. Irrespective of the type of modulation, receivers for "high fidelity" reproduction require equally costly high power, low distortion audio amplifiers and expensive loudspeakers and acoustical systems.

Present-day transmitters of NBC and others in the standard broadcasting band transmit much higher fidelity than is reproduced in moderate-priced receivers. To reproduce sound in the degree of fidelity which is now available, requires more costly receivers. Popular interest has been much more pronounced in low-priced receivers than in high fidelity at more cost. The problems of high fidelity are problems of cost and of widespread public appreciation of improved fidelity and not of a type or method of modulation. High fidelity eventually will receive the recognition it merits. Provision has been made for it in the ultra-high-frequency channel allocations.

In the NBC field test of frequency modulation attention was directed toward the evaluation of the frequency modulation system in the suppression of undesired noise and interference, using the amplitude modulated system as the reference, or standard of comparison.

Scope of the Tests

To carry out this project properly, and for the first time completely, it was decided that:—

1. The same transmitter, transmitting antenna, receiving antennas, receivers and measuring equipment should be utilized for each system of modulation.
2. The transmitter and receiver should be equipped for instantaneous switching to either amplitude modulation or frequency modulation and the transmitter should be of 1000 watts power.
3. The transmitter power should be continuously variable over a range of 10,000 to 1 on frequency modulation, and a means of accurately measuring it should be installed.
4. The most important comparisons should be between amplitude modulation, frequency modulation with a deviation of 15 kc (total swing of 30 kc), and frequency modulation with a deviation of 75 kc (total swing of 150 kc). A minimum of 15 kc deviation was chosen because it represents a deviation ratio (deviation divided by maximum audio frequency transmitted) of 1 and because a 30-kc i-f system would still be required for a smaller deviation, to accommodate the side-bands.
5. Order wires should be used between the transmitting and main

receiving points to expedite the work and insure accurate results.

6. The observations and measurements should be conducted at a number of scattered and representative receiving locations throughout the service area of the station.
7. The observations would include shared channel and adjacent channel operation of F-M stations.

The transmitter was equipped with relays for instantly selecting at will the condition of modulation desired. Since the degree of frequency deviation is directly proportional to the audio input voltage, pads, selected by a relay, served to produce various frequency deviations. Herein, for frequency modulation, the term "per cent modulation" refers to the passband of the system and is the ratio of the total swing being used divided by the total pass-band.

Tone modulation was used for most measurements. For measurements of distortion or signal-to-noise ratios, with modulation present, the tone output of the receivers was cleaned up by passing it through filters and then impressed upon RCA noise and distortion meters.

For brevity the following designations will be used herein:

AM .....Amplitude Modulation.
FM-15 .....Frequency Modulation with a deviation of 15 kc, or total swing of 30 kc.
FM-30 .....Frequency Modulation with a deviation of 30 kc, or total swing of 60 kc.
FM-75 .....Frequency Modulation with a deviation of 75 kc, or total swing of 150 kc.
S/N .....Signal-to-noise ratio.

Preliminary work on this field test was started in April, 1939, and the project was completed on May 4, 1940. The results of this work were submitted to the FCC during the F-M hearings of March, 1940 and presumably were of assistance to that body in formulating rules and standards of good engineering practice covering u-h-f broadcasting.

The Transmitter

The 1000-watt transmitter was an RCA unit, originally designed for amplitude modulation, which was modified and equipped to meet the requirements for both FM and AM. The receivers, of which there were four, were specially designed and built throughout so as also to meet the requirements.

Special temporary authorization was requested and obtained from the FCC to use either FM or AM on 42.6 Mc, ordinarily an exclusive F-M frequency.

The station was licensed as W2XWG and was constructed on the Empire State Building in New York City. It is shown in Figure 1. This transmitter utilized an FM modulator of the type developed by Murray G. Crosby of the RCA. Some manufacturers of F-M transmitters have adopted this system as standard equipment for commercial F-M transmitters. The transmitter was equipped with a reactance control in the primary circuits of the main rectifier which

Fig. 1

permitted continuously variable control of the carrier power from 1000 watts down to less than 1/10 of one watt during transmission with frequency modulation. A General Radio multi-range vacuum tube voltmeter was provided to measure the transmission line voltage and power accurately at any point over this wide range. Both of these instruments are illustrated by Figure 2.

The frequency modulator consisted of several tubes, including an

oscillator, a reactance tube, a crystal beating oscillator, a discriminator and a filter. The operation is as follows:—

The reactance tube produces a change of reactance in its output circuit when the grid voltage is actuated by d-c bias or a superimposed audio frequency such as tone or program. This reactance controls the frequency of a connected oscillator which drives the transmitter. The other items in the unit are provided to maintain accurate average carrier frequency stability. This is accomplished by a series of simple operations. A fixed crystal oscillator beats against the fre-

Fig. 2

quency of the modulated oscillator to produce an average difference frequency of 1500 kc. This 1500 kc is brought up in level through an amplifier and impressed upon a discriminator. When the carrier frequency is such that exactly 1500 kc is produced no voltage appears in the output of this discriminator. However, when the reactance tube average frequency changes, 1500 kc is no longer produced and a voltage is developed in the output circuit of the discriminator. A discriminator output filter removes all modulation frequencies to produce a direct current which controls the reactance tube d-c bias and thus the average frequency of the oscillator. The purpose of the filter is to prevent the reactance tube bias from changing with modulation.

The transmitter modulation characteristics are shown on Figure 3.

On this curve there are plotted a-c volts as the abscissa and carrier frequency as the ordinate. The audio input to the transmitter is superimposed on the reactance tube d-c bias.

Figure 4 shows the overall frequency response of W2XWG and the Bellmore receiver before pre-emphasis and de-emphasis were inserted.

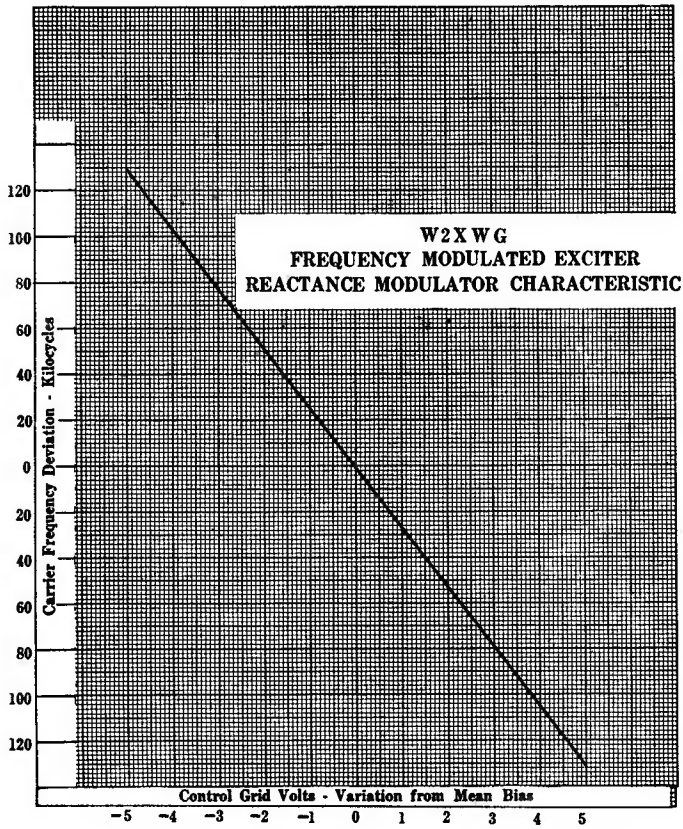


Fig. 3

The antenna normally used for transmitting for the field test was the television video unit on top of the Empire State Building.[1] However, an auxiliary antenna was built consisting of a folded dipole. This was directed eastward towards Bellmore and used when the video antenna was occupied for television schedules.

The photograph on page 5 shows the video antenna, which is 1300

[1] Television Transmitting Antenna for Empire State Building, Nils E. Lindenblad, RCA Review, pp. 387-408, April, 1939.

feet above the sidewalk. Figure 6 shows the folded dipole in operating position.

The circuit arrangement of the special 1000-watt transmitter is illustrated in the form of a block diagram in Figure 7. The frequency modulator comprises the five blocks at the upper right side, containing RCA 6V6, 6K8 and 6H6 tubes.

The degree of frequency deviation was measured by a method recently described in the RCA Review[2]. The method consists of applying a constant frequency tone to the transmitter input terminals and gradually increasing the voltage until the carrier amplitude drops to zero. When this occurs, the frequency deviation is 2.405 times the audio

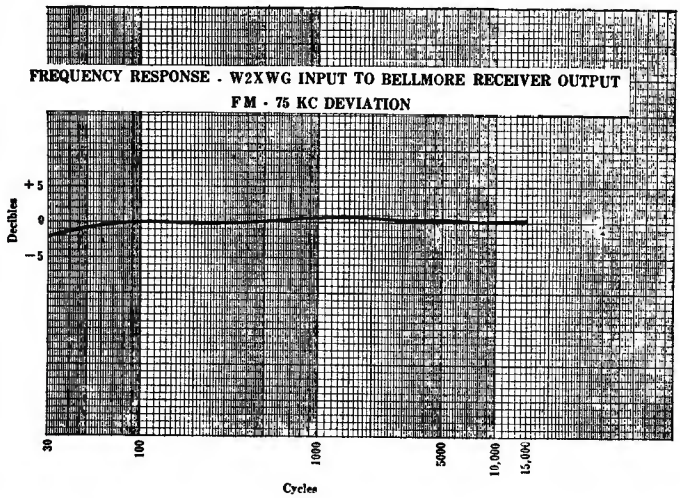


Fig. 4

modulating frequency. This method is simple and very satisfactory if reasonable precautions are taken.

The Receivers

In order that an exacting comparison of amplitude and frequency modulation can be made, it is very desirable that the receivers for the two systems be as nearly ideal and alike as possible. This practice minimizes otherwise possible errors due to differences in receiver performance. The ideal way to build such receiving systems would be the

[2] A Method of Measuring Frequency Deviation, M. G. Crosby, RCA Review, pp. 473-477, April, 1940.

common use of as many parts as possible. This was done. They were built on the same chassis. The r-f amplifier, first oscillator and the converter were used to drive two different i-f systems in parallel. Switching and chassis space was provided for a third i-f system but it was not used. The i-f systems were used separately and each was designed particularly for the reception of the modulating system to be tested. Each contained an F-M and an A-M detector and each detector had a separate audio output amplifier. The outputs of these audio

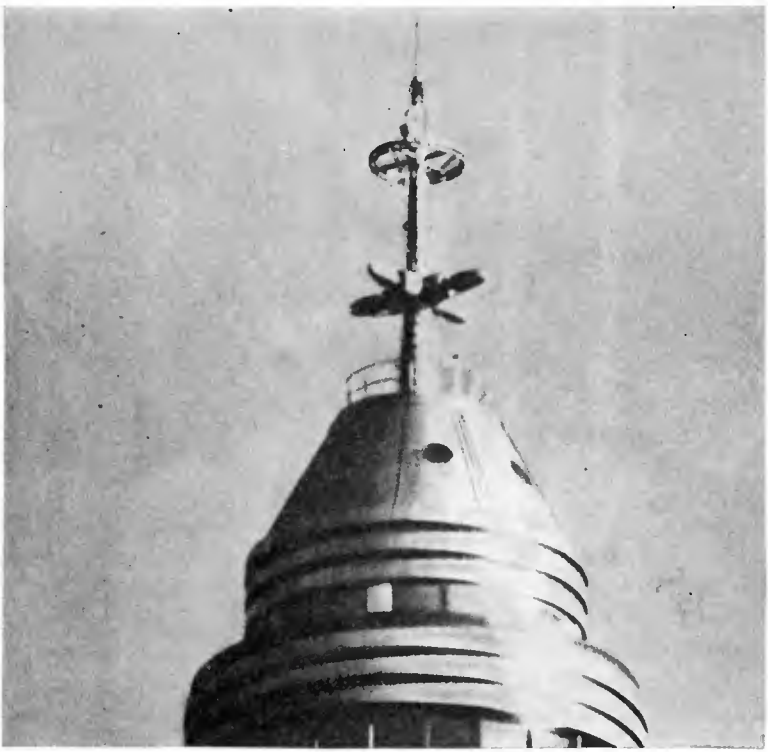


Fig. 5

amplifier tubes were all in parallel, but the system desired was selected by screen grid blocking of the other amplifiers. For AM and FM-15, parts of the same i-f system were used but the detectors and limiters were separate. For FM-75 a separate i-f system was used. This was also provided with a separate A-M detector, but it was little used because the 150-kc i-f passband was not representative of amplitude receivers. One meter was provided to show the diode currents and another was provided to show the discriminator currents. The latter is a zero-

center meter which indicates zero discriminator current when the receiver is exactly tuned for FM. Separate gain controls were provided for the i-f systems and also for the individual audio amplifiers. The receiver output circuit was provided with a very sharp 8-kc low-pass filter with a disconnecting key and the high-frequency de-emphasis circuit at this point was also provided with a disconnecting key. A single switch served to select the type of modulation which it was desired to receive.

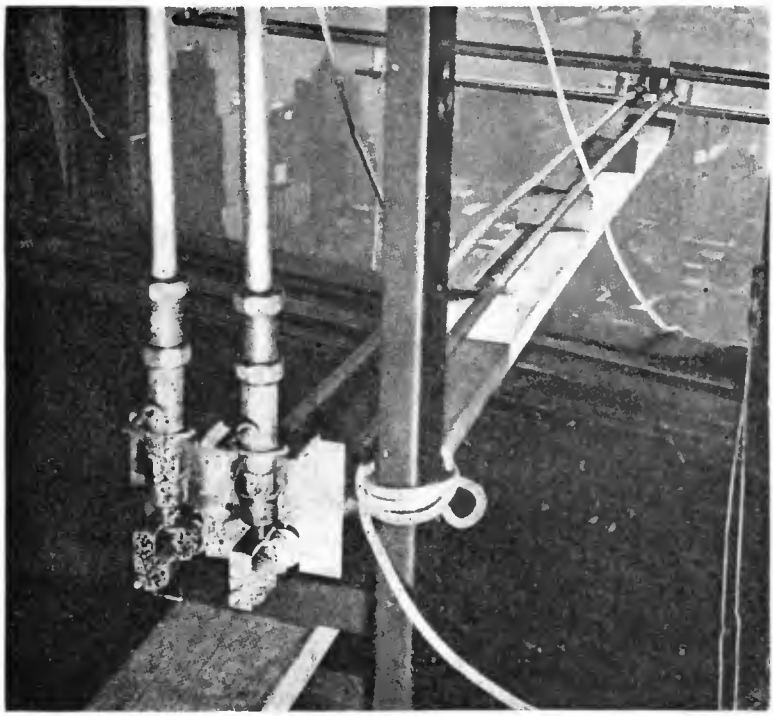


Fig. 6

Four receivers were specially built by the RCA Manufacturing Company for this test and would theoretically give full output with an r-f voltage of only 1/10 microvolt across the input terminals. The hiss level of 1 microvolt peak in these receivers was representative of the best modern receivers and was produced by thermal agitation in the antenna circuits and by tube hiss. They were made as good as receivers can be built in order that the final conclusions concerning frequency modulation as a system would not be erroneous due to apparatus shortcomings. The sacrifice of receiver design to price will not permit

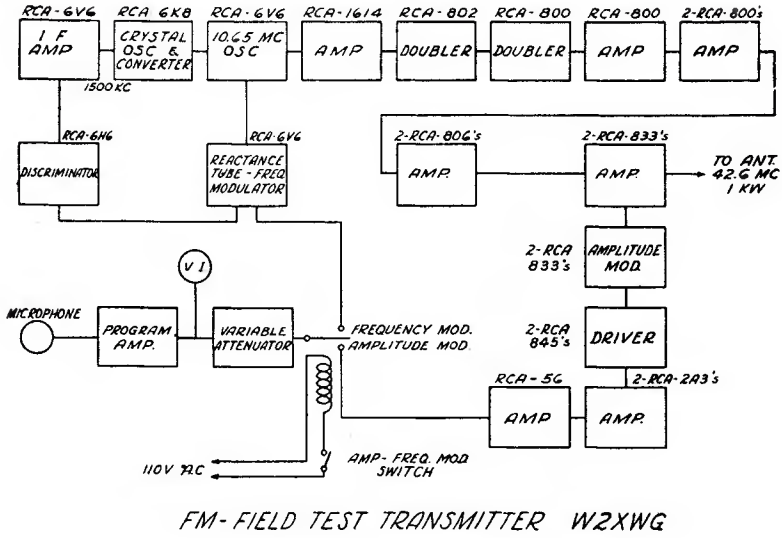


Fig. 7

the full gain of frequency modulation, as reported herein, to be realized.

A block diagram of the duplicate receivers is shown in Figure 8. A photograph of one of the receivers is shown in Figure 9. At Bellmore and also in the NBC laboratory in Radio City there were provided and used a special F-M receiver, dipole antenna and trans-

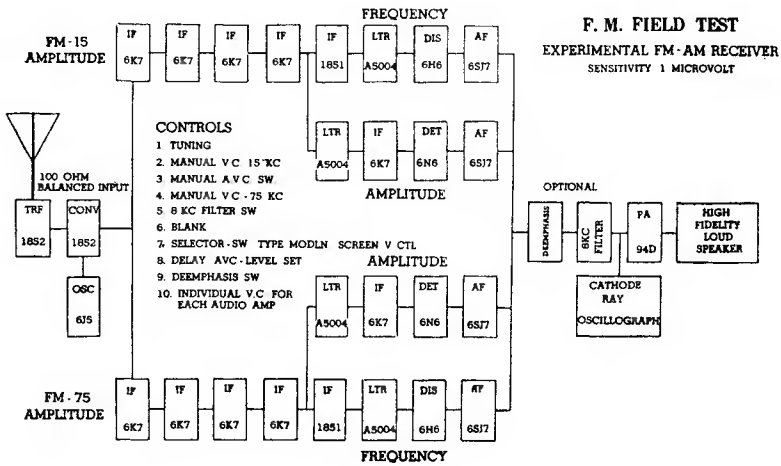


Fig. 8

mission line, commercial receivers of various makes, cathode-ray oscillographs, a special high-fidelity high-power audio amplifier, an RCA audio oscillator, an RCA noise and distortion meter, a harmonic analyzer, disc-recording equipment, a high quality loudspeaker, unweighted volume indicators, volume indicators weighted for ear response, a u-h-f signal generator, a u-h-f field intensity measuring (RCA type 301A) set, r-f transmission line calibrated attenuators, noise-producing devices including automobiles, diathermy machines,

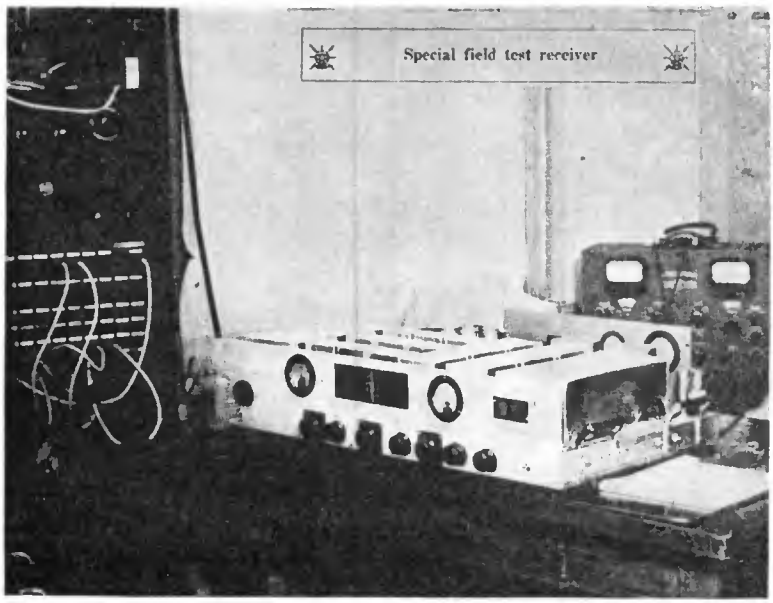


Fig. 9

etc., audio voltage amplifiers, balanced variable pads, and a microphone for recording.

The i-f systems in these receivers were painstakingly designed and adjusted for FM-15, FM-75 and AM, and at intervals during the field tests they were checked with respect to filter characteristics, i-f characteristics, limiter action, discriminator performance, audio frequency response, distortion, etc.

A shielded mutual inductance type antenna attenuator was built and calibrated for controlling the amount of signal or noise voltage reaching the Bellmore receiver input terminals from the antenna. By controlling the power of W2XWG, the antenna attenuator and the noise sources themselves, any desired carrier or noise voltages could

be produced. Separate means were provided for controlling the noise amplitudes without changing the character of the noise.

Figure 10 shows measurements made with the special RCA field test receiver and two commercial receivers. Receiver input microvolts are plotted against audio output level. The drop in the commercial receivers is due to r-f gains insufficient to operate the limiters at low input

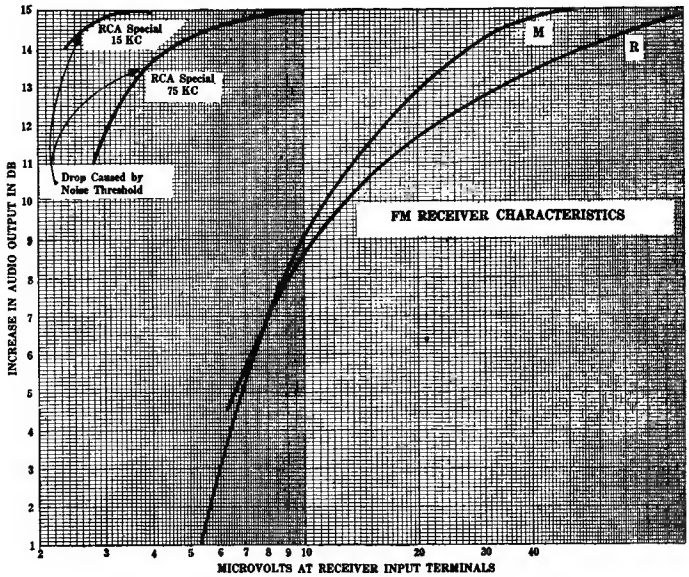


Fig. 10

voltages. The drop in the field test receiver is due to the noise threshold limitation and not lack of r-f gain.

The Field Intensity Survey of W2XWG

At the Bellmore receiving location all observations and measurements were correlated with field intensity in microvolts per meter, and also with the number of microvolts across the receiver input terminals. Similar measurements were also made of the noise voltages. In order that all of these measurements could be directly related to miles service radius, a field intensity survey was made of W2XWG. The measurements were carried out to 3.5 microvolts per meter, corresponding to a distance of approximately 85 miles, as shown on the survey map of Figure 11. The radiation index for W2XWG is 910. This is defined as the product of the antenna height, the antenna gain and the square root of the power in kilowatts.

One of the Radio Facilities engineering cars used by NBC for such measurements is shown in Figure 12. Each of these cars is a passenger sedan in which the body and interior has been modified to house a 75B RCA 500-kc to 20-Mc field intensity measuring set, a 301A RCA 18-Mc to 155-Mc field intensity measuring set, a modified RCA 302A

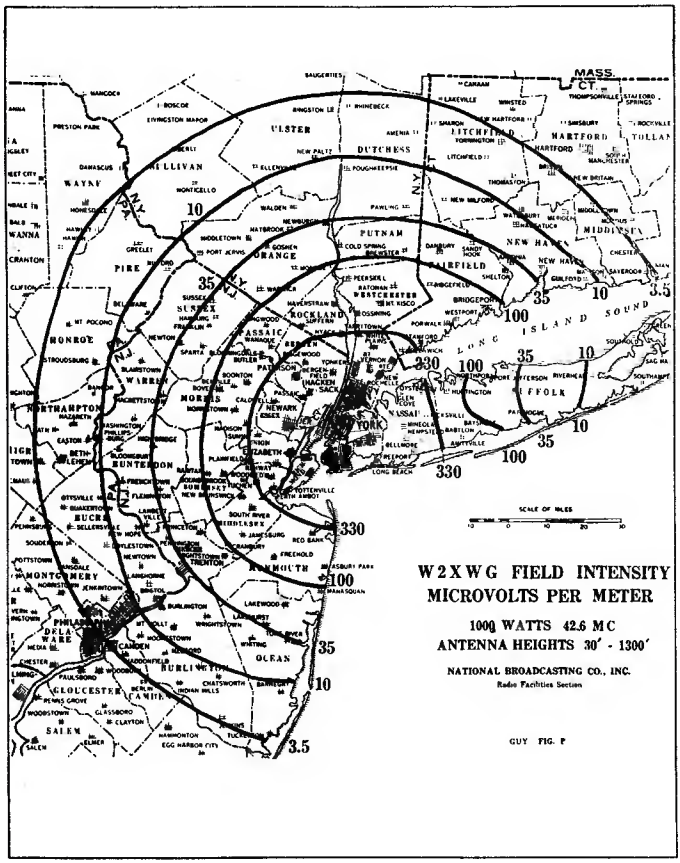


Fig. 11

noise meter, 2 Esterline Angus recorders, a driveshaft-to-recorder gear and clutch system, a special vernier speedometer and distance recorder, an automobile receiver, an aviation type broadcast loop, a u-h-f universally mounted antenna, a swivel-mounted searchlight, a compass and a steel mounting frame with aviation shock absorbers for instruments. Quick-demountable mechanisms permit easy removal of measuring gear.

The Measuring Locations

As a part of this project, field tests and electrical transcriptions were made under a variety of conditions at the following locations:—

Collingswood, N. J.	—85 miles	(temporary station)
Hollis, L. I.	—12 miles	(temporary station)

Fig. 12

Floral Park, L. I.	—15 miles	(temporary station)
Port Jefferson, L. I.	—50 miles	(temporary station)
Commack, L. I.	—36 miles	(temporary station)
Riverhead, L. I.	—70 miles	(temporary station)
Hampton Bays, L. I.	—78 miles	(temporary station)
Bridgehampton, L. I.	—89 miles	(temporary station)
Eastport, L. I.	—65 miles	(temporary station)
Bellmore, L. I.	—23 miles	(permanent station)
NBC Laboratory	— 1 mile	(permanent station)

In addition to making thousands of measurements at the permanent stations, field intensity measurements, listening tests and orthacoustic recordings were made at each of the temporary stations. The recordings compared on one disc, in each series, AM, FM-15 and FM-75. Several discs were recorded at each location under various noise conditions, including the random neighborhood noise encountered. For the observations at the temporary stations a two-car "F-M caravan" was assembled, equipped and moved from station to station. Figure 13 shows the two cars used to transport the equipment and also shows one of the typical locations selected. Figure 14 shows the Bellmore

Fig. 13

receiving station, at which most of the measurements reported herein were made. This receiving station is in a typical suburban neighborhood 23 miles from the transmitter atop the Empire State Building in New York City.

At the Bellmore station a simple horizontal dipole antenna was mounted 25 feet above the ground on a wooden pole. At the temporary receiving stations the receiving antenna consisted of a tripod-mounted dipole. The temporary station locations were selected over a range of distances which covered all grades of service ranging from zero to excellent and a sufficient number of stations were used to insure authoritative conclusions. The observations, orthacoustic recordings and comparisons made at these temporary stations provided a broad overall picture of u-h-f broadcasting without which this thorough field test would have been incomplete. The recordings remain as permanent exhibits of the results.

F-M Theory

The theory of FM has been presented in the literature[3,4,5]. However, since the following pages compare the actual performance in the field with the theoretical, there is presented for the convenience of the reader a very brief review of the reasons why FM is superior, and the nature of the superiority.

The advantages of FM over AM in noise suppression are contributed by three factors:

1. The triangular noise spectrum of FM.
2. Large deviation ratios.
3. The greater effect of de-emphasis in FM compared to AM.

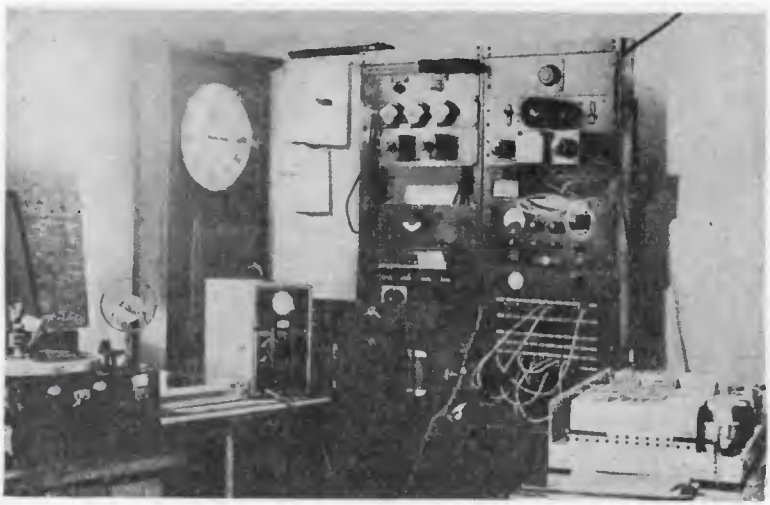


Fig. 14

The Triangular Noise Spectrum

An F-M system with a deviation ratio of one has an advantage in signal-to-noise ratio of 1.73 or 4.75 db for tube hiss or other types of fluctuating noise.

Tube hiss consists of a great many closely overlapping impulses. When combined with a steady carrier of fixed frequency, the noise

[3] A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, Edwin H. Armstrong, *Proc. I.R.E.*, pp. 689-740, Vol. 24, May, 1936.

[4] Frequency Modulation Characters, Murray G. Crosby, *Proc. I.R.E.*, pp. 472-514, Vol. 25, April, 1937.

[5] The Service Range of Frequency Modulation, Murray G. Crosby, RCA Review, pp. 349-371, January, 1940.

peaks beat with it. In the following it is convenient to consider an individual noise frequency as a separate carrier, of which there are many present at any time.

Since a combination of two station carriers differing in frequency is equivalent to a station carrier and a single noise voltage, both cases may be considered at the same time. The effect is most easily shown and understood by means of a simple vector diagram.

The desired carrier vector continuously rotates through 360 degrees and is indicated in Figure 15. A weaker carrier, or a noise voltage, rotates around the carrier vector at a frequency which is equal to the difference between the two. Amplitude modulation is produced as shown. As the undesired vector rotates around the desired vector,

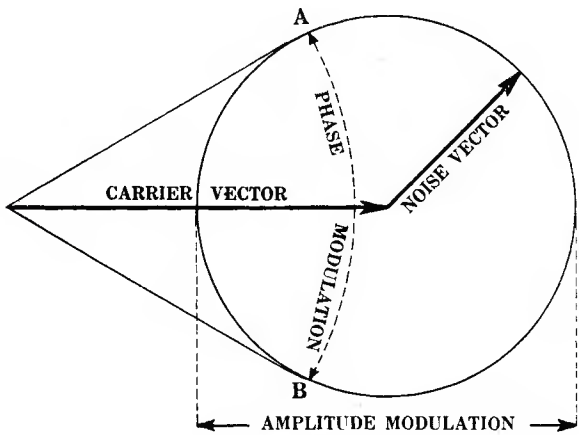


Fig. 15

phase modulation also is produced between the limits A and B. The faster the undesired vector rotates, or the faster the rate of phase change becomes, the greater becomes the momentary change in frequency and, therefore, the greater the frequency modulation becomes, because frequency modulation is a function of the first differential of phase modulation. Therefore, the amplitude of the F-M noise or beat note varies directly with beat frequency. This results in a triangular noise spectrum.

In amplitude modulation there is no such effect as this. All noise components combine with the carrier equally, resulting in a rectangular noise spectrum. The ratio of r-m-s fluctuation noise voltages in FM and AM is, therefore, the ratio between the square root of the squared ordinate areas of these spectrums. This ratio is 1.73 or 4.75 db.

The Deviation Ratio

The deviation ratio, or modulation index, is obtained by dividing the maximum carrier deviation by the highest audio frequency transmitted. For an F-M system, the suppression of fluctuation noise is directly proportional to the deviation ratio. In Figure 16 the A-M noise spectrum corresponds to the total hatched area below 15 kc because the i-f and a-f system would cut off there. The FM-75 receiver i-f system actually accepts noise out to 75 Mc and it has the usual F-M triangular noise characteristic. However, the audio amplifier and the ear respond only to noise frequencies within the range of audibility, around 15 kc, and reject everything else. Therefore, the FM-75 noise we actually hear corresponds only to the small cross-hatched triangle.

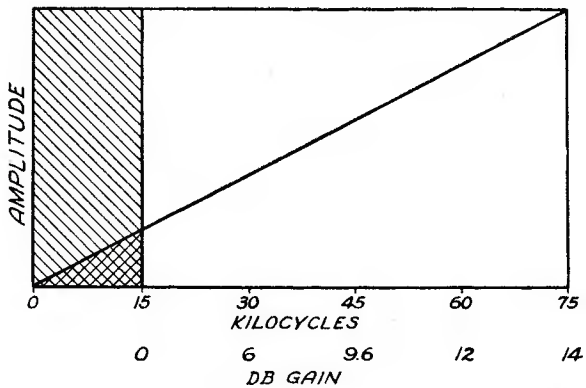


Fig. 16

The maximum height of this F-M triangle, corresponding to voltage, is only one-fifth of the height of an FM-15 triangle or the A-M rectangle. Such being the case, the FM-75 advantage over FM-15 is 5 to 1, or 14 db, and over AM it is 1.73 x 5 or 18.75 db.

De-emphasis

The use of a 100-microsecond filter to accomplish this high-frequency pre-emphasis and de-emphasis has been adopted as standard practice in television and u-h-f sound broadcasting by the Radio Manufacturers Association and recently by the FCC.

It was shown that in FM the noise amplitude decreases as its frequency decreases whereas in AM it does not. Therefore, de-emphasis is more effective in FM. This is shown in Figure 17.

The full rectangle at the left is the A-M noise spectrum. The full triangle at the right is the F-M spectrum. The application of de-

emphasis reduces these areas to those combining the hatched and black sections. Squaring those ordinates gives the black areas, corresponding to power. Extracting the square root of the ratios of these black areas gives the r-m-s S/N advantage of FM over AM. It is slightly over 4, corresponding to about 12.1 db. This includes the gains contributed by both the triangular noise spectrum and de-emphasis. The spectrum advantage is 4.75 db. Hence the de-emphasis advantage is 12.1 db minus 4.75 db or 7.35 db. To sum up, for hiss noise the F-M noise spectrum advantage is 4.75 db, the de-emphasis advantage is 7.35 db and the deviation ratio of FM-75 is 14 db, giving a total of about 26 db.

It has been reported[5] that the use of 100 microsecond pre-emphasis produces overswing or overmodulation of from 2.5 to 4.5 db with

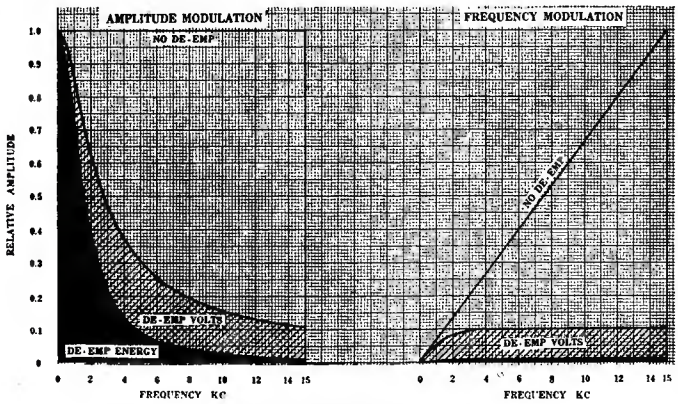


Fig. 17

program modulation, depending upon the character of the sound source. The NBC field test confirmed these conclusions. Therefore, in presetting transmitter gain controls, using low-frequency tone modulation, a 2.5 db correction should be made for program modulation.

Results of the F-M Field-Test Measurements

One of the first facts sought and determined was the lowest field intensity which could provide good service if no external r-f noise were present, and receiver hiss only were the ultimate limiting factor. Figure 18 shows the ultimate limit in service range due to the receiver noise alone, in the complete absence of any noise received on the antenna. The signal-to-noise ratio is shown for the three systems over a significant range of receiver input microvolts, field intensity in microvolts-per-meter and miles-service range. The signal-to-noise ratios were rated by subjective listening tests and the ratings are shown

on the right side of the figure. The noise threshold values are indicated in such a manner that they show where the increase of noise with modulation becomes severe but do not indicate the absolute value where an insignificant increase of noise results. The curves are shown dotted below the threshold value because the increase of noise with

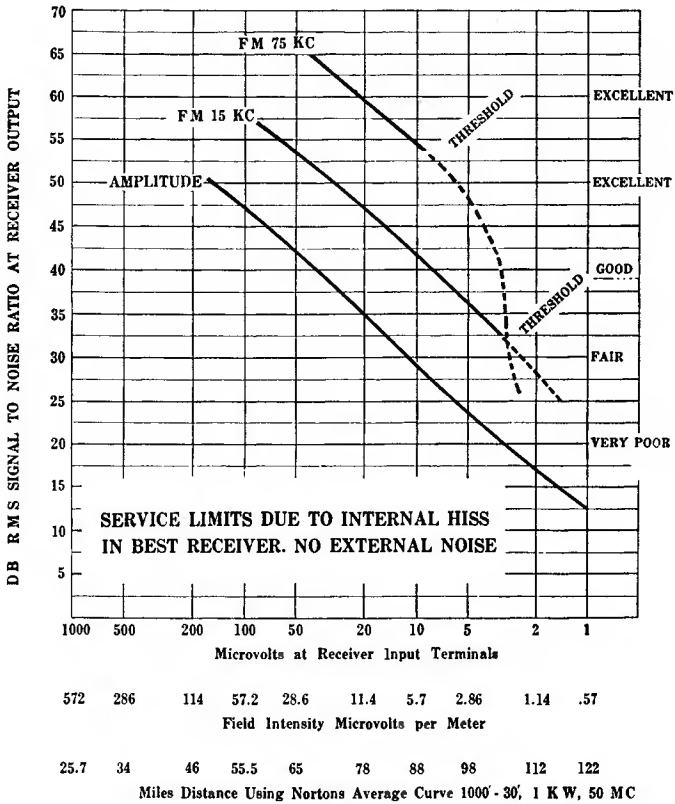


Fig. 18

modulation is very severe. A study of these curves will show the ratio of field intensity or power necessary to produce equivalent performance on the three types of modulation plotted. It is possible to determine directly the distance at which equivalent grades of service may be obtained with the three types of modulation plotted. The three curves may be projected on a straight line toward the upper left corner of the figure if it is desired to compare the results at very high signal-to-noise ratios.

The noise level in the receiver may be read directly from this figure. For instance, where there are two microvolts at the receiver input

terminals the signal-to-noise ratio on AM is 17 db. Therefore, the noise is 17 db below two microvolts RMS. This curve was made using a signal generator as a transmitter, feeding directly to the input terminals to block out all external noise. It may be seen that the full theoretical gain of FM-15 and FM-75 over AM was obtained.

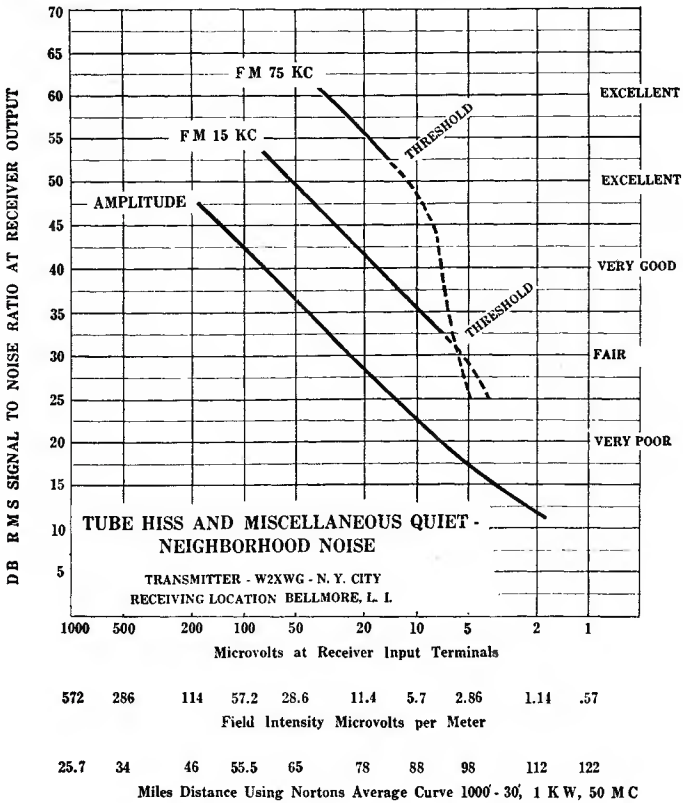


Fig. 19

Figure 19 shows the results of a series of measurements made under conditions which would be representative of a receiver operated in a home. The dipole was used and the noise received was a combination of random Bellmore neighborhood noise plus receiver hiss and thermal agitation. This receiver was located about 400 feet from the nearest public street and there was little automobile traffic in the neighborhood. The measurements were not made during a time when the noise was particularly low but the noise was taken as it existed, at random times. In a location with a higher noise level the three curves plotted on this figure would slide to the left but would other-

wise retain their characteristics, if the noise were predominantly steady in character. Here again it may be observed that the full theoretical gain of FM-15 and FM-75 over AM is obtained.

In order that this figure may be more easily interpreted, the same data has been plotted on a bar chart comprising Figure 20. This shows the number of miles service range for the types of modulation shown and also includes calculated values for a frequency-modulation system with a deviation of 30 kc. This also shows the service range with powers of 5 kw and 50 kw, assuming that the same neighborhood noise level existed at all receivers.

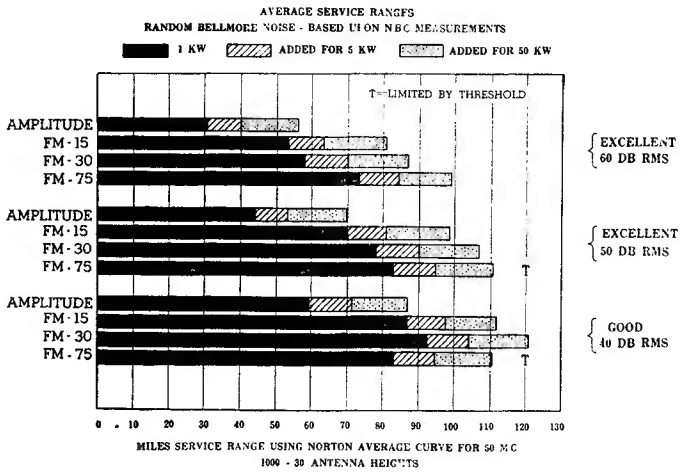


Fig. 20

It will be seen that with a signal-to-noise ratio of 40 db, FM-30 produces a greater service range than any other modulating condition, assuming that in each case the receiver is designed specially for the system being received. This would indicate that if 40 db r-m-s signal-to-noise ratio is considered satisfactory as a minimum, FM-30 would be about the optimum deviation to use. The service range for a swing of 150 kc does not extend beyond the limit imposed by the noise threshold and therefore FM-75 is limited to the distance at which a S/N ratio of 53 db is obtained in all cases.

The threshold effect actually starts on FM-75 at about 60 db but does not become severe until the unmodulated S/N ratio is about 53 db. Therefore, the threshold is indicated on the curves as 53 db.

If neighborhood noise were to be greater than that measured at Bellmore, these bars would all shorten, but would retain their relative

lengths. It will be noted that FM-75 is superior, even with the threshold limitation, at signal-to-noise ratios of 50 or more decibels.

Regardless of the carrier-to-noise or the signal-to-noise ratio coming out of the discriminator, the full benefits of FM cannot be obtained unless the audio amplifier hum level is sufficiently low. The advantages of FM do not extend into the audio amplifier.

Figure 21 shows measurements of tube hiss. The figure shows the ratio of r-m-s to peak values using an 8-kc audio band width and a 15-kc audio band width with the triangular F-M noise spectrum and de-emphasis. The hiss levels shown are representative of what can be expected from the best modern high-gain receiver.

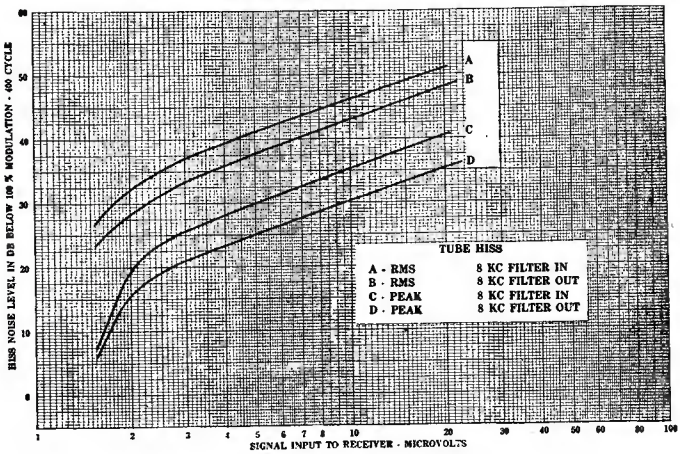


Fig. 21

From the curve, it can be observed that the ratio of peak to r-m-s values with 8 kc is slightly over 10 db. With 15 kc it is slightly over 13 db. It can also be seen that the r-m-s noise is reduced 3 db when the pass-band is reduced from 15 kc to 8 kc. The peak noise is reduced 5 db under the same conditions.

Operation of Two F-M Stations on the Same Channel

By referring to the section covering noise interference it can be seen that the worst condition of shared channel operation occurs when both stations are unmodulated, and a fixed beat note, therefore, results. It will also be seen that the higher this beat note the greater will be its amplitude up to about 5,000 cycles. Figure 22 was made on the basis of the worst conditions, which occur when the difference in carrier frequency reaches approximately 5,000 cycles. Were it not for the effect of de-emphasis in the receiver the beat note amplitude would contin-

uously rise with frequency. However, de-emphasis of the high frequencies prevents that from happening. The effect may be further understood by referring to the section on pre-emphasis and de-emphasis. It will be noted that the noise on the desired station caused by the undesired station varies inversely with the deviation ratio. Here again the theoretical advantage of FM was obtained in our field test.

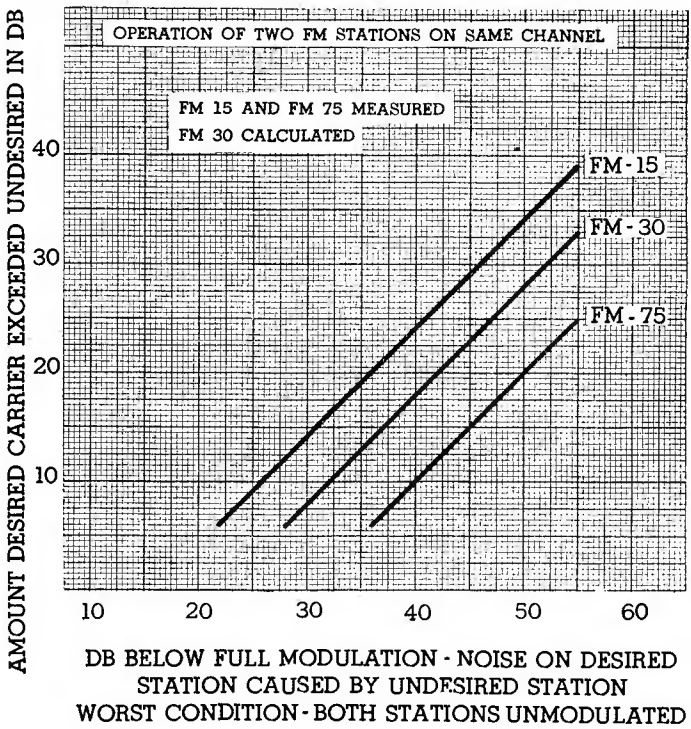


Fig. 22

When either of the stations producing the beat note becomes modulated, the beat note disappears, because one carrier sweeps across the other one. When the desired station is approximately 20 db stronger than the undesired station, all interference and cross talk effects become unnoticeable. At 12 db difference they are noticeable, but it is the opinion of some engineers that the 12 db ratio would be tolerable. Frequency modulation offers a great advantage over amplitude modulation in the allocation of stations on the same frequency. In AM the carrier amplitude of the desired station must be 100 times, or 40 db greater than the undesired carrier amplitude for a 40 db signal-to-beat

note ratio. For FM-75 it need be only 10 db, or 3 times greater; for FM-30, 17.5 db, or 8 times greater; for FM-15 24 db, or 10.5 times greater. The result is that F-M stations may be located much closer geographically than A-M stations, and therefore many more station assignments can be made per channel.

Operation of Frequency-Modulation Stations on Adjacent Channels

Figure 23 shows the results of listening tests on adjacent F-M channels using in each case FM-75. On the basis of this information it is

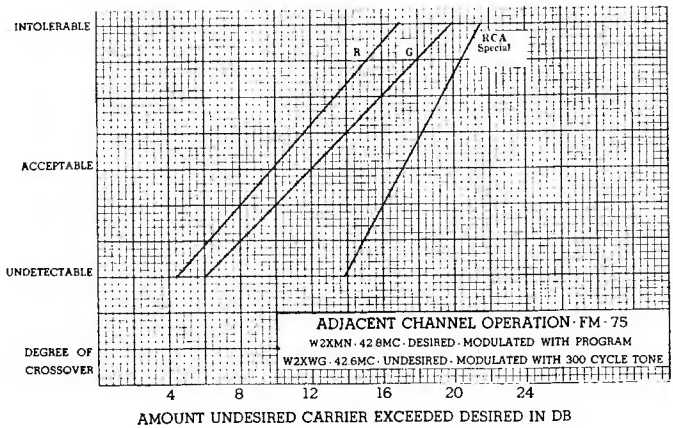


Fig. 23

seen that adjacent channel stations must not be located in the same geographical area. The channels were 200 kc wide and were adjacent, one being W2XMN on 42.8 Mc and the other W2XWG on 42.6 Mc. W2XMN was used as the desired channel, since its field intensity was constant at Bellmore. W2XWG was used as the undesired station. The field intensity ratios were adjusted as desired by varying the power of W2XWG. Observations were made of the special field-test receiver and also of two commercial receivers with the results shown.

The figure shows that the undesired carrier level should be not more than about 10 db greater than the desired carrier level to prevent objectionable cross-talk in the commercial receivers. The RCA special field-test receiver will give equivalent performance at a carrier level ratio of 17 db, but this receiver is more elaborately built and is superior in certain respects to commercial models.

Intolerable cross-talk occurs on all three receivers at carrier level ratios of from 17 to 21 db.

The Noise Threshold in Frequency Modulation

Crosby points out that an interesting series of events takes place in a frequency-modulated system when the noise peaks equal or exceed the peaks of the carrier.[4] The result is a rapid increase of the noise

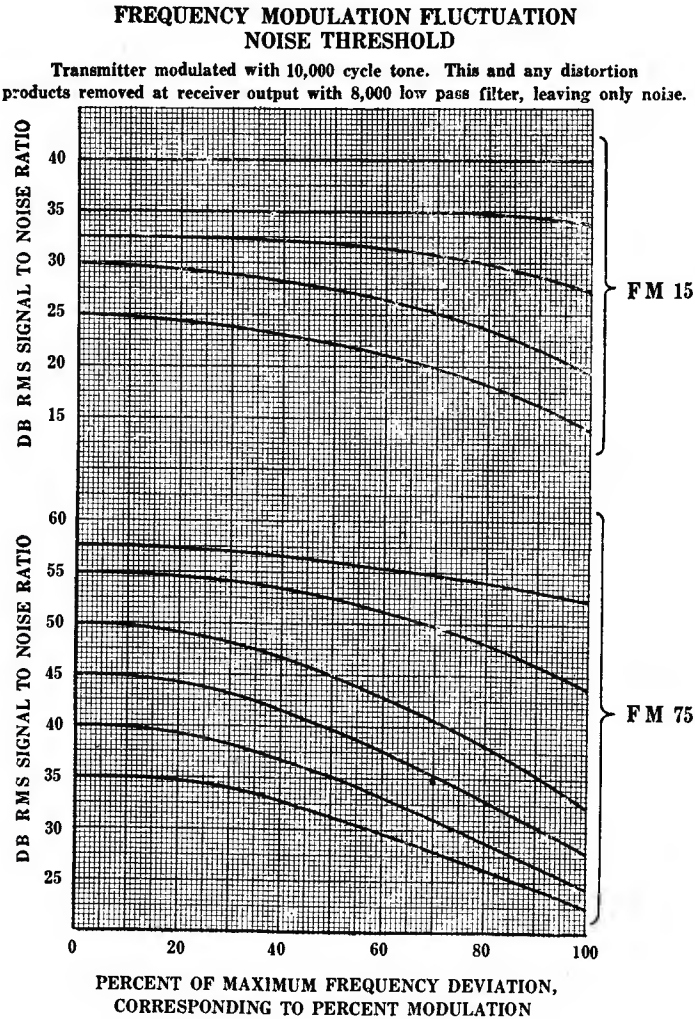


Fig. 24

level or decrease of the signal-to-noise ratio, with modulation. In frequency modulation wherein the maximum swing is 150 kc the point where this begins to occur is reached when the unmodulated signal-to-noise ratio is about 60 db. When the unmodulated signal-to-noise ratio

is less than about 60 db, or 1,000 to 1, the noise level rises with modulation, and, as the noise peaks exceed the carrier peaks by a considerable amount, this noise level may increase almost 20 db, or 10 times. When operating above the threshold limit the noise changes little as the

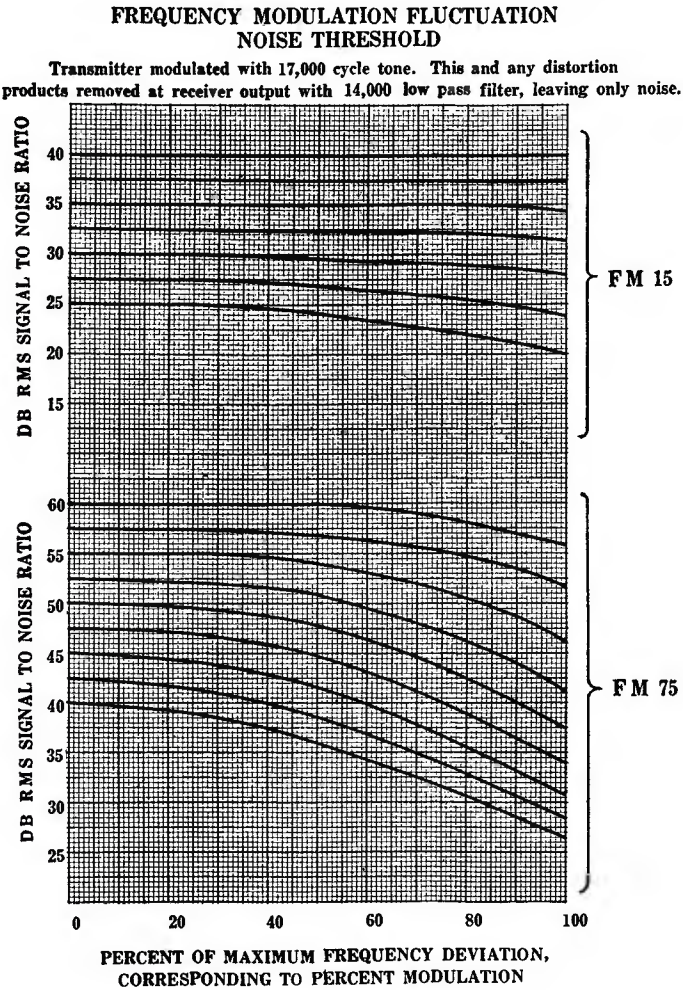


Fig. 25

station is modulated. Below the threshold limit the effect is not unlike severe harmonic distortion in an overloaded A-M transmitter.

In frequency modulation of a lesser swing, such as 30 kc, the same effect occurs. In this case, however, the threshold limit occurs at about 35 db signal-to-noise ratio. Figure 24 shows the results of some of the

measurements made at Bellmore. In order that the noise would not be confused with any small amount of inherent distortion in a man-made system, the measurements were made in such a manner that the effects of distortion were eliminated. This was done by modulating the trans-

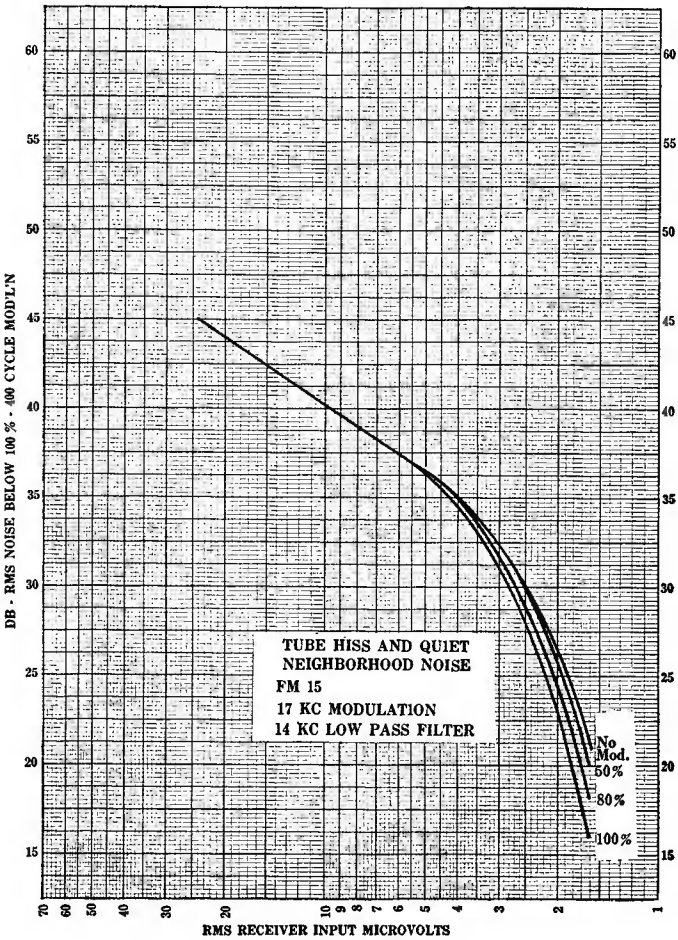


Fig. 26

mitter with a 10,000-cycle tone and eliminating at the output of the receiver, with an 8-kc low-pass filter, not only the fundamental modulating tone but also the distortion products, leaving only the noise. Figure 25 shows the results of another set of measurements made with a 17-kc modulating frequency, and a 14-kc low-pass filter to eliminate the fundamental tone and distortion products.

This effect has no doubt been observed by many without being understood. It is an inherent characteristic of a frequency-modulation system. The noise threshold in the case of an FM-40 system having a

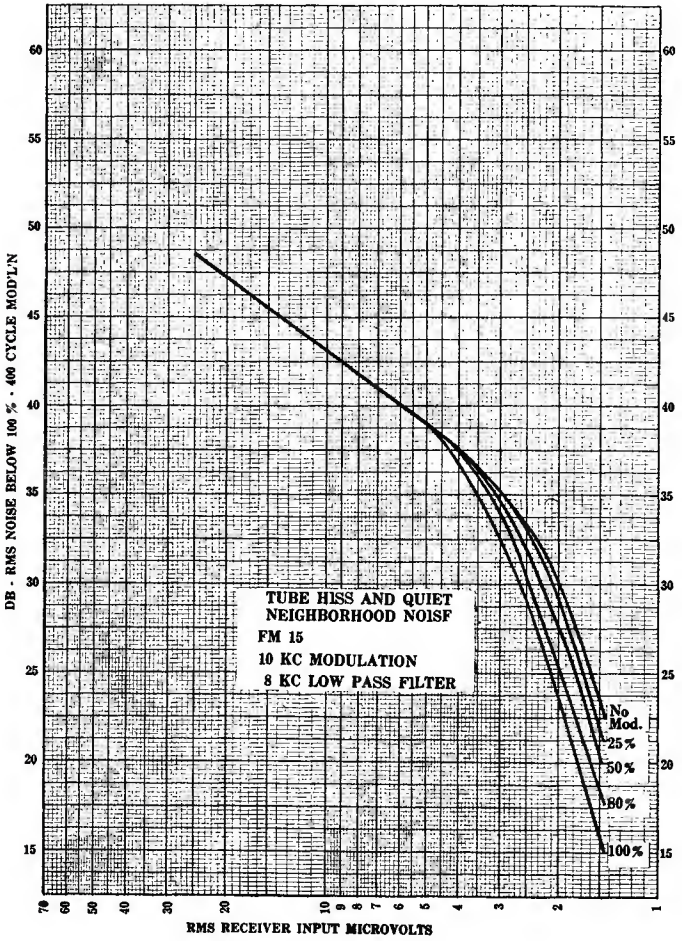


Fig. 27

total band width of 100 kc occurs at about 43 db. This provides a very good signal-to-noise ratio.

Figures 26-27 give additional results of threshold r-m-s measurements showing noise levels plotted against receiver input microvolts, with various percentages of modulation. The signal-to-noise ratio (ordinate) is the ratio of maximum 400-cycle modulation to noise.

Measurements of Peak Ignition Noise

Because of the peculiar wave shape and large crest factor of ignition noise it is preferable to measure the peak signal and peak ignition noise rather than the r-m-s values in order to establish, for one thing, the

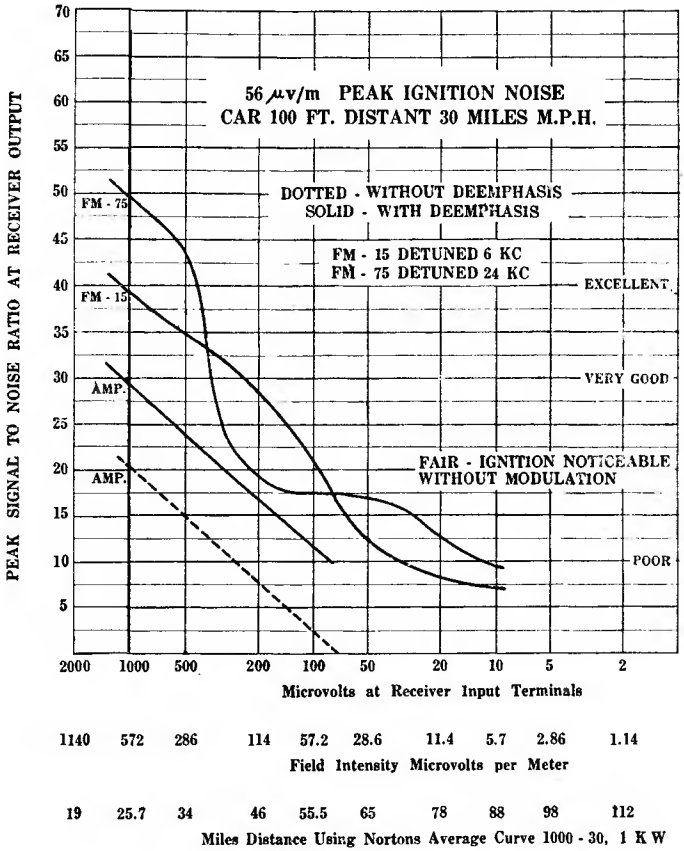


Fig. 28

threshold where they become equal. Because of the infrequent number of peaks, compared with hiss noise, much higher values of peak noise levels can be tolerated from ignition systems. In making measurements of ignition noise an actual automobile was used. Making such measurements is quite difficult because of the variation of peak noise amplitudes from an automobile system over short periods of time. Also, to show what the dynamic noise characteristics of a system would be without actually modulating it, it becomes necessary to de-tune the receiver or resort to some other expedient. This is necessary because high noise

peaks, if synchronized with an unmodulated carrier, do not show the existence of the noise threshold. Modulation, in effect, de-tunes the receiver and the threshold becomes evident. De-tuning the receiver in the absence of modulation is one expedient which produces a similar result and was the method used in obtaining the data shown on Figure 28.

Of particular interest in this figure is the rating of the signal-to-noise ratio as shown at the right. A 30 db signal-to-noise ratio, when measured with peak values, is equivalent to a 40 db r-m-s signal-to-noise ratio. Even with as low a signal-to-ignition-noise ratio as 20 db the service is quite fair, although the ignition would be noticeable without modulation. The relative infrequency of ignition peaks produces an audible result which is very deceiving. With signal-to-ignition-noise ratios of 10 or 12 db, service is still not completely ruined but could be tolerated if there were a special interest in the program material.

In general, ignition noise is transient, lasting for only a matter of seconds as a car passes by a residence. During that period the noise is not distressing. Furthermore, over a period of years, it may be expected that automobile ignition systems will be provided with suppressors which will reduce the u-h-f interference by at least 20 db. Ignition noise is of particular concern and is the predominant noise in suburban areas. In urban areas, the field intensity from a F-M station or an A-M station will ordinarily be high enough to over-ride the higher noise levels experienced. Figure 28 represents some of the results obtained with 56 peak microvolts per meter noise. The method of making these measurements was not ideal in all respects but the data is indicative of the results obtained in the presence of ignition noise. The curve of Figure 29 shows r-m-s ignition noise measurements made without modulation and illustrates that no threshold is found under such conditions. Since a system is of no value until modulated, this curve is shown only to illustrate the point that the noise threshold must be associated with modulation.

Figure 30 presents interesting data showing peak ignition noise measurements made with an 8-kc audio band width. Peak noise input microvolts are plotted against peak S/N ratio, the field intensity of the station remaining unchanged. The signal with which the noise was compared was 100 per cent 400-cycle tone modulation. The noise source in this case was an automobile ignition system built up and mounted on a lathe at Bellmore. The battery power was not varied to produce different field intensities of noise because this method changed the character of the noise. The field intensity was varied over a very wide range, without changing its normal characteristic, by orienting and

changing the length of the connected noise transmitting antenna, which was located about 50 feet from the receiving dipole. FM-15 represents a deviation ratio of 1.875 when the audio band width is 8 kc. As Crosby has shown[5], the noise spectrum advantage of FM is 6 db for impulse noise. The FM-15 threshold is shown. The FM-75 threshold is not shown because at the time the measurements were made a-c hum within

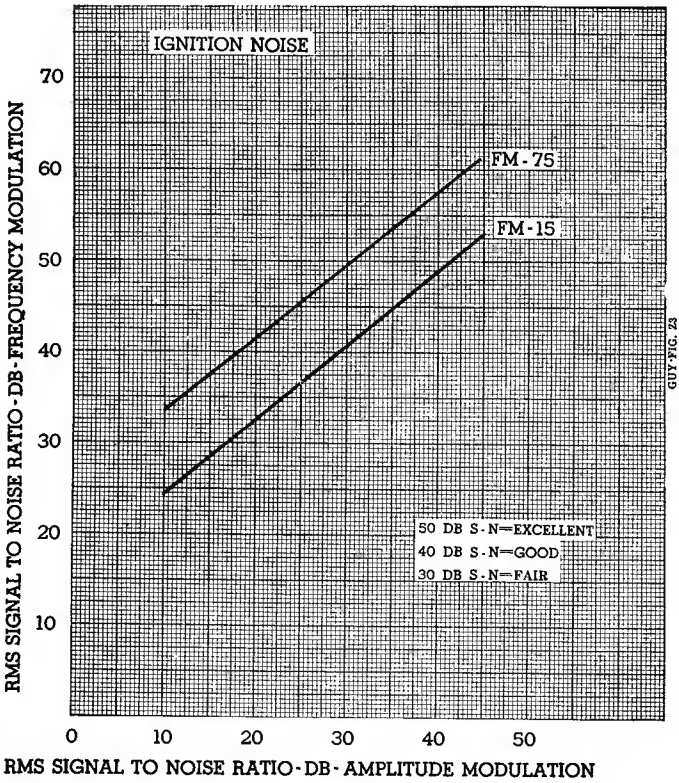


Fig. 29

the system made the accuracy of S/N measurements in the 60-db region uncertain. Of particular interest is the shape of the F-M curves at the lower right side of the figure. The explanation for it is given in the references.

This is to be expected from the character of ignition noise. The impulses are very short in duration, very high in amplitude and relatively widely separated. They literally blank out only small portions of the signal waves, without impairing the remainder. The short blanked out intervals of the signal change little over a wide range in

noise peak amplitude. Once an ignition peak has risen to the value required to control the receiver and blank out the signal a further rise in the noise level will not occur until the peak increases in breadth, or duration, or until there is a sufficient rise in certain low amplitude components of ignition noise having fluctuation noise characteristics. The peculiar shapes of such curves below the threshold values are due to the wave shapes and crest factors of ignition noise.

Results of Observations at Temporary Receiving Stations

The observations at the temporary receiving stations confirmed the measurements made at Bellmore. In going to progressively greater distances the FM-75 noise threshold distance was passed and FM-15 became superior. Then the FM-15 threshold distance was passed and AM became superior, although it was very noisy. At the limit of A-M intelligibility both FM-75 and FM-15 were completely smothered by noise. The service limits were all in excellent agreement with those shown by the bar chart, Figure 20, subject to spasmodic interference from passing automobiles. This type of interference was at all stations the predominating one, but was only intermittent and not as troublesome as might be expected.

Observations of Diathermal Interference in AM and FM

Diathermal machines vary considerably in their characteristics, some types using raw a.c. and others using partially filtered power supplies. Observations were made of interference on the three types of modulation using raw a-c machines. It was concluded from these observations that this type of interference is characterized by the transmission of a band of frequencies about 15 kc wide. With amplitude modulation the background interference is essentially equal to the carrier-to-noise ratio. With fairly weak interference from diathermy exactly centered on the desired carrier, FM-75 reception is 20 to 25 db superior to AM and FM-15 is 10 to 12 db superior. However, with the diathermy 5 kc off the desired carrier, AM was approximately equal to FM-15 and was in some cases superior to FM-75. With the diathermy carrier at the edge of the AM and FM-15 passband, the interference is highly attenuated. Under these conditions the FM-75 interference was extremely severe. With the diathermy carrier well outside of the passband of FM-15 or AM, the interference is noticeable only when the diathermy amplitude is extremely high. Under normal receiving conditions it would not be heard. However, if the diathermy under these conditions is within the FM-75 passband, the interference

is extremely severe. Thus, under such conditions, FM-75 was the worst of the three types of modulation. It was concluded that in locations having strong diathermy interference, narrow band receiving systems

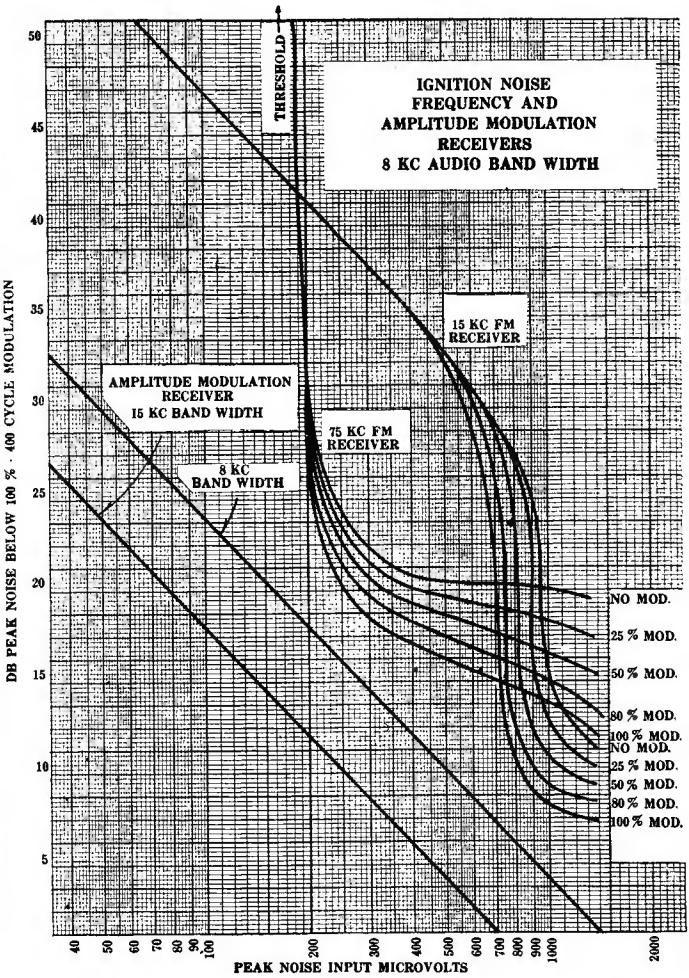


Fig. 30

would be far superior to others. It was also concluded that in locations having weak diathermy interference the wide band system would be much superior.

Noise Levels

In the foregoing, considerable data has been shown to indicate the field intensities which will provide good service for F-M and A-M sys-

tems. Figure 31 is shown because it is of particular interest in connection with this field test. It represents an accumulation of noise measurements made over a period of years by various RCA groups. These data were assembled by Dr. H. H. Beverage.

Noise levels vary considerably from time to time and it is not possible to give fixed values for any time or place. However, the informa-

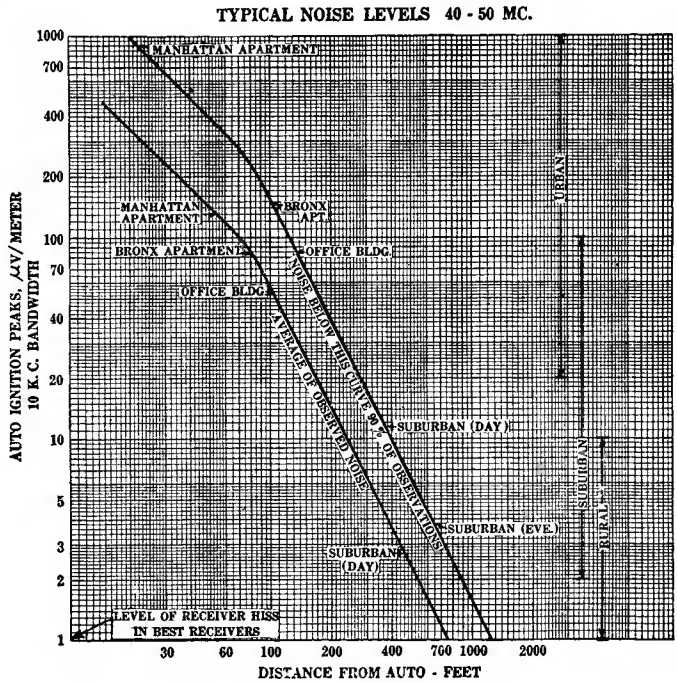


Fig. 31

tion shown on Figure 31 is indicative of the noise levels which may be encountered under a wide range of conditions with a 10-kc audio band width. For peak noise the amplitude varies directly with the frequency band. For fluctuation noise the amplitude varies as the square root of the band width.

Acknowledgment

The authors express their appreciation to Messrs. R. R. Beal and O. B. Hanson for their assistance in making this project possible. Grateful appreciation is also extended to many other NBC and RCA engineers, particularly Mr. L. A. Looney of NBC.

GENERATION AND DETECTION OF FREQUENCY-MODULATED WAVES*†

By

Stuart Wm. Seeley,‡ Charles N. Kimball, Allen A. Barco

RCA License Laboratory, New York, N. Y.

Summary—This paper contains a description of a frequency detector which is inherently linear in operation and, consequently, adaptable for precision monitoring measurements. In conjunction with a mixer and heterodyne oscillator, the detector is capable of operation at signal frequencies encountered in the usual FM transmitter. The monitor detector is discussed in the first part of the article.

The availability of a detector capable of demodulating frequency-modulated waves with the introduction of less than 0.1 per cent distortion makes possible the investigation of other low-distortion frequency-modulated devices as, for example, a generator of phase or frequency-modulated waves. The second part of the paper contains a description of a method of producing low-distortion frequency-modulated signals. The system was developed, and its low-distortion capabilities realized, with the aid of the aforementioned linear detector circuit, in conjunction with an audio-frequency wave analyzer.

Part I—Linear Frequency-Modulated Monitor Detector

Principle of Operation

Introduction

OPERATION of the detector unit may best be described with the aid of the basic circuit diagram of Figure 1. e_{in} is, in practice, a frequency-modulated signal whose carrier frequency lies in the range between 100 and 300 kc. The amplitude of e_{in} is sufficient to overswing the limits of cut-off and zero bias of the 807 tube. During portions of the cycle when this tube is cut off, the plate potential rises to the $+B$ level. During positive grid portions of any cycle the plate current rises to a certain maximum value, beyond which it does not increase with further increase in positive grid swing. (See plate family curves of beam pentode tube characteristics.) Thus, the minimum value as well as the maximum value of the instantaneous plate potential is constant, and the output wave is squared off on top and bottom at definite fixed potentials, in spite of possible variations in amplitude of grid swing.

With a square wave of plate voltage, having a repetition rate equal to that of the input signal e_{in}, the action is as follows: when the plate potential of the 807 reaches its peak value (equal to E_B) the small condenser C (25 $\mu\mu f$ approx.) in Figure 1 has charged through diode d_1 to a potential equal to E_B, with the diode end of C negative. At that

* Decimal classification: R630.1.

† Reprinted from *RCA REVIEW*, January, 1942.

‡ Now Manager, Industry Service Laboratory, RCA Laboratories Division, New York, N. Y.

part of the cycle when the plate potential of the 807 swings toward its lowest value, C discharges through diode d_2 in series with its load R_o; thus, one pulse of current flows through the load R_o of diode d_2 for each cycle of operation.

The total charge acquired by the condenser C once each cycle is CE_B (neglecting the contact potential of diode d_1). The portion of this total charge which passes through diode d_2 and R_o once each cycle is equal to that total charge (CE_B) minus the residual charge Ce_{pmin} (again, neglecting the contact potential of diode d_2). Then, as long as E_B and e_{pmin} remain constant and the time constants of the charge and discharge circuits are sufficiently small compared to the period of the input wave, the total quantity of electricity which flows through R_o in each cycle is constant. Thus, an increase in the repetition rate will increase the number of current pulses per unit of time and, therefore,

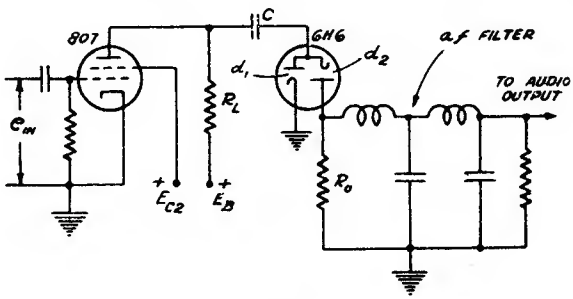


Fig. 1

increase the average value of current flowing through R_o. Conversely, a decrease in repetition rate decreases the average current through R_o, and the average potential across R_o is thus a perfectly linear function of the repetition rate or frequency of the input signal.

Obviously, then, dynamic operation of the detector will result in essentially distortionless detection over that part of the characteristic for which the previously discussed charge and discharge time-constant requirements hold.

Practical Detector Circuit

The circuit of the laboratory model of the monitor detector, shown in Figure 2, is essentially the same as the basic circuit of Figure 1, except that a 6AG7 amplifier stage has been added.

Certain pertinent factors relating to the development and use of the monitor detector are listed below.

(1) The sensitivity (audio volts developed per kc frequency deviation in the input wave) is approximately a linear function of the B voltage applied to the 807. Hence, variations in this B voltage will

introduce undesired components into the detector output. Therefore, the B supply must be well regulated.

(2) Approximately 20 peak-peak volts, at a center frequency of 150 kc, are required as input to the 6AG7, when the circuit of Figure 2 is used. Voltage at this frequency is obtained from that at any other signal frequency by means of the heterodyne oscillator and converter, shown in Figure 3. This particular circuit is designed for operation at signal frequencies in the 40-Mc range.

The a-c peak-to-peak grid voltage applied to the 807 must be at least 20 per cent greater than the value required to swing from zero bias to cut-off. The 3000-ohm series grid resistor in the 807 circuit aids in maintaining approximate equality of duration of the two halves of the plate-voltage wave, and also limits the peak grid current on positive grid swings.

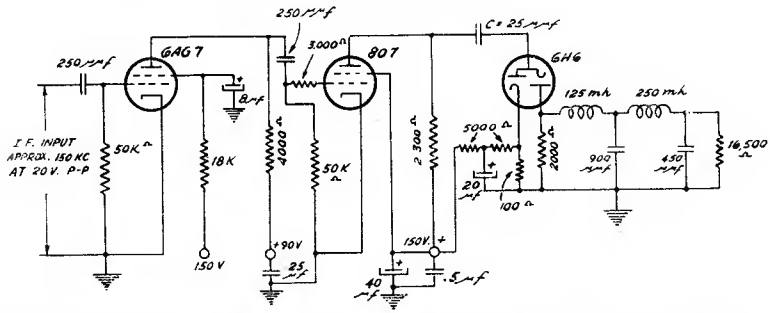


Fig. 2—90-volt and 150-volt supplies are regulated. One side of all heaters (supplied from 6.3-volt transformer) is grounded.

(3) The 807 plate load is adjusted to produce plate-current saturation at a peak grid signal just short of grid current.

(4) The value of capacitance C (in conjunction with other inherent circuit capacitances) is determined by the maximum permissible time constants in the charge and discharge circuits. These should both be sufficiently small compared to a half-period at the highest operating frequency to permit (a) e_p to rise to within 0.1 per cent of E_B in one-half cycle and (b) to permit e_p to fall to within 0.1 per cent of its normal minimum value, again within a half-cycle.

(5) The lower limit of operating frequency is determined by the cut-off of the low-pass filter in shunt to R_o (Figure 1). This filter is necessary both to remove undesired signal-frequency components from the output of the device and to allow the instantaneous potentials across R_o to rise and fall with each discharge-current pulse. If R_o were by-passed directly, to remove the signal-frequency components, the residual charge on the bypass condenser would act as a diode bias, and prevent complete discharge of condenser C. Then the discharge currents would

approach zero asymptotically as the frequency was increased, and the resultant curvature of the output voltage vs. input frequency would produce undesirable distortion.

The characteristic impedance of the filter is high compared to R_0. The filter is terminated at its output and where it may be connected to the grid of an audio amplifying tube.

(6) It has been found experimentally that a small bias voltage (approximately 1.5 volts) in series with the charge diode d_1 effectively overcomes the contact potential of both diodes and contributes somewhat to improved linearity of the device.

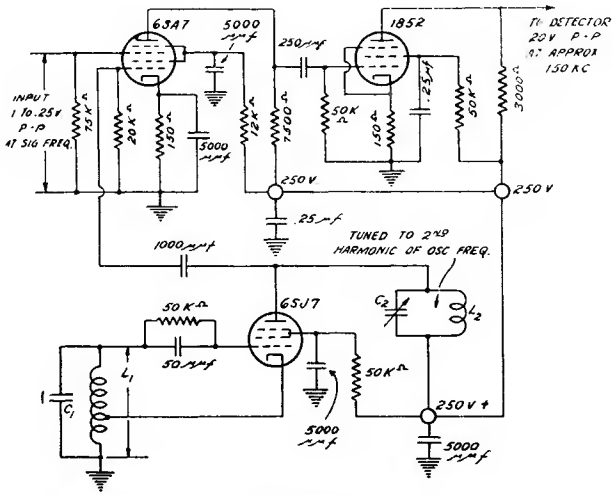


Fig. 3

Experimental Determination of Detector Linearity

One may expect representative linear performance from a monitor detector built according to the data presented. There may, however, be circumstances wherein an accurate knowledge of the capabilities of the device are required. For this reason a brief description of the calibration method employed is included here.

Figure 4 shows a circuit in which the monitor detector (without the heterodyne oscillator of Figure 3) is represented diagrammatically by a box with input and output terminals. A signal generator is employed (in conjunction with a calibrating crystal oscillator) to provide a point-to-point determination of d-c detector output vs. input frequency. An L. & N. Kohlrausch slide wire, fed from a 2-volt storage cell, is used in conjunction with a sensitive galvanometer in a potentiometer circuit which permits the d-c output of the detector to be

determined with a precision of about 0.02 per cent (approximate reading accuracy of the graduated scale of the slide wire).

The extreme sensitivity of the d-c measuring circuit requires the elimination of all sources of diode-current fluctuation. Hence, the *B* supplies feeding the 807 plate and 6AG7 screen are doubly regulated (for calibration purposes only), with the circuit shown in Figure 4, and the diode (d_1) bias obtained by bleeder current (as shown in Figure 2) is replaced by a 1.5-volt battery for determination of detector linearity.

The crystal oscillator of Figure 4 is used to check the calibration of the signal generator, by zero beat methods, with the detecting phones in the plate of the mixer tube. Calibration of the signal gen-

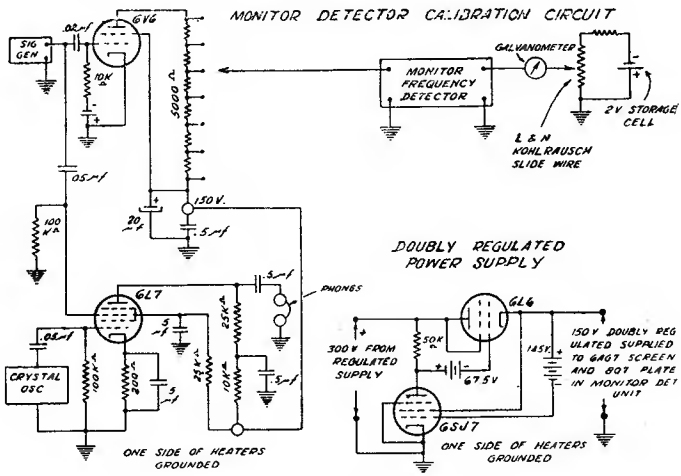


Fig. 4

erator at frequencies which correspond to fractional harmonics of the crystal is possible as a result of curvature in the mixer characteristics.

Measurements of d-c output voltage of the discharge diode in the detector circuit were made at regularly spaced input-frequency intervals. Any non-linearity in the plot of d-c volts vs. frequency was determined by plotting the readings, and comparing the resultant curve with a straight line having approximately the same mean slope. Then a curve of the difference in ordinates of the measured curve and the comparable straight line was again plotted against frequency.

By employing this method, it was found that the departure from linearity of the composite detector characteristic was less than 0.02 per cent over a frequency range of 50 to 250 kc. Thus, if a frequency-modulated wave of 100-kc deviation were applied to the detector at a carrier frequency of 150 kc, the detector would introduce only

approximately 0.02 per cent distortion in the audio output obtained from the detector.

Tests were also made to determine the susceptibility of the detector to amplitude modulation in the input signal. With 25 per cent amplitude modulation (and a mean value of input signal of 20 volts peak to peak) the output contains an undesired signal corresponding to less than 0.05 per cent of the audio voltage produced by 100-kc deviation.

The sensitivity of the detector is approximately one volt peak-to-peak (audio) for 100-kc deviation in the input signal (and 20 volts peak-to-peak applied signal). Therefore, if a single frequency signal of 100 kc is applied to the detector the d-c output should be approximately 0.5 volt. Knowledge of this sensitivity makes possible the adjustment of the heterodyne oscillator frequency (in Figure 3) to a value which causes the beat frequency to lie in the band for which the detector linearity is optimum. It is necessary merely to measure the d-c output of the detector as the heterodyne oscillator frequency is varied, and to adjust the latter to produce a d-c output voltage corresponding to 150-kc carrier input.

Applications

The monitor detector as thus constituted is an excellent instrument for studying the effects of the pass characteristics of a frequency-modulation receiver's r-f and i-f circuits on distortion, for checking the performance of frequency-modulation signal generators, for use as a monitor and peak deviation indicator at a frequency-modulation transmitter, and for many other applications.

Part II—Low Distortion Frequency-Modulation Generator

Introduction

Frequency-modulation signals can be derived from systems based on either frequency or phase modulation principles, provided that the audio potentials (or program material) be properly integrated before application to the phase modulator.

In a frequency-modulation system of the type employing a reactance tube modulator, the frequency deviation produced by a certain audio voltage is independent of the frequency of that audio voltage, and is determined only by the sensitivity of the reactance tube (kc deviation per audio volt), and by the subsequent frequency multiplication. Assuming that the circuits in the generator are sufficiently flat over the desired deviation band, the harmonic content of the output of a frequency-modulation generator of the aforementioned type is deter-

mined principally by curvature in the reactance tube characteristic, and is not a function of audio frequency.

Conditions obtaining in phase-modulation systems are quite different, since, for a given frequency deviation, the required peak-phase deviation varies inversely with audio frequency, i.e.

$$\phi \text{ in degrees at carrier freq.} = \frac{57.3 \times \text{freq. dev. } c/s \text{ at carrier freq.}}{\text{audio-freq. } c/s}$$

Thus, the audio voltage at the phase-modulator grid must vary inversely with audio frequency, to produce constant frequency deviation. If the distortion in the resultant frequency-modulated signal is caused by a non-linear relationship between the instantaneous phase of the generated signal and the instantaneous audio voltage, the distortion will generally vary in some inverse manner with audio frequency. Another source of distortion lies in the inability of the frequency multiplier circuits to pass all the required side components of the modulated signal. This deficiency will generally produce distortion increasing with audio frequency.

The distortion due to non-linearity in the phase characteristic can be reduced by employing frequency multiplication from a relatively low-frequency source of phase-modulated signals, to reduce the required phase deviation at the modulator. Side-band clipping distortion effects are suppressed by employing multiplier circuits having adequate band width.

Principle of Operation of Experimental Phase-Modulation Generator

In most generators employing phase modulation as a means for ultimately producing frequency-modulated signals, modulation by the audio components or program material is accomplished at a low value of signal frequency (generally about 100 kc), which is subsequently multiplied and heterodyned to produce the desired carrier frequency.

One system of phase modulation employed experimentally requires first the generation of amplitude-modulated signals at a low signal frequency, after which the carrier is shifted in phase with respect to the amplitude-modulated sidebands, to produce a phase-modulated signal. The inherently remanent amplitude modulation present in this developed signal is then removed by limiter action.

Another system of phase modulation is described in Patent No. 2050067, issued August 4, 1936 to Dr. Walter Van B. Roberts. This system makes use of an unique characteristic of a simple parallel tuned

circuit operated at $1/\sqrt{2}$ times its resonant frequency. An investigation of the operation of this circuit together with following frequency multipliers was made with the aid of the afore-described frequency-modulation monitor. The basic principle of operation of the circuit is as follows: A parallel resonant circuit, tuned to $\sqrt{2}$ times the frequency of an unmodulated low-frequency carrier (generally about 100 kc) is used as the plate load in a pentode whose grid is fed with voltage at the source frequency (100 kc). See Figure 5. The phase of the plate voltage of the pentode is then varied by changes in the series resistance of the resonant circuit. Figure 6 shows the manner of variation of ϕ with R; between the limits of $R = 0$ and $R = \infty$ the total phase change is 180°. One of the important features of the system is that the absolute value of the plate-load impedance is constant and

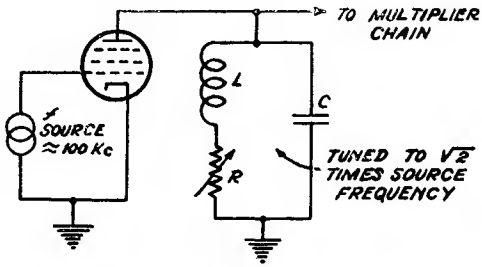


Fig. 5

equal in magnitude to the reactance of the circuit capacitance, for all values of R from zero to infinity. Hence, no amplitude modulation is present, regardless of the extent or amount of phase modulation. Appendix I contains a development of the equations of phase vs. R, and also a proof of the constancy of plate-load impedance as R is varied.

In practice, R in Figure 5 is the dynamic plate impedance of a tube, whose control grid is supplied with audio-frequency potentials. Since the variation of ϕ with R is non-linear, as shown by Figure 6, a linear relationship between the instantaneous phase of the pentode's plate voltage and the instantaneous value of modulator grid voltage is obtained only over an interval for which a curve of Eg vs. R_p (for the modulator tube) has the same shape and curvature of that of Figure 6. The problems associated with matching the curves of ϕ vs. R and R_p vs. Eg will be discussed in greater detail later.

The phase-modulated voltage (at a carrier frequency of 100 kc) is fed from the tuned circuit of Figure 5 through a chain of harmonic multipliers to develop a frequency-modulated or phase-modulated signal

at the desired carrier frequency. The phase deviation is, of course, multiplied by the same factor as is the center frequency.

The phase deviation required at the modulator is determined by the desired phase deviation and by the order of frequency multiplication. Thus, an expression for modulator phase deviation (at the 100 kc level) for any audio frequency is

$$\phi^\circ = \frac{57.3 \times \text{freq. deviation at ultimate carrier freq.}}{\text{audio frequency} \times \text{order of frequency multiplication}}$$

It is desirable to maintain the maximum required phase deviation in the modulator circuit at as low a value as possible, to minimize non-

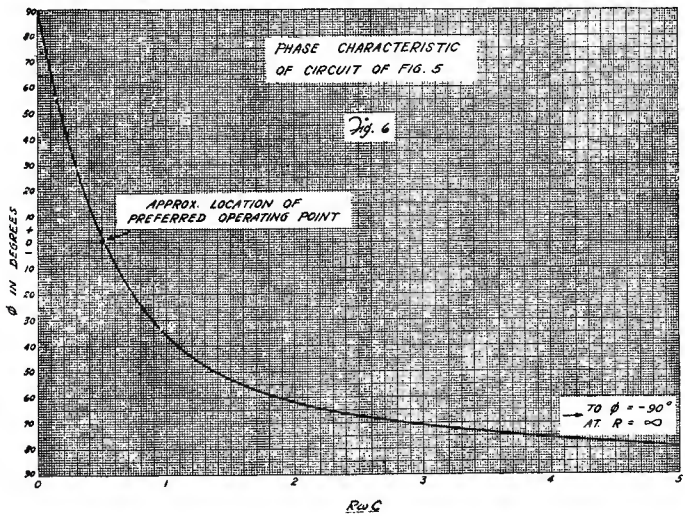


Fig. 6

linear effects at low audio frequencies (which require the greatest phase deviation for a given frequency deviation). Two methods of minimizing this low-frequency-modulator phase deviation are available. The first involves starting at a reduced value of signal frequency (lower than 100 kc) to obtain 2 or 3 times additional frequency multiplication, thereby reducing the required maximum modulator phase deviation by a factor of 2 or 3. The limiting low value of source frequency must be greater than the highest audio-modulating frequency.

A better way to reduce the low audio-frequency phase shift required in the modulator is to employ a heterodyne action at some point in the multiplier chain, by which the phase-modulated signal is heterodyned from say, 1500 kc, down to a lower frequency, after which additional multiplication is applied. This permits the realization of practically

unlimited multiplication of the phase deviation without resorting to very low source signal frequencies in the modulator circuit.

The heterodyne system was employed in the laboratory generator, as shown in the block diagram, of Figure 7. The source of signals is a crystal oscillator operating at 100 kc. This is fed through a filter to a 6SJ7 with the tuned circuit of Figure 5 as its plate load.

The grid of the 6L6 modulator tube (triode connected) is supplied with audio potentials, and its plate resistance constitutes the variable R of Figure 5. The resultant phase-modulated signal is multiplied successively by factors of 3 and 5 to produce a 1500-kc voltage, whose phase deviation is 15 times that effected by the modulator. In a

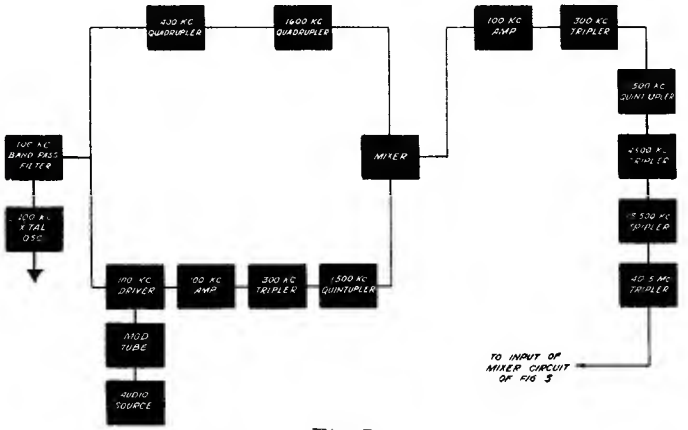


Fig. 7

parallel chain of multipliers, an unmodulated 100-kc signal (from the same crystal oscillator) is multiplied by factors of 4 and 4, to produce a 1600-kc voltage, with no phase deviation. The 1600-kc signal serves as a constant-frequency heterodyning source, converting the phase modulated 1500-kc signal to the difference frequency of 100 kc, which is taken from the mixer plate through a low-pass filter. The phase deviation in this 100-kc signal is increased over the modulator circuit deviation by a factor of 15. Hence, the apparent sensitivity of the modulator (degrees phase shift/audio voltage) is increased by 15 to 1, allowing a reduction of non-linear distortion products (at low values of audio frequency) by a factor greater than 15 to 1.

The 100-kc heterodyne signal is then fed through a subsequent chain of multipliers which raises the carrier frequency to 40.5 Mc, with a phase deviation of $405 \times 15 = 6075$ times that produced at the modulator. For distortion measurements this output signal is applied to the converter of Figure 3, which feeds into the linear monitor detector.

Figure 8 is a diagram of the phase-modulator circuit employed in the laboratory generator. The component circuits indicated by the blocks in Figure 7 are, in general, self-explanatory, but a few remarks determined from development experience are in order.

The coupling transformers for the 300 kc, 400 kc, 1500 kc, and 1600 kc stages are double-tuned transformers employing inductive coupling. The coupling and damping of each unit was adjusted to provide adequate band-pass characteristics. In this respect it must be noted that those stages following the heterodyne mixer require greater percentage band width than those preceding the mixer by a factor equal to the heterodyne ratio (15 in this case).

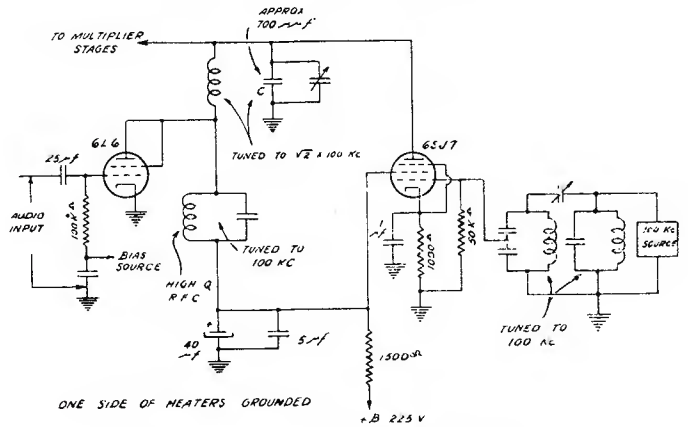


Fig. 8—Phase modulator circuit.

The 4.5 Mc, 13.5 Mc, and 40.5 Mc stages utilize single resonant circuits, each tuned with a 3-30 $\mu\mu f$ trimmer and all loaded or damped sufficiently to make the pass band essentially flat and thus to minimize the generation of amplitude modulation and phase distortion in these stages.

Jacks are provided in each stage to facilitate the measurement of d-c grid current, as a means for determining correct tuning conditions in preceding stages. The grid bias resistors are adjusted to produce maximum harmonic output in each stage. Thus, for example, in a tripler whose plate load is tuned to 300 kc, the tripling action is maximized by adjusting the tripler's grid leak to produce maximum grid current in the following tube. This bias adjustment is carried out in all stages. The grid-circuit time constants of all stages are made sufficiently low to permit the multiplier grids to follow any a-f variations in the amplitude of the signals, thereby producing limiting action..

Experimental Procedure

An essentially distortionless audio source is required for measurements of phase distortion in the frequency-modulation generator. This happens because the amplitude of the nth harmonic in the developed frequency deviation is proportional not only to the amplitude of the nth harmonic in the audio-source voltage, but also to the order of the harmonic, since the frequency deviation is proportional to the *rate* of phase deviation. Thus, 1 per cent 3rd harmonic in the audio input would produce (in an otherwise distortionless system) 3 per cent harmonic in the resultant frequency deviation, which would then appear as 3 per cent 3rd harmonic in the audio output of the monitor detector.

This problem may be circumvented in two ways, i.e., first integrate the audio voltage before application to the modulator, in which case the device produces frequency-modulated signals, requiring constant audio input for constant-frequency deviation as the audio frequency is varied. The source harmonics are then reduced in magnitude by an amount proportional to their order. The alternative method was used in the experimental work; an audio source of good waveform was used in conjunction with a series of band-pass filters, each capable of passing an octave of frequencies, starting at 30 cycles and ending at 8000 cycles. No integration was employed: hence, the generator produced phase-modulated signals.

A General Radio Type 636-A wave analyzer was used for all distortion measurements. The harmonic content of the audio source was found to be less than 0.05 per cent for all audio frequencies employed. The wave analyzer was then connected permanently across the monitor-detector output terminals. In parallel with this a 9-inch oscilloscope was connected to facilitate preliminary adjustments. In general the audio input to the modulator was adjusted to produce 100-kc deviation (for all audio frequencies) at the final carrier frequency of 40.5 Mc. The heterodyne oscillator in the detector unit was adjusted to produce 150 kc as the intermediate frequency for application to the monitor detector. This had previously been ascertained as lying in the region of optimum detector linearity.

As pointed out in a preceding section, there are two sources of distortion, in a generator of this type; the first occurs at low audio frequencies (less than 400 cycles) and is due to non-linear relationships in the modulator, which are accentuated by the large audio swings required.

This type of distortion is generally a function of the modulator bias, which controls the degree of matching between the curves of ϕ vs. R (Figure 6) and R_p vs. E_g of the modulator tube.

At higher audio frequencies (greater than 1000 cycles) the distortion is generally independent of modulator bias (since only a small audio swing is applied), but is determined, rather, by sideband clipping in the multiplier tuned circuits.

In adjusting the unit for minimum distortion products the audio input was set at 3200 cycles, and, with the monitor detector output indicating 100-kc deviation, and the wave analyzer reading 2nd harmonic distortion, adjustments were made by loading and tuning of the multiplier-load circuits until the distortion due to sideband clipping was minimized. This procedure was repeated at 7500 cycles. The overall distortion was reduced to less than 0.5 per cent rms, for 100-kc deviation, at these higher audio frequencies.

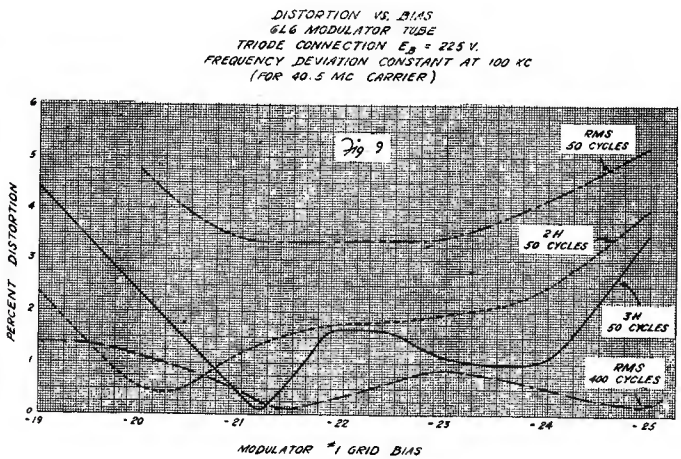


Fig. 9

The audio frequency was then reduced to 400 cycles, for distortion measurements at relatively larger modulator audio swings. With the frequency deviation adjusted for approximately 100 kc and the wave analyzer set for 2nd harmonic determinations, it was found that the amplitude of this harmonic was a function of modulator bias, and that a minimum value of less than 0.1 per cent could be attained with critical modulator bias adjustments. This sharply defined minimum is due to the dependence of the localized curvature of the R_p-E_g curve of the modulator upon its bias, and to the fact that, for a certain bias, correct matching between this curve and that of ϕ vs. R is obtained. Figure 9 shows curves of rms distortion vs. bias at 400 cycles modulating voltage.

At 50 cycles the applied audio voltage required for 100-kc deviation is 8 times that required at 400 cycles, and as a result, the degree of matching of the E_g vs. R_p and ϕ vs. R curve is less complete over an

audio cycle. However, the rms distortion at 50 cycles is relatively constant at 3.5 per cent over a 2-volt range in bias, for 100-kc deviation, as shown in Figure 9.

A large number of tube types (6J7, 6K7, 6V6, 6AC7/1852, 6AC7/1853, 6J5, 6L6) were used experimentally as modulators. The 6AC7/1852, for example, has higher sensitivity (kc deviation per audio volt), but the dependence of distortion on bias is more critical. The 6L6 was finally chosen as a result of its lower rms distortion at 50 cycles modulating frequency, and the relative constancy of distortion with bias.

Analysis of the modulator circuit shows that distortion produced at low audio frequencies, caused by mismatch of the R_p vs. E_g and ϕ vs. R curves, is due principally to insufficient curvature in the R_p vs.

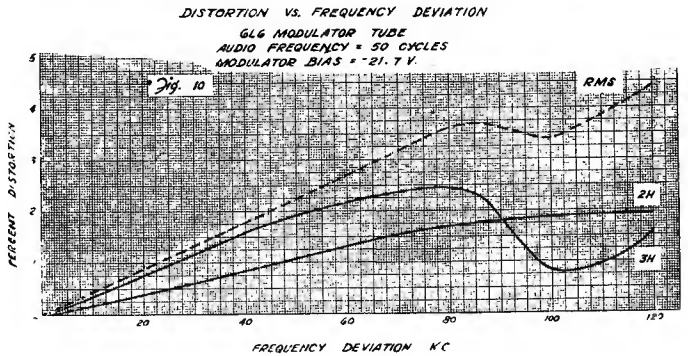


Fig. 10

E_g curve. This can be corrected, if desired, by some types of regenerative audio circuits, which tend to increase non-linearity in the modulator tube R_p curve, or by the use of an additional modulator to supply 2nd harmonic variations in R_p to produce additional curvature.

Experimental Results

Figure 10 shows the distortion incurred at 50 cycles modulating frequency vs. frequency deviation (in the 40.5 Mc carrier) taken at a value of bias for which the 2nd harmonic distortion is at a broad minimum.

Figure 11 is a plot of rms distortion vs. audio frequency for two values of frequency deviation. Note that the distortion is essentially independent of deviation except at frequencies below 200 cycles, for which the audio swings (approximately 20 peak-peak volts at 50 cycles) are large, causing considerable mismatch in the phase and R_p curves, which, at lower swings, are normally coincident.

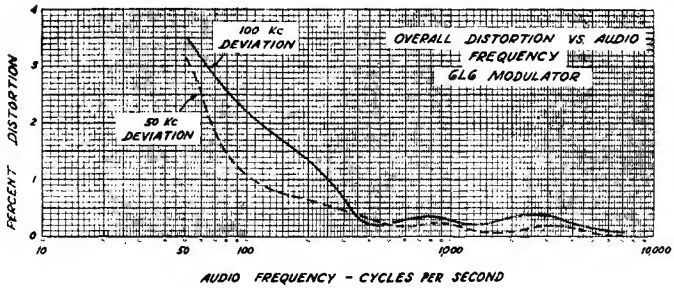


Fig. 11

The distortion at high audio frequencies is independent of modulator bias, but is very much a function of tuning, even with sufficient circuit damping. This is particularly the case in the 300-kc tripler following the mixer stage, which must accommodate a band of frequencies 15 times as great as that encountered in circuits directly following the modulator.

Attention must be directed toward minimizing amplitude modulation in the frequency-modulated wave, particularly at the lower signal frequencies, for which the phase deviation is small. As shown in Appendix II, n per cent amplitude modulation can produce n per cent 2nd harmonic distortion in the frequency deviation, due to sideband cancellations, if the phase deviation is sufficiently low to permit the representation of the wave by a carrier and a pair of first-order sidebands.

Appendix I

Analysis of Phase Modulator Load Circuit:

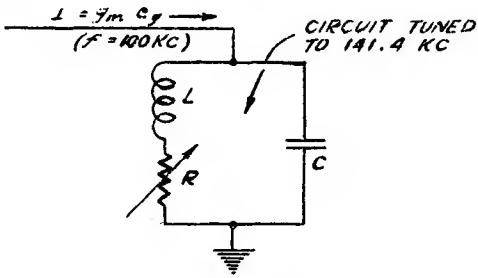


Fig. (a)

(1) To demonstrate the constancy of tuned impedance as R is varied. $\omega_s = 2\pi \times$ input frequency (100 kc)

$$Z = \frac{(R + jL\omega_s)\dfrac{1}{jC\omega_s}}{R + j\left(L\omega_s - \dfrac{1}{C\omega_s}\right)} = \frac{(RC\omega_s + jLC\omega_s^2)\dfrac{1}{jC\omega_s}}{RC\omega_s + j(LC\omega_s^2 - 1)}$$

The circuit is resonant at $\sqrt{2} \times$ the source frequency, i.e., C in the above figure is half the value required for circuit resonance at the source frequency. Therefore, $LC = \dfrac{1}{2\omega_s^2}$

and

$$\not{Z} = \frac{\left(RC\omega_s + j\dfrac{\omega_s^2}{2\omega_s^2}\right)\dfrac{1}{j\omega_s C}}{RC\omega_s + j\left(\dfrac{\omega_s^2}{2\omega_s^2} - 1\right)} = \left(\frac{2RC\omega_s + j1}{2RC\omega_s - j1}\right)\frac{1}{j\omega_s C}$$

The absolute magnitude of $\not{Z}$ is therefore $|\not{Z}| = \dfrac{1}{\omega_s C}$, which is independent of R.

(2) The phase vs. R equation of the circuit is readily determined from the above, viz:

$$\not{Z} = \left(\frac{2RC\omega_s + j1}{2RC\omega_s - j1}\right)\frac{1}{j\omega_s C} = \frac{1}{\omega_s C}\left(\frac{1 - j\,2RC\omega_s}{2RC\omega_s - j1}\right)$$

total phase shift ϕ = phase shift in numerator θ minus phase shift in denominator α

$$\phi = \theta - \alpha$$

$$\phi = \tan^{-1}(-2RC\omega_s) - \tan^{-1}\left(\frac{-1}{2RC\omega_s}\right)$$

$$\tan\phi = \frac{\tan\theta - \tan\alpha}{1 + \tan\theta\tan\alpha}$$

Hence,

$$\tan\phi = \frac{-2RC\omega_s + \dfrac{1}{2RC\omega_s}}{1 + \dfrac{2RC\omega_s}{2RC\omega_s}} = \frac{1}{2}\left(\frac{1}{2RC\omega_s} - 2RC\omega_s\right)$$

$$\phi = \tan^{-1}\left(\frac{1}{4RC\omega_s} - RC\omega_s\right)$$

This is plotted in Figure 6

APPENDIX II

Proof that amplitude modulation in a low-deviation phase-modulated wave can produce frequency-modulation distortion equal approximately to the per cent amplitude modulation, if this factor is small.

First assume that the phase deviation is sufficiently low to justify the use of only the first-order sidebands. The signal then has the form of Figure (b).

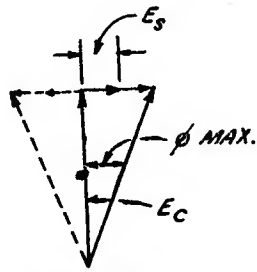


Fig. (b)

$$\phi = \phi_{max} \sin \omega_a t$$

With no added amplitude modulation

$$\phi = \tan^{-1} \frac{2E_s}{E_c} \sin \omega_a t$$

where E_s = phase-modulation sideband amplitude
E_c = carrier-voltage amplitude
$\omega_a = 2\pi \times$ modulating frequency

Now let the carrier be amplitude modulated at the same audio frequency, viz:

$$e_c = E_c(1 + m \sin \omega_a t)$$

The phase angle is given by

$$\phi = \tan^{-1} \frac{2E_s \sin \omega_a t}{E_c(1 + m \sin \omega_a t)}$$

Let $\frac{2E_s}{E_c}$ be a constant, determined by the amount of phase modulation.

Then

$$\phi = \tan^{-1} K \frac{\sin \omega_a t}{1 + m \sin \omega_a t}$$

For small angles it is permissible to write

$$\phi = K \frac{\sin \omega_a t}{1 + m \sin \omega_a t}$$

also for m, (per cent amplitude modulation $\div$ 100) less than 0.10, it is sufficiently accurate (within 1 per cent) to write $\phi = K \sin \omega_a t\ (1 - m \sin \omega_a t) = K \sin \omega_a t - Km \sin^2 \omega_a t$. This can be expanded to give

$$\phi = K\ (\sin \omega_a t - \frac{m}{2} + \frac{m}{2} \cos 2\omega_a t)$$

The ratio of coefficients of the 2nd harmonic and fundamental audio terms is $m/2$, which implies the existence of $\frac{m}{2} \times 100$ per cent 2nd harmonic in the phase-modulated wave.

The frequency deviation is equal to $\frac{1}{2\pi} \frac{d\phi}{dt}$

Performing this operation we get

$$\frac{d\phi}{dt} = K\ (\omega_a \cos \omega_a t - 2\omega_a \frac{m}{2} \sin 2\omega_a t)$$

$$\Delta f = \frac{1}{2\pi} \frac{d\phi}{dt} = Kf_a\ (\cos \omega_a t - m \sin 2\omega_a t)$$

where Δf is the frequency deviation (instantaneous) and f_a is the modulating frequency. It is seen that m per cent amplitude modulation will therefore produce m per cent 2nd harmonic distortion in the frequency deviation of the system.

A NEW EXCITER UNIT FOR FREQUENCY-MODULATED TRANSMITTERS*†

By

N. J. Oman

Engineering Department, RCA Victor Division
Camden, N. J.

Summary—*This paper describes an exciter unit for use in frequency-modulation transmitters. The new exciter unit is capable of producing a frequency modulated carrier of excellent linearity and low noise level. The carrier frequency is automatically maintained to an accuracy close to that of a crystal-controlled oscillator, the frequency of the latter being used as a reference.*

THERE are two systems now in general use for producing a frequency modulated signal which meets the Federal Communications Commission's requirements for frequency modulation broadcasting. One method, making use of phase-shift modulation, was developed by Major E. H. Armstrong. The operation of this system is shown in schematic form in Figure 1.

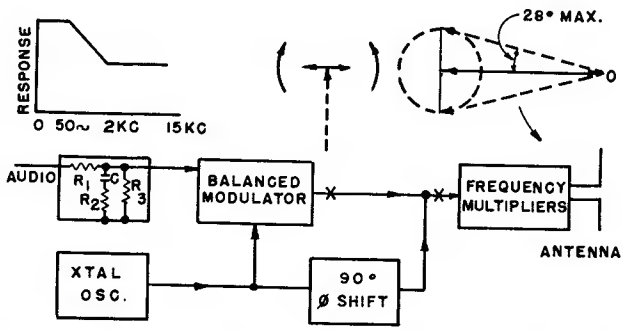


Fig. 1—Operation of Phase-Shift Modulator, Armstrong Type FM Transmitter.

The crystal frequency and the audio input are fed to a balanced modulator. The balanced modulator is so constructed that only the sidebands appear in its output. The sidebands are pictured at the output of the modulator as two small vectors rotating in opposite directions. The relation of the carrier, although not actually present, is

* Decimal Classification: R423.8
† Reprinted from *RCA REVIEW*, March, 1946.

shown dotted in the proper phase relation. A portion of the output of the crystal is shifted in phase by 90 degrees, and added to the sidebands, as shown in the top right of the illustration. The effect of the sidebands is to advance and retard alternately the position of the carrier vector. The whole picture rotates about the point *O* at an average angular velocity determined by the crystal frequency. The magnitude of frequency modulation produced depends on the modulating fre-

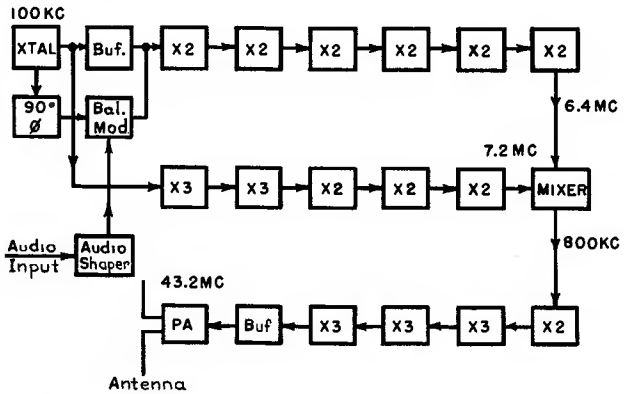


Fig. 2—Block Diagram of Armstrong Type FM Transmitter.

quency, and on the amplitude or length of the sideband vectors relative to the reintroduced carrier. It can be seen that, if the length of the sideband vectors is fixed, the frequency modulation will increase six decibels per octave as the modulation frequency is increased.

In order to obtain pure frequency modulation it is necessary to decrease the amplitude of the audio input six decibels per octave, with rising frequency. This is shown in the response curve, at the upper left, in the modulating frequency range from 50 to 2,000 cycles. The audio response is flat for frequencies above 2,000 cycles. This means that the frequency-modulated output increases at six decibels per octave above 2,000 cycles, to give the 75 micro-seconds standard pre-emphasis of high frequencies agreed upon by the Radio Manufacturer's Association, as a measure to improve signal-to-noise ratio at the receiver. The receiver has a corresponding drop in response above 2,000 cycles to produce a flat response for the overall system. The audio response is again flat for frequencies below 50 cycles, resulting in six decibels per octave drop in frequency-modulated output below 50 cycles. This restriction is imposed by the fact that the distortion in the frequency-modulated output is five per cent, when the phase displacement is plus or minus 28 degrees, and increases rapidly with greater phase shifts. These

representative figures call for a frequency multiplication of about 3,000 to produce a frequency swing of plus or minus 75 kilocycles at 50 megacycles.

Figure 2 is a more complete block diagram of an Armstrong type of transmitter making use of an arrangement to minimize radio carrier drift.

A second method of producing a frequency modulated signal employs a reactance tube to produce a direct change in the frequency of a master oscillator. The basic circuit is shown in Figure 3. The oscillator circuit is illustrated at the right and the reactance modulator is shown at the left. The plates of the two tubes are tied together. The grid of the reactance tube is driven from the plate circuit, but is 90 degrees out-of-phase with it. The plate current of the reactance tube is therefore 90 degrees out-of-phase with respect to the plate voltage. The tube reactance becomes either capacitive or inductive in nature, depending on whether the 90-degree shift in grid excitation

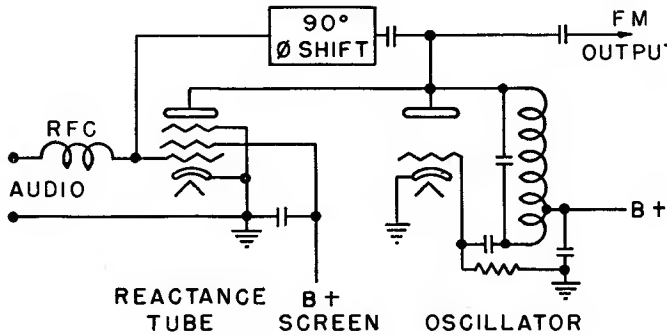


Fig. 3—Reactance Tube.

leads or lags the plate voltage. The audio voltage changes the bias of the reactance tube grid in accordance with the signal voltage. As the reactance tube is a pentode, this change in bias causes a corresponding change in mutual conductance of the tube. The net result is that the reactive current of the tube varies as the impressed audio signal. In a high-fidelity circuit it is customary to use two reactance tubes, one acting as a capacitive reactance and the other as an inductive reactance, in order to obtain maximum frequency stability and linearity of response. The circuit is capable of a frequency swing of more than plus or minus 10 kilocycles at 5 megacycles so that very little frequency multiplication is required to obtain a plus or minus 75 kilocycles swing.

Because of the advantages of the reactance tube modulator, considerable effort has been devoted towards obtaining such an exciter

unit that would also have excellent frequency stability. The frequency correction should be effected independent of the modulation process, and, if possible, provision should be made to correct frequency manually if there should be a failure of the automatic frequency control. A good crystal-controlled oscillator should be used as a standard reference frequency source for the automatic frequency-control circuit.

One type of frequency modulation exciter (illustrated by the block diagram of Figure 4) beats the master oscillator frequency down to one megacycle, using a crystal oscillator as the beat-frequency source. The one-megacycle beat frequency acts on a discriminator to provide a direct current control potential which is applied to a reactance modulator tube grid so as to compensate any drift of the master oscillator frequency. The time constant of the correction is of sufficient duration to prevent it responding to modulation.

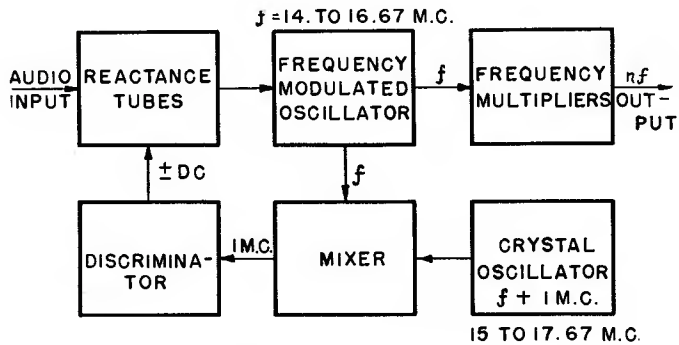


Fig. 4—Reactance Tube FM Exciter.

Certain points, in regard to this method of frequency control, however, should be mentioned. There is fluctuation in the direct current control potential from the discriminator caused by variations in contact potential in the diodes. Circuit components change in value with changes in temperature; therefore the discriminator circuit and reactance tubes have to be placed in an oven, as illustrated in Figure 5, to provide an accurately controlled temperature. The functions of modulation and frequency control are performed by the same reactance tubes, which does not allow optimum adjustment of the reactance tubes for minimum distortion. If the modulating signal is unsymmetrical, there is a change in control potential from the discriminator, causing a shift in oscillator frequency. The frequency correction is degenerative in action, and therefore acts only to reduce the magnitude of frequency drift of the master oscillator.

Data on many other circuits intended to accomplish automatic frequency control have been published. A number of these make use of the scheme shown in Figure 6. The signal from the oscillator to be controlled is fed to two amplitude-modulation detectors or mixers. The signal from the crystal-controlled reference oscillator is also fed to the detectors, but through circuits that shift phase so that there is a 90-degree displacement of the reference frequency between the two detectors. The resulting beat frequency of the two oscillators will appear in the output of each detector, and because of the 90-degree phase shift of one of the exciting frequencies, the beat frequencies will also be displaced by 90 degrees. The interesting feature of the circuit is that,

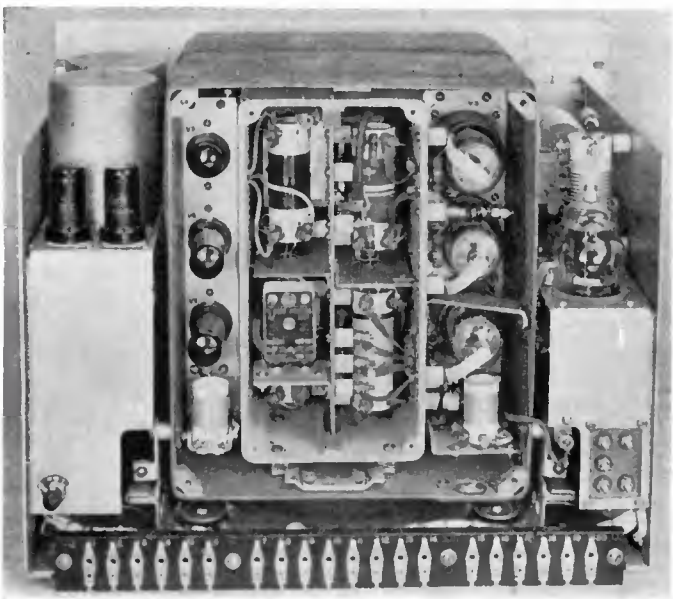


Fig. 5

as the frequency of the oscillator, being compared to the frequency of the crystal, is caused to change through the synchronous or zero-beat condition, there is a reversal in phase of one beat-frequency output with respect to the other.

There are many ways in which to make use of this phase reversal to provide an indication as to whether the controlled oscillator frequency is high or low. Several methods were investigated. The most promising idea was to make use of the fact that the two beat notes displaced by 90 degrees constitute a source of two-phase power and

could, therefore, be used to set up a rotating magnetic field in an induction motor. Inasmuch as one phase reverses when the controlled oscillator frequency passes through zero beat, the rotation of the field

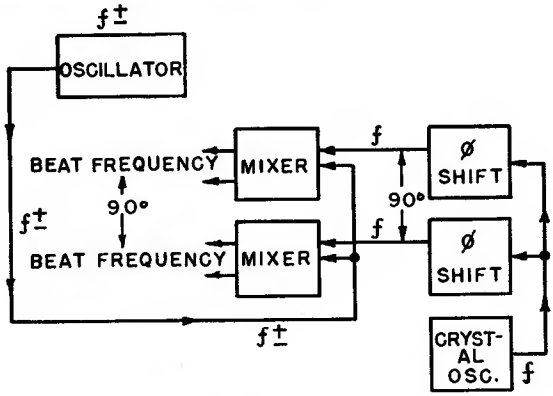


Fig. 6—Fundamental Frequency Control Circuit

and therefore of the motor, will reverse. The motor can be used to drive a variable condenser or other tuning means to bring the controlled oscillator in step with the crystal oscillator as shown in Figure 7. This

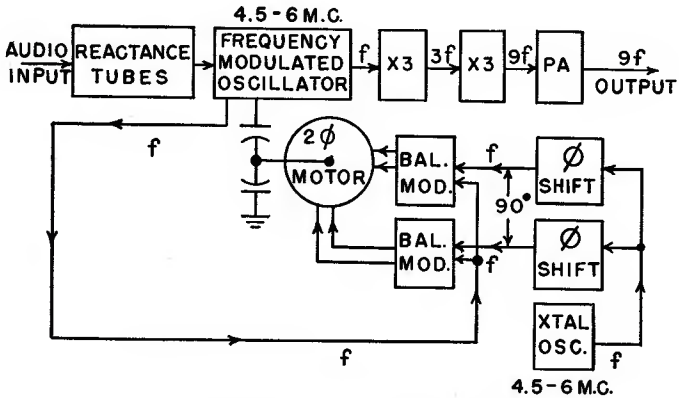


Fig. 7—Preliminary Frequency Control FM Exciter

scheme was tried in just this form. The induction motor developed torque enough to run well at frequencies from a few cycles per second to two or three hundred cycles per second. This does not constitute a very wide range of control at five megacycles. On a cold start it was necessary to change the frequency of the controlled oscillator so that the

beat frequency would fall in the range where the motor had torque enough to take control. Frequency modulation of the controlled oscillator had no apparent effect on the motor except to reduce its torque. The system was operated with program modulation from a local broadcast station and ran for several days without once losing control. There was, however, a constant danger that a sudden jump in oscillator frequency, especially in the presence of modulation, might cause loss of control. Some improvement could be obtained by speeding up the rate of frequency correction. In these first tests it had been necessary to drive the control condenser through a gear train so that the frequency correction would be gradual and without overshooting, which would result in hunting.

The slow speed of the motor, added to the backlash and friction in the gear train, gave rather poor performance in following minor de-

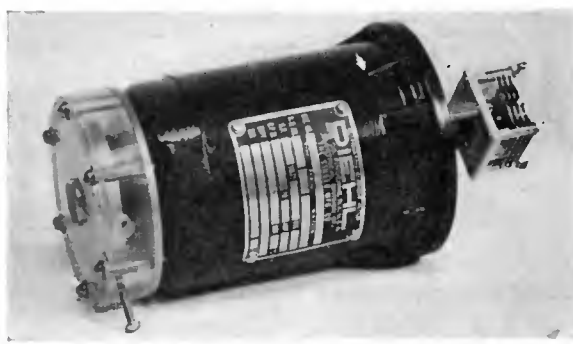


Fig. 8

partures in frequency. To eliminate some of these troubles, the condenser was mounted directly on the motor shaft as illustrated in Figure 8. The tendency to hunt was eliminated by fitting a dash-pot to the opposite end of the motor shaft. The friction of the tuning condenser bearings was avoided by using a split stator condenser with the rotor on the motor shaft and insulated therefrom. One set of stator plates was grounded and the other set was connected to the plate of the controlled oscillator. This procedure eliminated all friction save that in the ball bearings of the motor. The range of operation was, thereby, extended at both high and low frequencies, the motor being able to take control at 1,000 cycles above or below zero-beat. The oil damping gave much more rapid control action without causing trouble from hunting. These changes yielded an enormous improvement in the operation of the frequency control.

Since the limit of the control range is the high-frequency-response limit of the motor, an increase in the range of control can be obtained by dividing the modulated oscillator frequency, so that it could be compared to a crystal oscillator at a lower frequency. The range of control would thus be increased by a factor equal to the division in frequency. An exciter unit, such as described by J. F. Morrison in his U.S.P. #2,250,104, accomplishes this by using ten stages, each divided by two, yielding a total division of 1024. Each divider stage is a duplication of the one schematically shown in Figure 9. The circuit employs the principle of regenerative-modulation whereby a subharmonic is obtained by a modulation process. Since the output energy is obtained by a modulation process involving both the input and output waves, the output wave will appear only when an input wave is applied,

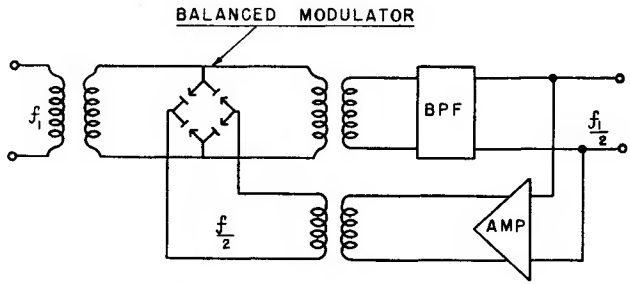


Fig. 9—Frequency Division Pat. By. R. L. Miller 2,159,595

and bears a fixed frequency ratio with respect to it. The circuit makes use of copper-oxide modulators and the output circuit is tuned to one-half the input frequency. The tuned circuits are made sufficiently broad to permit substantial output voltage over a frequency range of ± 1.5 per cent. While the circuit is rather complicated and expensive, the output waveshape is good and the fact that there is no output wave when there is no input signal, might prove to be of value. The division per stage is low, although division ratios greater than 2 are possible under special circumstances.

Another possible way to obtain frequency division would be to make use of multivibrators. The waveshape would not be good, but it would be possible to get higher division per stage.

The circuit finally adopted for use in the new exciter unit is shown in Figure 10. This circuit minimizes the number of divider circuit components. The waveshape of the output of the divider, described in greater detail by G. L. Beers in his U.S.P. #2,356,201, has been found

to be entirely satisfactory for automatic frequency control purposes. The circuit can be tuned over the required range by means of an adjustable iron slug located within the field of the coil. The lock-in range may be as high as plus or minus five per cent. Division ratios as high as 12 have been successfully obtained.

Having selected a circuit for frequency division the next consideration was to fix the frequency at which the comparison of frequencies was to be made. A crystal and holder with excellent performance designed to operate in the frequency range of 90 to 125 kilocycles is used as the frequency standard for the exciter units.

A preliminary test of the performance of the frequency control with frequency comparison, made at 100 kilocycles, gave excellent results except for a loss of control in the presence of tone-frequency modulation

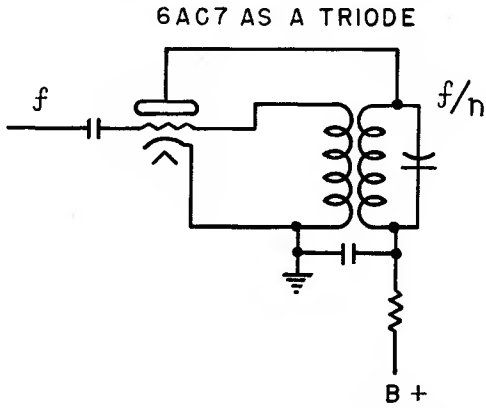


Fig. 10—Locked-In Oscillator Frequency Divider

when the modulating frequency was below 100 cycles. A fundamental characteristic of frequency modulation is that the energy in the modulated signal does not change with the degree of modulation. However, as the amount of modulation is increased from zero, there is a transfer of power from the carrier to the sidebands lying above and below the carrier frequency. This effect continues with increasing modulation until a point is reached at which all the energy is in the sidebands and there is no power in the carrier. With further increase in modulation, or frequency swing, the carrier again reappears. With the ± 75-kilocycle swing of frequency, used for broadcasting, there are several cycles of this effect at low modulating frequencies. In the first experiments, where the frequency comparison was made at five megacycles, this phenomenon was detected only in a loss of torque in the tuning motor.

Theoretically, it should have been possible to find conditions of modulating frequency, and degrees of modulation, where there was no torque. Actually, these points were so critical and so sharply defined that it required considerable patience to find them. In actual operation, with program modulation, the condition of no torque would not exist long enough to lose control.

The effect of frequency division is to reduce the swing of frequency with modulation, along with the reduction of carrier frequency. This is carried far enough when the frequency comparison is made at 100 kilocycles so that carrier loss occurs only at frequencies of modulation below 100 cycles and, for 100-cycle modulation, only with full frequency swing or 100-per-cent modulation. This effect was not serious in program modulation, but in taking performance data on a transmitter,

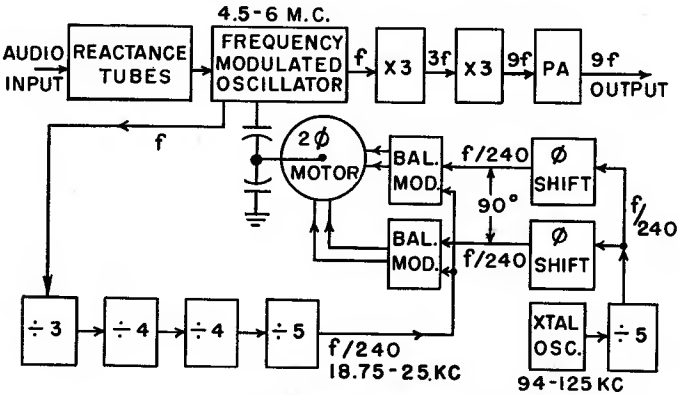


Fig. 11—New FM Exciter

with tone modulation, it would be readily apparent. An additional division by five, to 20 kilocycles, eliminates this problem except for modulation below 20 cycles, which appears low enough not to be a source of complaint.

The circuit of the current model of the exciter unit is illustrated in Figure 11, as a block diagram. The items in the top row are (from left to right) the reactance tubes, a modulated oscillator, a frequency tripler, a second tripler, and a power amplifier. The power amplifier is used to obtain sufficient power to feed the main transmitter through a transmission line. The output frequency will fall in the range of 40.5 to 54 megacycles. Multiplication by two, in the main transmitter, gives a possible output frequency range from 81 to 108 megacycles.

A lead shown at the left of Figure 11 serves to conduct a synchronizing voltage from the modulated oscillator to the first divider at the

lower left. The dividers are arranged as shown, with four stages, giving a total division of 240. This places the output frequency of the last divider in the range of 18.75 to 25 kilocycles. The output of the last divider is connected directly to the two balanced modulators. The crystal oscillator shown at the lower right may operate at any frequency between 94 and 125 kilocycles. The crystal output synchronizes a divider at one-fifth the crystal frequency. This frequency is also fed to the balanced modulators, but in this case a phase-shifting network is included in the lead to each modulator, adjusted to maintain a 90-degree displacement in-phase between the modulators, over the range of frequencies involved.

Each balanced modulator has a pair of 6L6 tubes biased to cut-off, and connected in push-pull. An induction motor with high-impedance

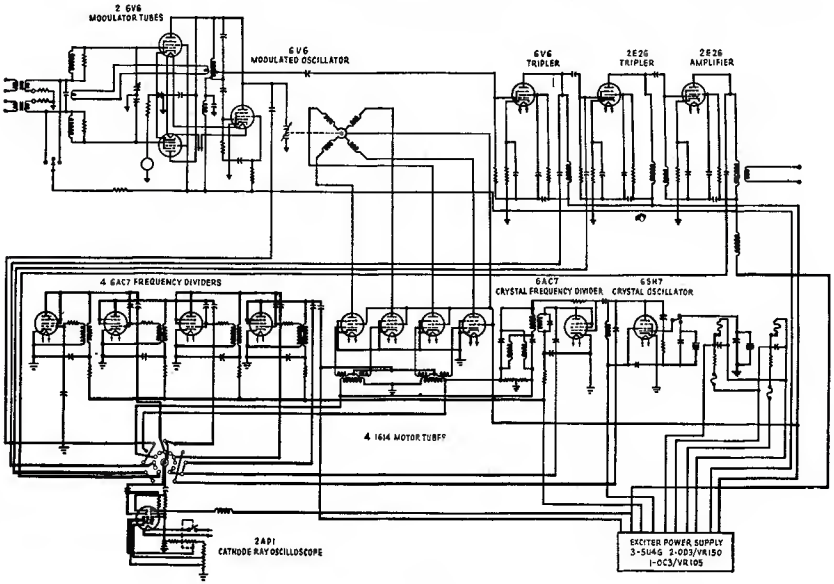


Fig. 12

center-tapped windings on each phase is used in the plate circuit of the modulator tubes, the use of matching transformers then being unnecessary. In this way the motor receives full voltage to direct current beat-frequency. The need for this is evident when one considers that the motor must respond to a beat-frequency lower than one part in a million at 20 kilocycles, which is .02 cycles per second or 1.2 cycles per minute. This rather phenomenal performance from an induction motor is largely made possible by the elimination of any load on the motor. The absence of gearing and the use of viscous damping establishes a condition in

which there is little or no resistance to slow rotation of the motor shaft. This motor responds on application of voltage to frequencies up to 1000 cycles.

In order to make tuning and performance checking of the frequency control simple and rapid, the necessary test equipment is built into the exciter. A cathode-ray oscilloscope, with the required switching mechanism, is provided. By selecting the proper switch position it is possible to check the division in each divider and also the tuning and multiplication of the tripler stages, by means of Lissajous figures.

The accuracy of the frequency control is so exact that it is limited by the heat cycle of the crystal oven. During test, this effect produced

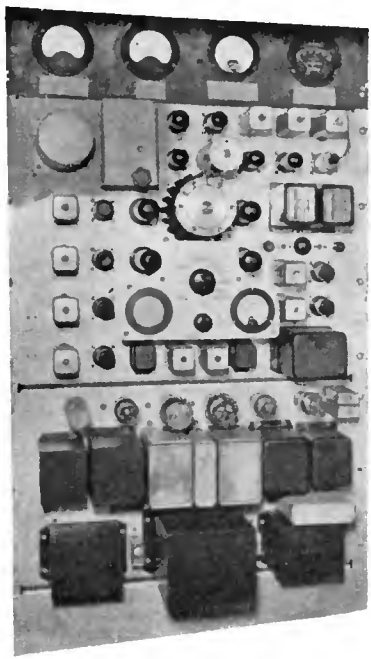


Fig. 13

a regular variation of ± 40 cycles at 100 megacycles as the thermostat of the crystal oven went on and off. The control action is smooth and rapid. There are no critical adjustments. The range of control can be as high as ± 1000 cycles at 20 kilocycles—that is, ± 5 per cent. At 100 megacycles this amounts to ± five megacycles. In case of failure of the frequency control during a program, operation can be continued with manual frequency control by locking the motor shaft and adjusting the frequency with a vernier tuning control located on the master oscillator tank coil.

The distortion in the frequency-modulated output of the exciter is of the order of 0.5 per cent for modulating frequencies from 30 cycles to 15,000 cycles. The noise level in the output is 74 decibels below 100-per-cent modulation.

A schematic diagram of the new exciter unit is shown in Figure 12. Figure 13 is a front view of the new exciter unit with the major components and controls identified. Figure 14 is a rear view of the exciter unit.

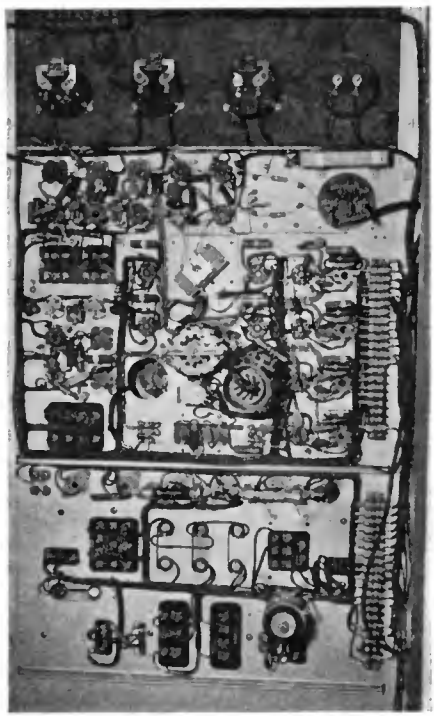


Fig. 14

In summary, the principal advantages of this new exciter unit are:

(1) Its relative simplicity of circuit design.
(2) Use of a minimum number of tubes.
(3) Accurate frequency control.
(4) Reliability of operation.
(5) Excellent performance characteristics.
(6) Ease of maintenance and service.

A PRETUNED TURNSTILE ANTENNA*†

By

George H. Brown and J. Epstein

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*Electrical and mechanical design refinements provide an ultra-high-frequency unit that can be adjusted before being erected. Elements are directly grounded for lightning protection and heaters are provided to prevent icing.*

THE purpose of this paper is to describe a new turnstile antenna design for ultra-high-frequency broadcasting.

The original turnstile[1] was so constructed that the elements were fed by means of open wires twisting around the supporting mast. Adjustment of the phase relationships and current magnitudes was accomplished by means of two properly matched transmission lines cut to the proper lengths and combined in a common terminal. The adjustment of these lines at the base of the antenna proper was rather critical and involved a certain amount of patient effort.

With the advent of frequency modulation on the ultrahigh-frequencies, the turnstile antenna found many applications. It soon became apparent that many antennas would be placed on the tops of extremely tall supporting structures, where the matching and phasing adjustments become very difficult if not impossible. With these factors in mind, the development of a new type of turnstile was undertaken.[2]

The most important feature of the new antenna is the fact that it is completely pretuned during the fabrication of the individual elements so that no work of an engineering nature is necessary to put the antenna into operation.

The 90-degree phase relation is accomplished in the construction of the antenna elements themselves, so that no adjustments at the base are needed to obtain the circular pattern.

The antenna elements are so constructed that, while acting as insu-

---

* Decimal Classification: R321.32.

† Reprinted from *Electronics*, June, 1945.

[1] Brown, George H., "A Turnstile Antenna for Use at Ultra-High-Frequencies," *Electronics*, Apr. 1936.

[2] Brown, George H., and Epstein, J., "A Turnstile Antenna for Ultra-High-Frequency Broadcasting," presented at the I.R.E. Summer Convention, Detroit, Mich., June 23, 1941.

lated members for the radio-frequency signal, they are actually grounded to the pole to afford lightning protection. Because of the grounding feature, it is a simple matter to include sleet-melting units in each radiator.

Two concentric feed lines connect the several units, replacing the open wires formerly used. By means of standard concentric line, these two feed lines are fed in push-pull from a single concentric feed line which runs back to the transmitter. This feed line is terminated to eliminate standing waves.

Principle of Turnstile Antenna

The fundamental objectives of the turnstile antenna are twofold,

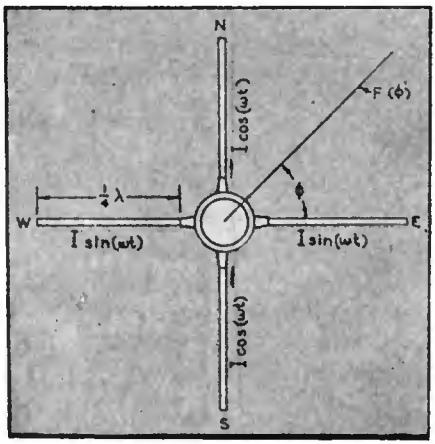


Fig. 1—Elemental turnstile unit.

first, to produce a radiation pattern which has circular symmetry in the horizontal plane, and second, to concentrate the energy in the vertical plane so that the signal strength toward the horizon for a given power input will be considerably greater than that obtained from a single half-wave vertical antenna with the same power input.

The principle involved in producing a circular pattern can best be explained by referring to Figure 1, which shows an elemental unit that could be used to produce the required pattern. As can be seen, it consists of four quarter-wave-length radiators symmetrically oriented in space, carrying equal currents, and so phased that the East-West radiator voltages are out of phase with each other and in time quadrature with the North-South radiator voltages, which are likewise out of phase with one another. The field at any point in the horizontal plane, due to the East-West radiators, is equal to

$$F_1 = (60\ I/r)\ \sin\ (\omega t)\ \sin\ (\phi) \tag{1}$$

where ϕ and r are the coordinates at the point. The field due to the North-South radiators is

$$F_2 = (60\ I/r)\ \cos\ (\omega t)\ \cos\ (\phi) \tag{2}$$

The sum of Equation (1) and Equation (2) gives the total resultant field

$$\begin{aligned} F(\phi) &= F_1 + F_2 \\ &= (60\ I/r)\ [\sin\ (\omega t)\ \sin\ (\phi) + \cos\ (\omega t)\ \cos\ (\phi)] \\ &= (60\ I/r)\ \cos\ (\omega t - \phi) \end{aligned} \tag{3}$$

Thus the total field at any distance r is constant in magnitude and changes in phase as ϕ changes, giving a circularly symmetrical horizontal pattern.

The concentration of energy in the vertical plane is obtained by stacking a number of elemental turnstile units along the vertical axis. The vertical radiation pattern of n units spaced a half-wave length apart and cophased is equal to

$$F(\theta) = \frac{\sin\ [(n\pi/2)\ \sin\ \theta]}{n\ \sin\ [(\pi/2)\ \sin\ \theta]} \tag{4}$$

where $F(\theta)$ = vertical radiation pattern

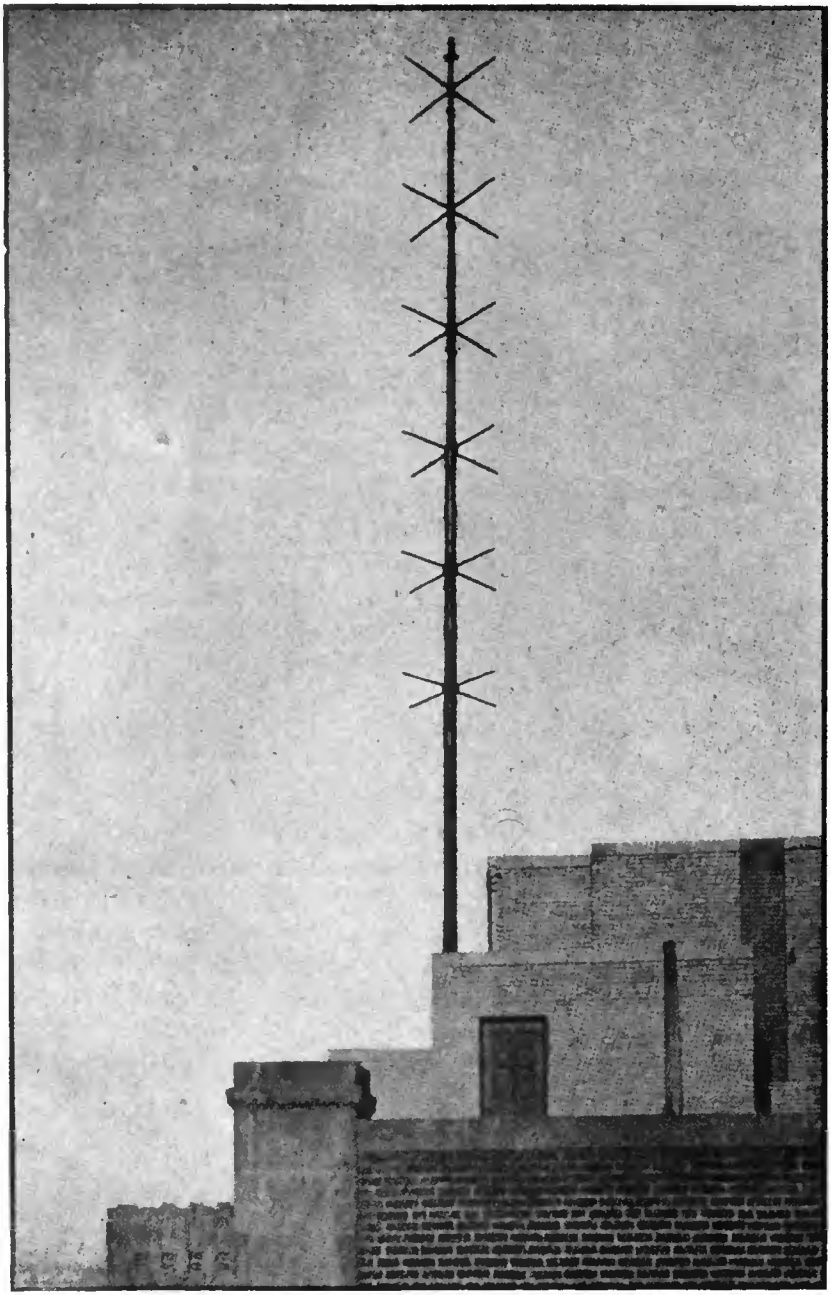
n = number of layers

θ = elevation angle measured with respect to the horizontal plane.

Equation (4) shows that maximum radiation occurs for $\theta = 0$, or in the horizontal plane.

Radiator Elements and Feed Lines

The radiator elements for the new turnstile design were constructed from copper tubing having a two-inch diameter. One of these elements is shown in Figure 2(a) in its proper position with respect to the supporting pole. Since the radiator is made slightly less than one-quarter wave in length, the impedance measured between the end adjacent to the pole and the pole consists of a resistance component and a capacitive component, as shown in the lower section of Figure 2(a).



A modern turnstile antenna, serving WCAU's f-m audience in the Philadelphia area.

The method of supporting the antenna is shown in Figure 2(b). Here a one-inch tube extends from the support pole out through the radiator, and a metal shorting-plug connects the inner tube to the outer sleeve. Thus the impedance of the radiator is shunted by the inductive reactance of the transmission line formed by the inner surface of the outer sleeve and the outer surface of the inner tube.

If the shorting plug is placed in the proper place, the entire combination will be tuned to parallel resonance and the impedance will be a pure resistance, as shown in the lower part of Figure 2(b). The magnitude of this pure resistance will be determined by the length chosen for the outer sleeve. If R_a is the resistance of the outer sleeve and X_a is the capacitive reactance of this same sleeve, the parallel resonant resistance is

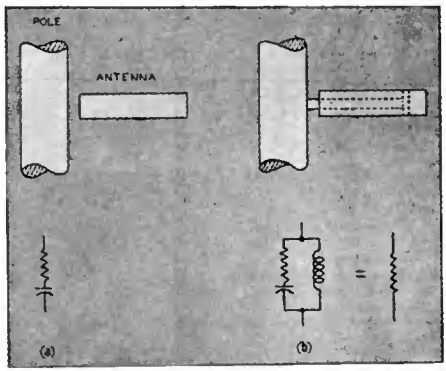


Fig. 2—By properly mounting antenna elements on the supporting pole impedances may be controlled and zero reactances obtained.

$$R = R_a^2 + X_a^2/R_a \tag{5}$$

and we see that R increases as the length is decreased.

The inductive reactance necessary to tune to parallel resonance is

$$Xp = (R_a^2 + X_a^2)/X_a \tag{6}$$

Figure 3 shows the antenna length and shorting-plug position as a function of the desired parallel resistance for a typical radiator in the neighborhood of 45 megacycles.

It will be seen that we have arrived at a radiator structure which may be metallically connected to the support pole and grounded to static or lightning, yet offering a controlled resistance to the radio-frequency

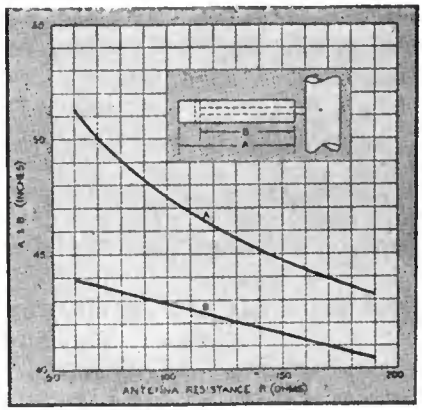


Fig. 3—By changing the antenna length and the shorting-plug position a wide range of parallel antenna resistance is available.

signal. Heating elements and the feed wires to these elements are easily placed within the inner tubing which is fastened to the flag pole.

The means of securing equal currents, with a ninety-degree phase relation, in adjacent radiators is shown in Figure 4. Here a piece of concentric transmission line whose outer conductor has a diameter of seven-eighths of an inch is soldered along the side of the antenna. As seen in Figure 4(a), the inner conductor of this piece of line on the East radiator ends in a short-circuit, while the corresponding element on the North radiator is open-circuited. Thus, the concentric feed line shown alongside the pole feeds the East radiator through a series inductance, while the North radiator is fed through a series capacitance.

The equivalent circuit is shown in Figure 4(b). In Figure 4(c), the parallel resonant circuit of the radiator and its shunt-support stub has been replaced by the pure resistance R, and the series-capacitance and

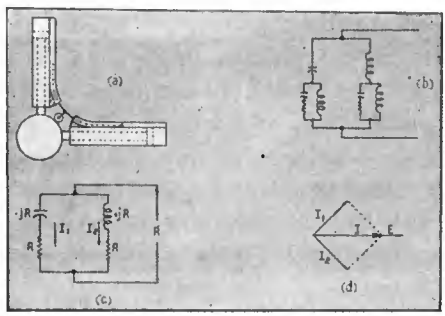


Fig. 4—How phase relations and current ratios in the dipoles are controlled.

series-inductance elements each have a reactance equal in value to R. In a previous analysis of this circuit,[3] it was shown that

(1) I_1 and I_2 were equal in magnitude.

(2) I_1 leads I_2 by ninety degrees.

(3) The input impedance of the total circuit is a pure resistance of R ohms.

The first condition, equality of the two currents, is evident from an inspection of Figure 4(c). A vector diagram of the currents and the driving voltage is shown in Figure 4(d), and helps to show the quadrature relation as well as the fact that the input impedance is a pure resistance.

A complete single layer of a turnstile combination is shown in Figure 5. Here the East and North radiators are fed from one con-

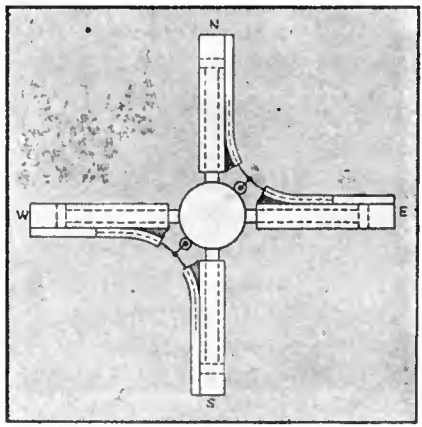


Fig. 5—A complete turnstile layer.

centric line, while the West and South radiators are fed by another concentric line, which is out of phase with the first line by 180 degrees. This arrangement fulfills the conditions required to give a circularly symmetrical horizontal pattern.

Details of one radiator are shown in Figure 6. It will be noted that a cylindrical insulator is added at the mouth of the antenna to furnish additional support and to seal off the interior of the inductive section. A smaller cylindrical insulator is used on the phasing section.

Figure 7 shows a photograph of the feed lines and coupling network of a typical turnstile antenna. It may be seen that the feed lines spiral around the support pole. This is necessary since the radiator

[3] Brown, George H., and Baldwin, John M., "Adjusting Unequal Tower Broadcast Arrays," *Electronics*, Dec. 1943.

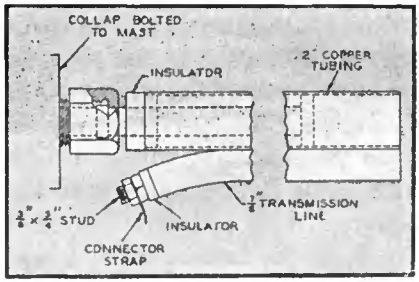


Fig. 6—Details of one radiator element.

layers are one-half wave apart and it is desired to feed these layers in corresponding phase.

The concentric feed lines are shown laid out in a plane in Figure 8. It will be noted that the feed line on the left is one-half wave longer than the feed line on the right. This insures the necessary push-pull

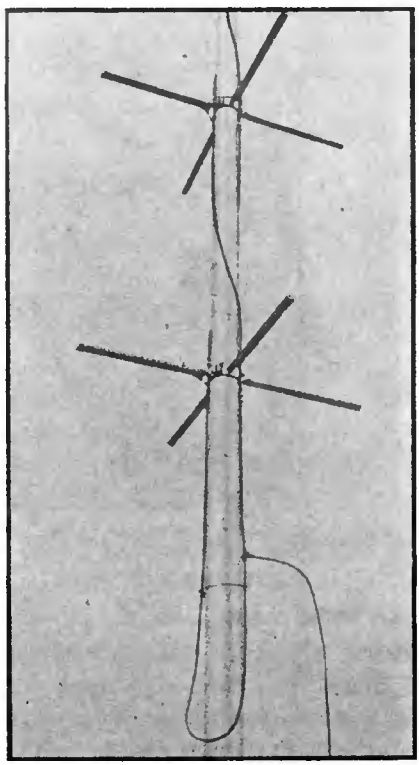


Fig. 7—Individual layers are connected together by means of two concentric lines which spiral around the pole. A simple matching network is also shown.

feed to each layer. Since each radiator has a resistance R, a resistance of R ohms is present at each little end-seal. The half-wave connecting lines transfer impedance without conversion, so that the impedance at point a on the right-hand line is that of the four sets in parallel, that is, $R/4$. In the general case, with n layers of radiators, the impedance at point a would be R/n ohms. The input impedance of a line which is one-quarter wave in length and of characteristic impedance Z_o is Z_o^2/Z out, so the impedance at point b, looking up the

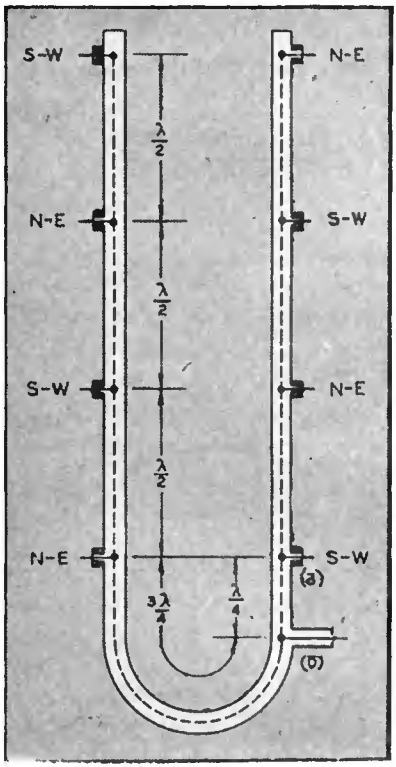


Fig. 8—A simplified layout of feed lines for four layers or less.

right-hand line, is $Z_o^2/(R/n)$, or nZ_o^2/R.

An equal impedance is seen at point b looking into the left-hand transmission line. Since these two impedances are equal, the total impedance at point b is $nZ_o^2/2R$. Now, if we choose the individual radiator resistance such that

$$R = nZ_o/2 \tag{7}$$

the input resistance is equal to Z_0 and the main feed line is matched in its characteristic impedance. The following table shows the values of R which would be chosen for a number of layers of Turnstile:

TABLE I

Number of Layers (n)	R/Z_0	R (ohms) when Z_0 is 70 ohms
1	0.5	35
2	1.0	70
3	1.5	105
4	2.0	140

While we could theoretically extend this procedure to many more layers, it was found that the antenna dimensions became quite critical when more than four layers were used. In order to obtain high parallel resistance, the radiators had been shortened to a point where the resistance changed a great deal with a slight change in antenna length.

A practical feed system used for six or more layers is shown in Figure 9. Here the main feed line enters at a point midway in the antenna structure. To match the main feed line, the radiator parallel resistance must satisfy the condition

$$R = nZ_0/8 \tag{8}$$

The table below shows the appropriate value of R:

TABLE II

Number of Layers (n)	R/Z_0	R (ohms) when Z_0 is 70 ohms
6	0.75	52.5
8	1.0	70.0
10	1.25	87.5
12	1.5	105.0

In order to provide a design which would be installed in the field without adjustment, tests were made on typical arrangements to assure ourselves that close tolerances on dimensions were not necessary. Our objective was to adjust and test an antenna of, say, four layers at some frequency in the f-m broadcast band, learn what scale factor should be applied to each dimension in order to build an antenna for any frequency in the band, and provide charts so that any future antennas which were to be provided could be constructed from these charts.

Method of Adjustment

The main flag pole was laid out horizontally on two wooden supporting structures so that the pole center was approximately twelve feet above the ground, as shown in Figure 10. The antenna system was assembled on the pole. Then the North and South radiators were disconnected from the feed lines and the series inductance sections of the East and West radiators were shorted out directly across the

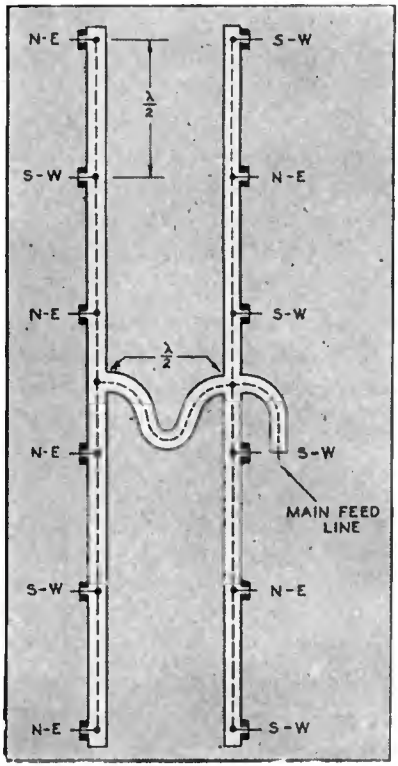


Fig. 9—A practical feed system for six layers or more.

small cylindrical insulators at the mouths of the inductance sections. Next, the antenna length and the position of the shorting plug on the support rod (see Figure 3) were changed on each East and West radiator until the main feed line was properly terminated in its characteristic impedance. This condition was observed by means of a probe voltmeter sliding along a slotted measuring line. The antenna length adjustment was made by sliding very thin-walled sleeves over the end of the radiator, while shorting plugs with spring-contact fingers were used to adjust the shunt tuning inductance.

This adjustment insured that the correct value of R was present for each radiator. It should be noted that even though mutual impedance was present between radiators, this method of adjustment takes account of the mutual impedances without even measuring the specific values.

We were now ready to adjust the series-inductance and series-capacitance sections. The shorting straps were removed from the series-inductance sections of the East and West radiators. The North and South radiators were connected to the feed lines. The North and South radiator lengths and position of the shorting plugs were made to correspond with the dimensions found to be desirable for the East and West radiators. Then, the approximate length for the series-capacitance sections was calculated and the capacitance sections were cut a little longer than this length. The series inductance sections were

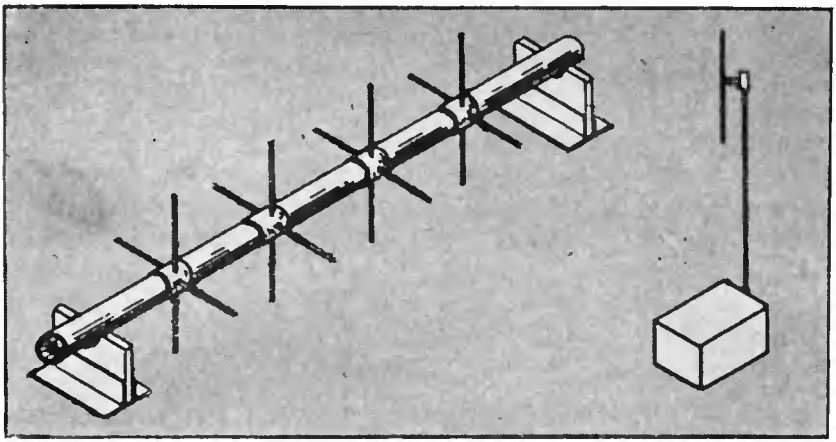


Fig. 10—Arrangement used for adjusting the new turnstile antenna on the ground.

next adjusted by means of variable shorting-plugs until the main feed line was matched.

Under this condition, the currents in the North and South radiators will be in quadrature with the currents in the East and West radiators, but may be somewhat greater in magnitude than the latter. A field-intensity meter was placed a few hundred feet away from the antenna. The turnstile antenna was then rotated about the axis of the flagpole and observations of field intensity were made at the remote point. If the maximum field intensity was obtained when the North and South radiators were pointing straight up and down, it was necessary to shorten the capacitance sections to obtain a circular pattern. When the capacitance sections were shortened, the series-inductance

sections were readjusted to again match the main feed line. Then observations were again made of the field intensity as the pole was rotated. Three or four adjustments of this type are sufficient to obtain a circular pattern.

Measurements of a typical well-adjusted turnstile are shown in Figure 11. This may be regarded as the horizontal field pattern of the antenna.

Vertical Field Intensity Patterns

With the pole mounted horizontally above the ground, we were afforded an excellent opportunity to measure the vertical radiation pattern by simply moving the field-intensity meter on the circumference

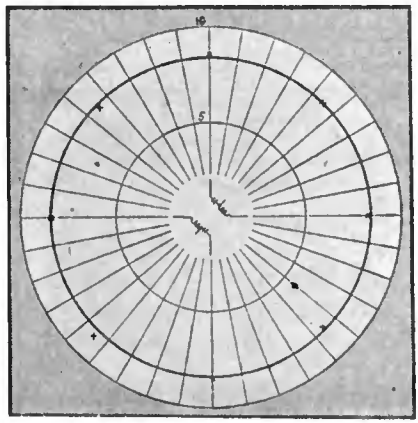


Fig. 11—Experimental data showing that the horizontal pattern of a turnstile is circular.

of a circle, where the center of the circle corresponds with the midpoint of the antenna.

Figure 12 shows the vertical pattern of a four-layer turnstile. The solid line is the calculated characteristic. The radiators are not quite one-half wave apart since the velocity of propagation on the feed lines is about ninety-two percent of the velocity in free space, and the feed lines were shortened to take this factor into account.

At an angle of thirty-five degrees, two experimental points are shown as solid triangles. It was not possible to tell from the field-intensity meter readings whether this point should be shown positive or negative.

A vertical field intensity pattern for a six layer turnstile is shown in Figure 13.

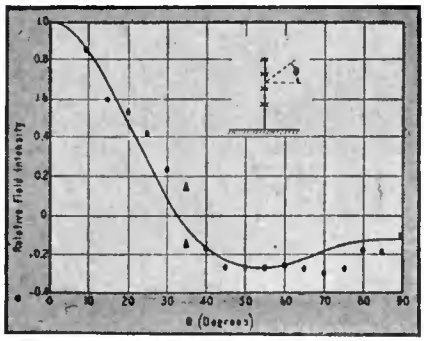


Fig. 12—Vertical field intensity pattern of a four-layer turnstile.

Impedance Characteristics

After the antennas were adjusted, observations were made of the standing-wave ratio on the main feed line, that is, the ratio of minimum to maximum voltage. This standing-wave ratio on a four-layer antenna which had been adjusted at 45.3 megacycles is shown in Figure 14. Observe that the curve is rather broad.

A similar curve for a six-layer turnstile which had been adjusted at 46.9 megacycles is shown in Figure 15. It is seen that the curve is somewhat sharper for the six-layer turnstile than for the four-layer, but curves are more than adequate for transmission of an f-m signal.

Voltage and Currents on the Feed System

Since the feed lines twist about the pole, we are limited as to the size of the feed lines. It is quite possible to bend concentric lines whose outer conductor has a diameter of seven-eighths of an inch, but when one considers the possibility of using a tubing diameter of

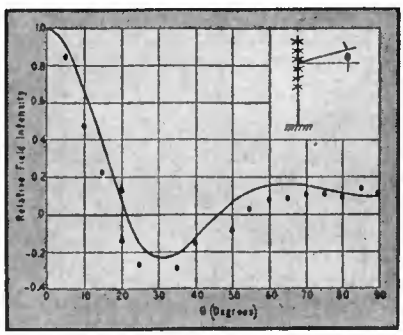


Fig. 13—Vertical field intensity pattern of a six-layer turnstile.

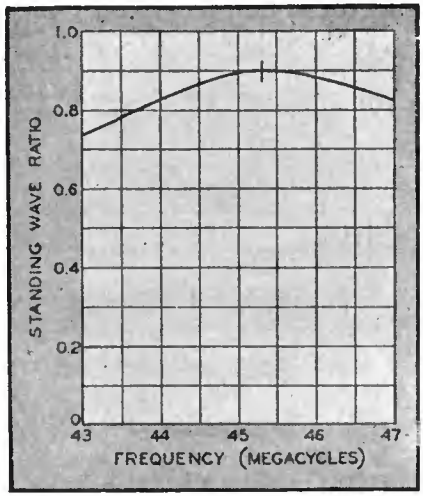


Fig. 14—Standing-wave ratio on the main feed line of a four-layer turnstile.

one and one-half inches, the prospects appear discouraging. It thus at first appears that this antenna design is limited in power-handling capabilities. An examination of the currents and voltages appearing on the feed lines and on the radiators will, at this point, prove interesting and illuminating.

Let us take, as an example, the case of a six-layer turnstile operating with a power of 10,000 watts. In the equivalent circuit (Figure 4c) the resistance R is 52.5 ohms. Since the power into each radiator is

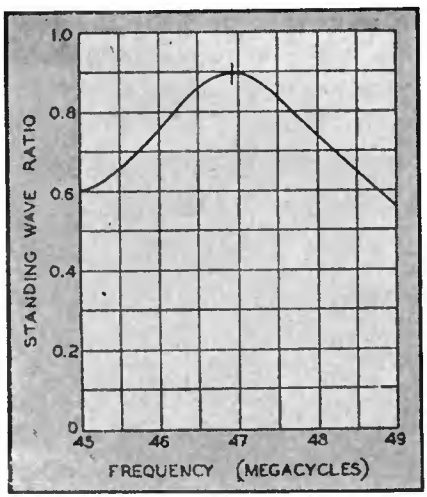


Fig. 15—Standing-wave ratio on the main feed line of a six-layer turnstile.

one twenty-fourth of the total power, the voltage from each radiator to the flagpole is 149.0 volts, with a similar voltage across each series—capacitance and series—inductance. The feed straps carry approximately 2.8 amperes, with a voltage from feed strap to ground of 209.0 volts. The voltage and currents along the feed lines are shown in Figure 16. These values speak for themselves.

It is the authors' wish to acknowledge the helpful assistance of their colleagues Donald W. Peterson and O. M. Woodward, Jr.

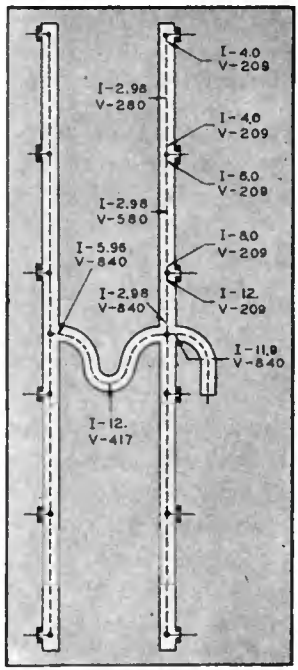


Fig. 16—Voltage and currents on the feed lines of a six-layer turnstile, operating with a power of 10,000 watts.

CHARACTERISTICS OF THE PYLON FM ANTENNA*†

By

ROBERT F. HOLTZ

Engineering Products Department, RCA Victor Division,
Camden, N. J.

Summary—The pylon antenna is a stack of slotted cylindrical sections. A description of the physical aspects and performance characteristics is given.

GENERAL DESCRIPTION

STRUCTURALLY, mechanically, and electrically this new antenna has been reduced to strictly functional elements. Each pylon section, Figure 1, is a metal cylinder approximately 13 ft. high and 19 ins. in diameter. It carries no dipoles, no loops, no appendages of any kind. The radiator is the cylindrical structure itself. A single transmission line, running up the inside of the cylinder, along the slot to the mid-point, is the feed line.

The cylinder is rolled from a single aluminum sheet, bolted at the top and bottom to cast rings which give it great mechanical strength and provide means for securing the pylon section to the supporting tower or to additional stacked sections.

As many as four, and perhaps more of these basic cylindrical sections can be stacked to provide a high-gain antenna of remarkably simple design. Such an assembly is shown in Figure 2. It should be noted that the bottom section is mounted directly on the building roof, without the use of a supporting tower.

ORIGIN OF THE DESIGN

The cylindrical pylon antenna has been undergoing development since early in 1944 at RCA in Camden. The principle of the so-called slot antenna has been employed in a number of applications where existing plane metal surfaces had to be adapted to radiating high-frequency energy.

When it was recognized that these radiating surfaces could be rolled into a cylinder and used for the radiation of omni-directional, horizontally-polarized waves, steps were taken to explore the further possibilities of such a design. As Figure 3 shows, the pylon antenna

* Decimal Classification: R326.611 × R630.
† Reprinted from *FM and Television*, Sept. 1946.

is, in effect, made up of a large number of circular elements, each one of which is a radiating member. When power is fed to the slot, Figure 4, the slot functions as a transmission line, and currents flow around the circular paths shown by the dotted line.

Early tests were conducted on 200-mc. models, fabricated of light wire screen. Although it may appear that the choice of a radiator

Fig. 1—Single pylon antenna section.

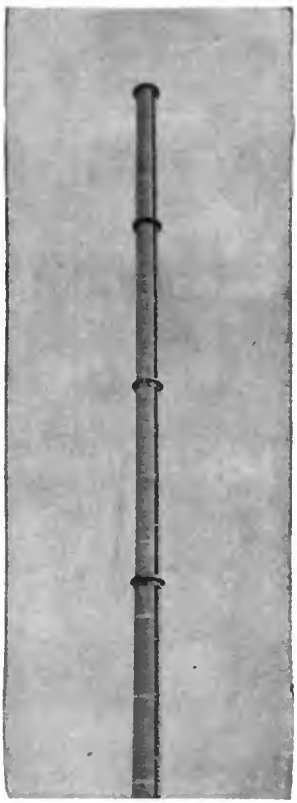


Fig. 2—A 4-section pylon antenna, mounted on the roof of a building.

½ wavelength in circumference and 1 wavelength long was an obvious one for such an antenna, this is by no means true. A given set of dimensions for a cylinder yields a radiator which can be operated at any one of a number of nodes. The cylinder diameter is intimately associated with the horizontal pattern. On the other hand, the diameter is also a controlling factor in the length of cylinder for resonance.

A compromise was struck at this stage of the development. Investigations revealed that the slot width could be used to control the absolute value of the impedance level at the feed point. This characteristic was explored carefully, in order to arrive at an optimum design. Careful tests were carried out to make certain that a round cylinder would yield the optimum characteristics. It was found that no reasonable degree of ovality improved the performance in any way. Multiple-slot cylinders, with more than one slot on the periphery, were con-

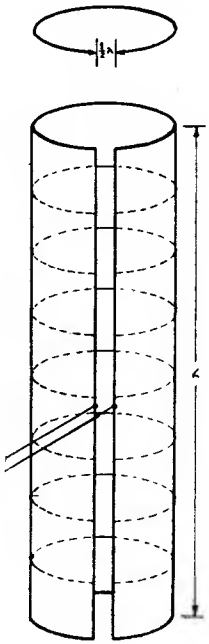


Fig. 3—Each section serves as a number of circular elements.

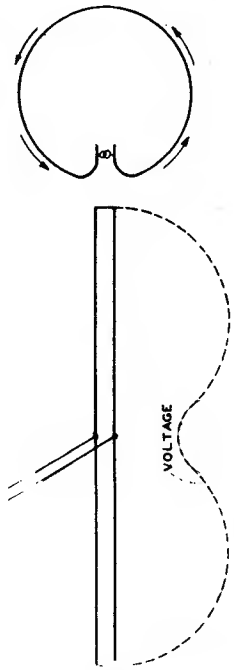


Fig. 4—The slot functions as a transmission line when fed with power.

sidered and rejected because the structural and electrical simplicity were lost.

Characteristics

Pylon sections can be stacked to provide higher gain. Figure 2 shows a 4-section assembly which has a power gain of 6. As many as 8 sections can be stacked in the same manner to give a power gain of 12.

Figure 5 illustrates the very simple feeder system required for 1, 2, and 4 sections. The lines are actually run inside the pylon sections,

so that they are protected from exposure. The feed system has an advantage when the antenna is being erected. That is, each pair of sections can be joined on the ground, and the interconnecting line put

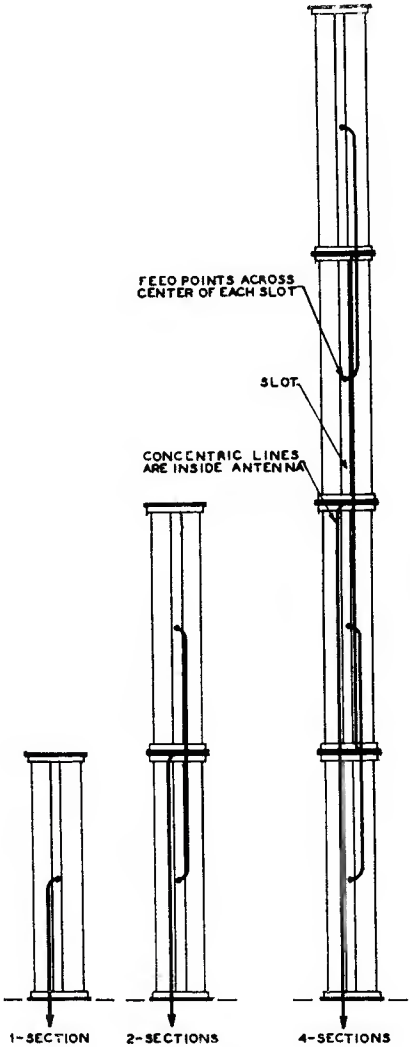


Fig. 5—Feeder system used for 1, 2, and 4 sections is exceedingly simple.

in place. Then it is only necessary to make 2 or 3 connections "in the air."

Successive sections are joined by bolting the end flanges together. All construction work involving arms or loops is eliminated, and no

supporting mast is required since the pylon sections are at once mast and radiators.

The power gain is indicated in Figure 6 and the horizontal radiation in Figure 8. While the pattern is not exactly circular, this is generally an advantage since, in any area, there is always one direction

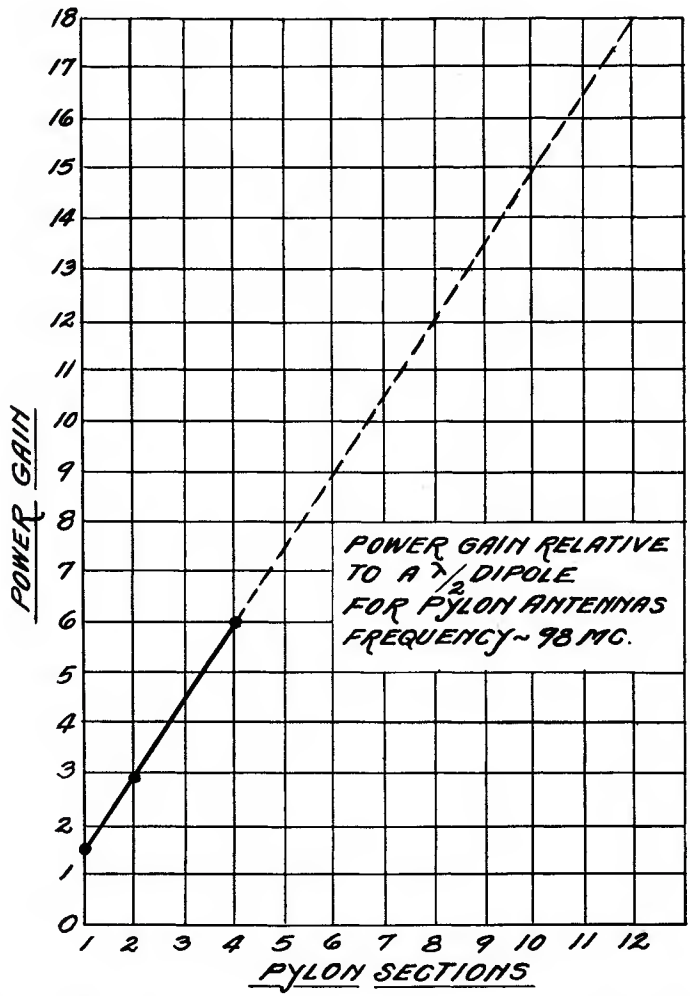


Fig. 6—Power gain of pylon, relative to standard half-wave dipole.

in which it is desirable to direct extra power to overcome topographical conditions, or to obtain a little extra distance. The vertical field patterns are plotted in Figure 7.

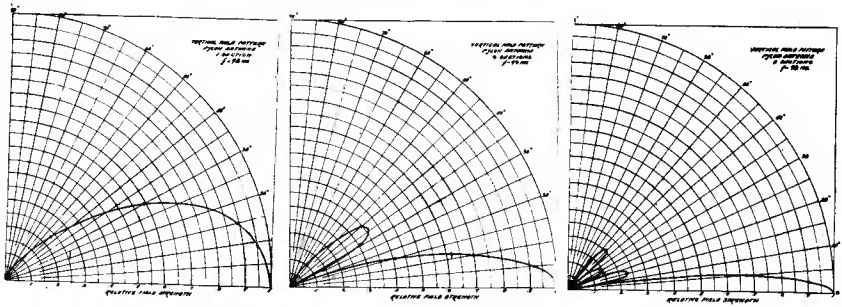


Fig. 7—Horizontal radiation pattern for antennas of 1, 2, and 4 sections.

One size of radiator, or pylon section, and two sizes of transmission line lengths cover the whole FM broadcast band. No tuning or adjusting is required either on the ground or in the air. The wide bandwidth is obtained by the use of line elements, Figure 9.

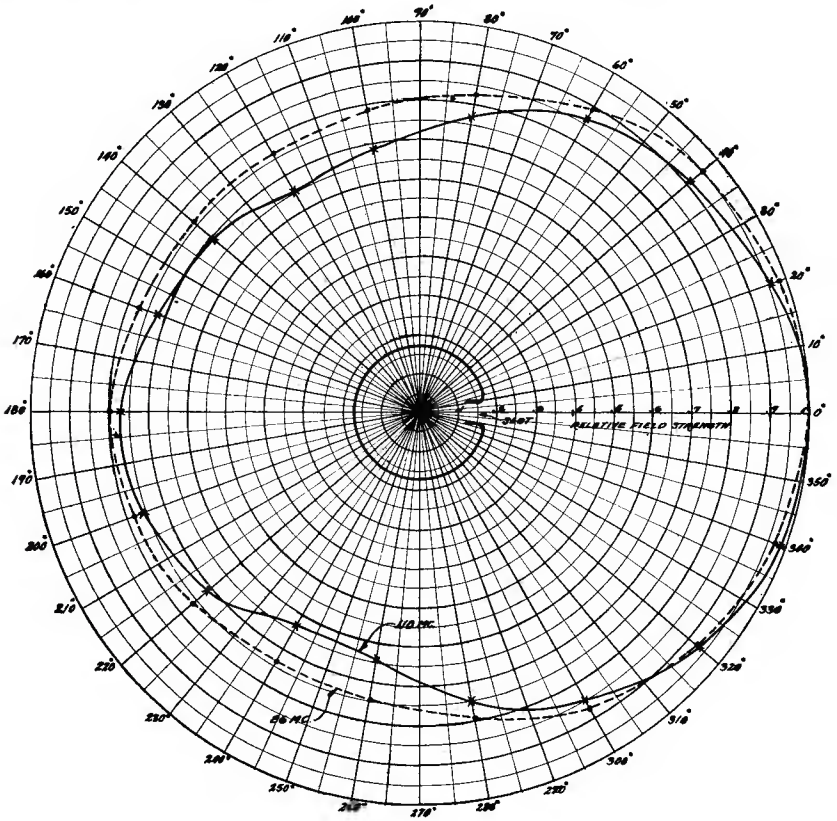


Fig. 8—Plot of the horizontal radiation, at frequencies of 86 and 110 mc.

A single section, Figure 1, weighs only 350 lbs., or 700 lbs., for two section, while a 4-section assembly, made of heavier material, comes to 2,000 lbs. These weights include the radiators, transmission line, beacon, steps, and hardware.

The ice problem is negligible because the transmission lines are inside the cylinders, where ice formation is unlikely. Any formation on the outside will add little to the weight and loading. Maintenance is made easy by the simple feed-line arrangement, the small number of end seals, and by the fact that the lines are enclosed within the antenna sections.

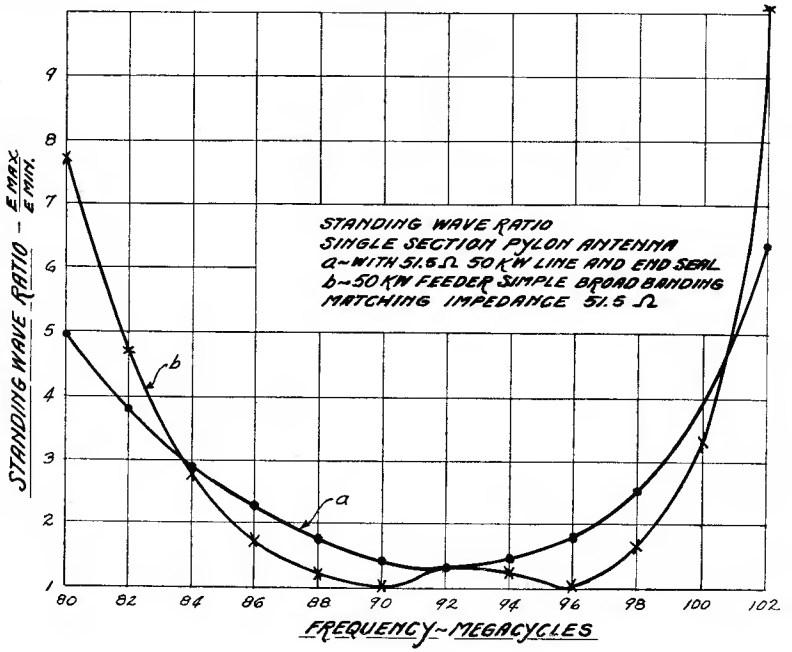


Fig. 9—The wide band width is obtained by the use of line elements.

A standard 300-mm. code beacon can be mounted on the plate which covers the top section of the antenna. When it is necessary to replace the beacon lights or the lines, they can be reached by climbing the steps on the cylinders. The slot is wide enough to give access to the feeder lines.

When an antenna of this type is carried at the top of a supporting tower, the load is made light by the smooth and unbroken exterior of the pylon sections. This can be seen from the relatively low value of R_1, Figure 10 shown in the table of specifications. Furthermore no unbalanced load is present in this design. Accordingly, there is a

saving in the cost of tower construction, if required, particularly in areas where icing and its attendant top-loading must be taken into consideration.

The wind load at the center of the radiator sections is based on 20 lbs. per square foot of the projected area and assumes a circular cross-section. The value of R_1 includes the 300-mm. beacon.

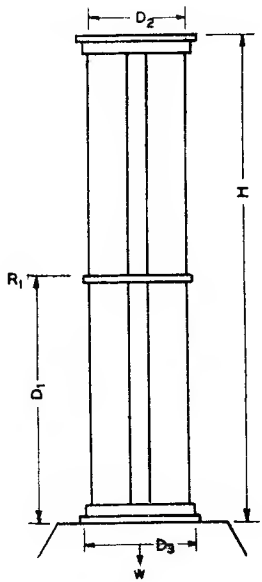


Fig. 10—Values for these design factors are given in the table here.

SPECIFICATIONS

The following table gives the specifications for pylon radiators of 1, 2, and 4 sections:

	1 Sec.	2 Sec.	4 Sec.
Power Gain, nominal	1.5	3	6
Field Gain, nominal	1.23	1.73	2.45
Weight, lbs.[1]	350	700	2000
Height, ft.	13.5	27	54
R_1, lbs.[2]	501	950	1868
D_1, ft.	7½	14	27⅛
D_2, ins.	19½	19½	19½
D_3, ins. bolt circle[3]	22⅝	22⅝	22⅝

[1] Total weight including radiator, transmission lines, beacon, steps, and hardware.

[2] Total wind load on radiator, including beacon. Based on 20 lbs. per sq. ft. of projected area, all sections assumed round.

[3] Mounting is by means of eighteen ¾-in. bolts on 22⅝-in. circle.

REACTANCE-TUBE FREQUENCY MODULATORS*†

By

Murray G. Crosby

RCA Communications, Inc.,
New York, N. Y.

Summary

An improved reactance-tube modulator is described which utilizes a push-pull circuit to balance out the carrier frequency instability due to power supply variations.

Design considerations are given for the type of reactance-tube modulator which uses automatic frequency control to stabilize the mean frequency.
(8 pages; 3 figures)

* Decimal Classification: R148.2.
† *RCA REVIEW*, July, 1940.

A TRANSMITTER FOR FREQUENCY MODULATED BROADCAST SERVICE USING A NEW ULTRA-HIGH-FREQUENCY TETRODE*†

By

A. K. Wing and J. E. Young‡

RCA Manufacturing Co., Inc., Harrison and Camden, N. J.

Summary

This article describes a 1-kilowatt transmitter developed for the frequency-modulated broadcast service. A description of the circuits, tubes and mechanical arrangement of the apparatus is included. The electrical performance is described.
(10 pages; 6 figures)

* Decimal Classification: R355 × R630.
† *RCA REVIEW*, January, 1941.
‡ Now with Engineering Products Department, RCA Victor Division, Camden, New Jersey.

DRIFT ANALYSIS OF THE CROSBY FREQUENCY-MODULATED TRANSMITTER CIRCUIT*†

By

E. S. Winlund

RCA Manufacturing Co., Inc., Camden, N. J.

Summary

Component drift, sensitivity, and band-width expressions are combined in an over-all expression for frequency stability of the Crosby circuit. Using experimentally obtained constants in this expression, an equation is derived for drift in terms of frequency and frequency multiplication open to choice by the designer, and the results are shown as design curves. The equation is checked against actual conditions existing in a Crosby exciter unit.

(9 pages; 7 figures; 2 tables)

* Decimal Classification: R148.2 × R630.

† *Proc. I.R.E.*, July, 1941.

ANTENNAS FOR F-M STATIONS*†

By

John P. Taylor

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

The importance of proper antenna choice is emphasized, and the significance of antenna gain is discussed. The paper covers: advantages of multi-layer antennas, limitations of multi-layer antennas, interrelation of height, gain and power, how antenna gain is obtained, meaning of field gain and power gain, practical types of FM antennas, original Brown turnstile, improved Brown turnstile, modifications of the turnstile, circular or "ring" antennas, and square-loop antennas. The summary lists a number of check-off items to guide station engineers in planning an FM station.

(8 pages; 15 figures)

* Decimal Classification: R320 × R630.

† *Broadcast News*, August, 1944.

F-M AUDIO MEASUREMENTS WITH AN A-M RECEIVER*†

By

R. J. Newman

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

The procedure for determining modulation level and making audio response measurements by the "Bessel Zero" method is covered in detail. "Bessel Zero" points are explained and the problems of finding them and how to use this information is discussed. The paper then deals with the procedure for making measurements of the AF response.

(3 pages; 2 figures; 1 table)

* Decimal Classification: R254.12.

† *Broadcast News*, August, 1944.

A SQUARE LOOP F-M ANTENNA*†

By

John P. Taylor

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

A typical installation of an FM antenna built around the top of an AM tower rather than on a flagpole extension is discussed. This construction adds little weight or wind resistance. Theory, details of construction, and tuning procedures are described.

(5 pages; 11 figures)

* Decimal Classification: R320 × R630.

† *Electronics*, March, 1945.

AUDIO FREQUENCY RESPONSE AND DISTORTION MEASURING TECHNIQUES FOR F-M TRANSMITTING SYSTEMS*†

By

R. J. Newman

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

Necessary procedures for making the response and distortion measurements required for "proof of performance" of F'M installations is dealt with in this paper. The equipment required is indicated, and the subject of pre-emphasis is discussed. Topics covered include: arrangement of equipment, calibration of the audio oscillator, frequency response measurements, use of distortion and noise meter for frequency response measurements, harmonic distortion measurements, and FM and AM noise level measurements.

(6 pages; 5 figures; 1 table)

* Decimal Classification: R630.1.

† *Broadcast News*, June, 1945.

F-M ANTENNA COUPLER*†

By

John P. Taylor

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

An FM antenna at the top of a base-insulated tower used as an AM radiator is fed through a concentric line without short-circuiting the tower. The coupling unit which accomplishes this is described in detail. The coupling problem is dealt with both in theory and with regard to construction. The tuning procedure is discussed in a thorough manner.

(3 pages; 8 figures)

* Decimal Classification: R320.51.

† *Electronics*, August, 1945.

ISOLATION METHODS FOR F-M ANTENNAS MOUNTED ON A-M TOWERS*†

By

Robert F. Holtz

Engineering Products Department, RCA Division,
Camden, N. J.

Summary

The problem of mounting FM radiators atop AM towers is dealt with for both shunt-and series-fed AM towers. The general desirability of FM antennas atop AM towers is discussed. No particular problem other than structural requirements exists in the case of shunt-fed AM towers. However, in the case of series-fed AM towers, the more-usual arrangement, an isolator unit is required. Such a unit is described and its function and operation discussed in detail.

(2 pages; 4 figures)

* Decimal Classification: R320.

† *Broadcast News*, October, 1946.

SLOT ANTENNAS*†

By

N. E. Lindenblad

Research Department, RCA Laboratories Division,
Rocky Point, N. Y.

Summary

The development of flush-type radiators of the slot and pocket type is described. Special emphasis is made of types applicable to aircraft. Specific solutions to altimeter and marker beacon pickup antennas are described. Reference to application in other fields is also made.

The general aspects of the phenomena which are involved are examined and it becomes evident that workable solutions, in the majority of cases, can be obtained only by means of actual experiment, since variations in the surroundings have first-order influence upon such vital characteristics as radiation patterns, slot impedance, and bandwidth.

Progress before and during the war is described in somewhat chronological manner. It is pointed out that while this progress has been considerable, an appreciable amount of skillful investigation remains to be done before slot antennas can be brought to maximum usefulness.

(7 pages; 19 figures)

* Decimal Classification: R326.21 × R630.

† *Proc. I.R.E.*, December, 1947.

THE SERVICE RANGE OF FREQUENCY MODULATION*†

By

MURRAY G. CROSBY

Engineering Department, RCA Communications, Inc.

Summary—An empirical ultra-high-frequency propagation formula is correlated with experimentally confirmed frequency modulation improvement factors to develop a formula for the determination of the relation between signal-noise ratio and distance in a frequency-modulation system. Formulas for the distance between the transmitter and the point of occurrence of the threshold of frequency-modulation improvement, as well as formulas for the signal-noise ratio occurring at this distance, are also developed.

Results of calculations and listening tests are described which evaluate the signal-noise ratio gain obtained by applying pre-emphasis to the higher modulation frequencies of both amplitude and frequency modulation systems.

Examples using a typical set of conditions are given and discussed.

IT HAS BEEN shown that the propagation characteristics of frequency modulation are such as to confine the use of this type of modulation insofar as telephony is concerned to the ultra-high frequencies which do not use the ionosphere as a transmission medium.[1] Consequently, for most purposes, calculations of service range of a frequency modulation transmitter may be based totally on the propagation characteristics of ultra-high-frequency waves and in this paper that base will be used. Formulas and empirical data now available make it possible to calculate the field strength at the receiver if transmitter power, antenna heights, and distance are known. Consequently, the signal-noise ratio for amplitude modulation may be determined if the field strength of the noise is known. From this known amplitude-modulation signal-noise ratio, the signal-noise ratio obtainable with a given frequency-modulation system may be determined by multiplying the corresponding amplitude-modulation signal-noise ratios by the frequency-modulation improvement factors as given in the author's previously published paper[2] on frequency-modulation noise characteristics. It is the purpose of this paper to perform this correlation of these propagation and frequency-modulation improvement formulas and develop formulas so that from the known constants of the frequency-modulation system, the signal-noise ratio at a given distance may be directly calculated.

* Decimal Classification: R630.11.
† Reprinted from *RCA REVIEW*, January, 1940.

Amplitude-Modulation Signal-Noise Ratio vs. Distance

The following empirical formula has been given by H. H. Beverage[3] for calculating the field strength when the receiver is within the optical distance of the transmitter:

$$E \text{ (r-m-s volts per meter)} = \frac{88\sqrt{W}\,ah}{\lambda D^2} \qquad (1)$$

where W = effective watts radiated = power in antenna times antenna power gain over a one-half wave dipole,
a = the receiving antenna height in meters,
h = the transmitter antenna height in meters,
D = the distance in meters,
λ = the wavelength in meters.

This formula is to be used for calculating the field strength for distances within the horizon only. For distances beyond the horizon, Beverage used a graphical method of plotting the curve of field strength versus distance. In this graphical method the field strength versus distance curve according to equation (1) was plotted for distances out to the horizon and then the curve was continued for distances beyond the horizon, but with a slope of $1/D^n$ instead of $1/D^2$. The exponent "n" was determined empirically and varies with frequency in the manner shown in Figure 1 which is reproduced from Beverage's paper.

In place of the graphical construction of the curve for distances beyond the horizon, the formula given by (1) may be revised to be applicable to all distances, whether they be within or outside of the horizon, as follows:

$$E \text{ (r-m-s volts per meter)} = \frac{88\sqrt{W}\,ah\,D_h^{n-2}}{\lambda D^n} \qquad (2)$$

in which the exponent "n" is equal to two for distances within the horizon and is chosen from the curve of Figure 1 for distances beyond the horizon. D_h is the distance to the horizon in meters and is equal to $2.21\sqrt{h} + 2.21\sqrt{a}$ where a and h are the receiving and transmitting heights in meters. Thus, where $D < D_h$, $n = 2$, and where $D > D_h$, n is taken from Figure 1.

When the units of the formula given by (2) are converted to feet, microvolts, miles, and megacycles, the formula becomes:

$$E \text{ (r-m-s microvolts per meter)} = \frac{0.01052\sqrt{W}\,ahf\,D_h^{n-2}}{D^n} \qquad (3)$$

where W = effective watts radiated = power in antenna times antenna power gain over a one-half wave dipole,

a = receiving antenna height in feet,

h = transmitting antenna height in feet,

D = distance in miles,

D_h = distance to the horizon in miles = $1.22\sqrt{h} + 1.22\sqrt{a}$,

f = frequency in megacycles.

In comparing the calculated curves with the experimental data, Beverage states that "scattering and absorption, even in open country,

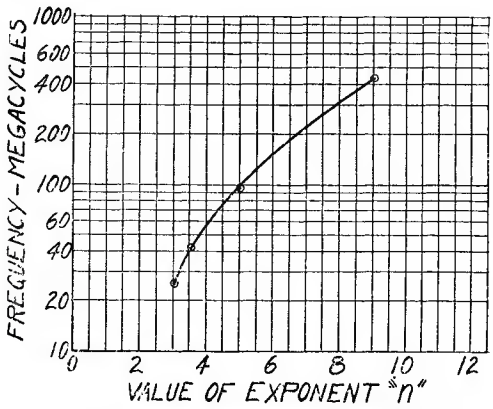


Fig. 1—Variation of exponent "n" in propagation formula when transmission is beyond the horizon.

tend to reduce the average intensity to something in the order of thirty to sixty per cent of the calculated value." In the following derivations an average experimental factor of forty-five per cent will be included to take into account this absorption and scattering.

It will be noted that this formula gives the average field intensity and does not take into account fading. More recent work by MacLean and Wickizer[1] shows the range of fading which may be expected for one set of transmission conditions. Thus, for transmission conditions reasonably close to the case treated by MacLean and Wickizer, the signal intensities at the fading minimums may be determined by applying a correction obtained from Figure 12 of the MacLean and Wickizer paper which is reproduced herewith as Figure 2. In the present paper, the formulas will be derived for the case of the average signal intensity and the fading correction will be applied to the examples given.

The peak carrier-noise ratio* obtained at a given distance may be determined by converting the voltage given by (3) into peak values (multiply by 1.414) and dividing by the peak noise voltage, N. N is the noise field strength as determined by means of a field-strength meter having a pass band characteristic equal to twice the band width of the audio spectrum which it is desired to receive. The experimental scattering and absorption factor may be taken into account by multi-

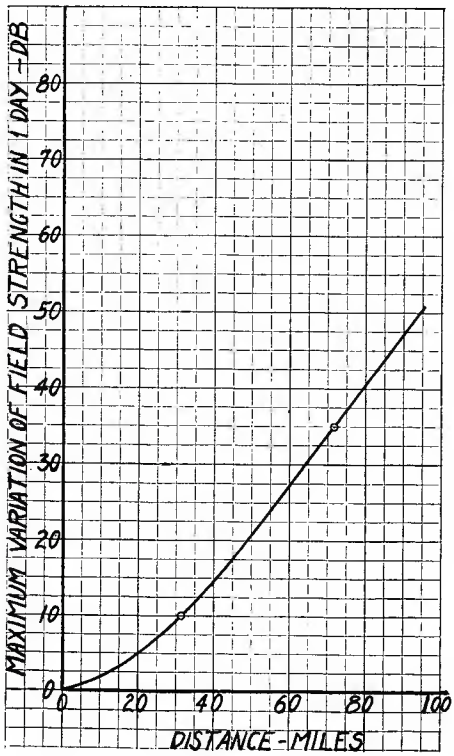


Fig. 2—Fading range of 50-megacycle transmission, transmitter antenna height = 1300 feet.

plying by 0.45. Thus, the amplitude-modulation peak carrier-noise ratio, which is approximately equal to the amplitude-modulation peak signal-noise ratio at the receiver output, for the case of one hundred per cent modulation, is given by:

* Throughout this paper carrier-noise ratio will refer to the ratio between the carrier and noise voltages as measured at the output of the intermediate-frequency channel of the receiver. Signal-noise ratio will refer to the ratio between the signal and noise voltages at the output terminals of the receiver. In an amplitude-modulation receiver the signal-noise ratio is usually substantially equal to the carrier-noise ratio, but in a frequency-modulation receiver the two quantities may differ greatly.

$$C/N = \frac{0.0067\sqrt{W}\,ahf\,D_h^{n-2}}{N\,D^n} = S_a/N_a \qquad (4)$$

where C = peak carrier-field strength,
N = peak noise-field strength as above defined,
S_a = peak audio-signal voltage, amplitude-modulation receiver,
N_a = peak audio-noise voltage, amplitude-modulation receiver.

Frequency-Modulation Signal-Noise Ratio vs. Distance

The author's previously cited paper[2] on frequency-modulation noise characteristics gives the improvement factors effected by a frequency-modulation system over an amplitude-modulation system. For the case of fluctuation noise the factor is $\sqrt{3}$ times the deviation ratio μ. (The deviation ratio, μ, is equal to F_d/F_a where F_d is the peak-frequency deviation due to modulation and F_a is the width of the audio channel of the system.) This factor holds for equal carriers fed to the two receivers and for the condition of a peak carrier at least 4 decibels above the peak noise in the frequency-modulation receiver intermediate-frequency channel. Thus, the signal-noise ratio at the output of the frequency-modulation receiver may be found by multiplying the signal-noise ratio given by (4) by the improvement factor or:

$$S_f/N_f = \sqrt{3}\,\mu\frac{0.0067\sqrt{W}\,ahf\,D_h^{n-2}}{N\,D^n}$$

$$= \frac{0.0116\,\mu\sqrt{W}\,ahf\,D_h^{n-2}}{N\,D^n} \text{ (fluctuation noise)} \qquad (5)$$

where S_f = peak audio-signal voltage, frequency-modulation receiver,
N_f = peak audio-noise voltage, frequency-modulation receiver.

When the received noise is impulse noise the improvement factor is equal to twice the deviation ratio or 2 μ. The corresponding signal-noise ratio is then:

$$S_f/N_f = \frac{0.0134\,\mu\sqrt{W}\,ahf\,D_h^{n-2}}{N\,D^n} \text{ (impulse noise)} \qquad (6)$$

It will be noted that in deriving the signal-noise ratio formulas of (5) and (6), the power gain normally effected by frequency modulation at the transmitter is automatically taken care of by the fact that W appears in both the amplitude- and frequency-modulation formulas. In the case of amplitude modulation the value of W used would normally be less than the corresponding value used in the frequency-

modulation formulas by a factor equal to the frequency-modulation power gain at the transmitter.

Effect of the Improvement Threshold

The presence of the phenomena called the "improvement threshold" places a rather definite maximum service range on a frequency-modulation system. The improvement threshold occurs at the point where the peak voltages of the noise and carrier in the intermediate-frequency channel of the frequency-modulation receiver are equal. The experimental work of the author's previously cited paper[2] shows that the full frequency-modulation improvement for fluctuation noise is not obtained until the carrier is at least four decibels above the improvement threshold or where the carrier-noise ratio is about four decibels. The nature of the improvement threshold is such that the signal-noise ratio drops rapidly as the carrier falls below the four decibel carrier-noise ratio. For the case of fluctuation noise, which is of a continuous nature, the noise smothers the signal in a manner which has been described in detail before[2]. Consequently the distance at which the improvement threshold occurs for fluctuation noise may be taken as the maximum service range of the frequency-modulation system. However, for impulse noise of the type which consists of sharp impulses having a low rate of recurrence, the situation is somewhat different. It has been pointed out to the author by V. D. Landon of RCA Manufacturing Company, that for the condition of no modulation present, if the receiver is carefully tuned so that the incoming carrier is approximately synchronized with the oscillation frequency of the impulse, the frequency-modulation improvement is maintained for impulse noise which is stronger than the carrier. This effect is shown in the experimentally determined curve of Figure 3 which also shows the effect of a slight detuning from the point of synchronism. For these curves, the peak carrier-noise ratio and the peak signal-noise ratios were measured by means of an oscilloscope coupled to the intermediate-frequency channel output and the audio output of the receiver, respectively. Curve B was taken with the receiver carefully tuned to the impulse noise minimum. Curve A was taken with the carrier detuned about 20 per cent of the maximum frequency deviation of the receiver used. It will be noted that the improvement threshold manifests itself sharply when the receiver is detuned. Hence, for all except the very low passages of modulation the threshold would be present since application of frequency modulation corresponds to a momentary detuning. However, the reduction of the noise during the idle periods of the modulation is undoubtedly very helpful.

The curves of Figure 3 also show the value of carrier-noise ratio required to obtain the full frequency-modulation improvement for impulse noise. It can be seen that for all practical purposes it can be assumed that the full improvement is obtained at a peak carrier-noise ratio of unity.

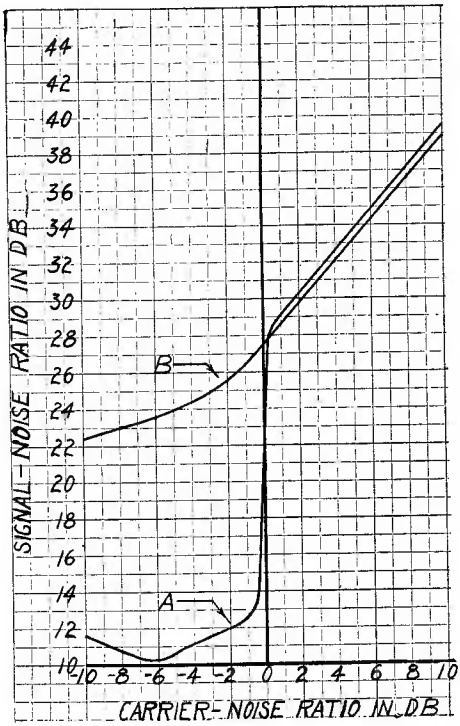


Fig. 3—Measured signal-noise ratio versus carrier-noise ratio characteristics with impulse noise. Receiver deviation ratio = 2.8.
Curve A = receiver detuned by an amount equal to 20 per cent of the maximum frequency deviation.
Curve B = receiver tuned for impulse noise minimum.

A further characteristic of the effect of impulse noise when it is stronger than the frequency-modulation carrier is the noise-silencing action which is present. This produces a sort of a minimum signal-noise ratio for carrier strengths below the improvement threshold when the carrier and pulse frequency are not synchronized. It has been shown[2] that this minimum signal-noise ratio is equal to the deviation ratio of the system. Thus, it may be seen that although the signal-noise ratio will become poorer below the improvement threshold, the signal

may still be serviceable. On the other hand, if the impulse noise is of a continuous nature like fluctuation noise, the noise discriminates against the signal below the improvement threshold and, therefore, smothers the signal in the same manner that fluctuation noise does.

The formula for the distance at which the improvement threshold occurs may be obtained by developing a formula for the distance from the transmitter at which carrier-noise ratios of four and zero decibels are obtained in the intermediate-frequency channel of the frequency-modulation receiver for fluctuation and impulse noises respectively. First—the equation for the carrier-noise ratio in the frequency-modulation intermediate-frequency channel must be determined. This may be derived from (4) which gives the peak carrier-noise ratio as defined for an amplitude-modulation system. For a given radiated carrier, the carrier-noise ratio in the frequency-modulation receiver is less than that defined by (4) by a factor which depends upon the ratio of the effective band widths of the two receivers and upon the type of noise.

For the case of fluctuation noise,

$$C_o/N_o = C/N \times \frac{1}{(F_{fi}/2F_a)^{1/2}} \text{ (fluctuation noise)} \qquad (7)$$

where C_o = carrier level in i-f amplifier of the frequency-modulation receiver,

N_o = noise level in i-f amplifier of the frequency-modulation receiver,

F_{fi} = i-f band width of frequency-modulation receiver,

F_a = band width of audio spectrum it is desired to receive.

For convenience let $Z_f = F_{fi}/2F_a$ for fluctuation noise. Substituting in (7), we obtain:

$$\frac{C_o}{N_o} = \frac{C}{N} \times \frac{1}{\sqrt{Z_f}} \text{ (fluctuation noise)} \qquad (8)$$

The equation corresponding to (8) for impulse noise, which varies directly with the band width instead of as the square-root of the band width, will be:

$$\frac{C_o}{N_o} = \frac{C}{N}\frac{1}{Z_i} \text{ (impulse noise)} \qquad (9)$$

where $Z_i = F_{fi}/2F_a$ for impulse noise.

The factors $\sqrt{Z_f}$ and Z_i, which are the ratios between one-half the intermediate-frequency channel width and the audio channel width of the frequency-modulation receiver for the two types of noise, are

noise-determining ratios and, therefore, must be expressed in terms of *equivalent* channel widths. The equivalent channel widths are different for the two types of noise; hence, a separate factor is used for each type of noise. For fluctuation noise, the equivalent channel width is determined by dividing the area under the energy response curve (the selectivity curve plotted with the ordinates squared) by the height of the curve at resonance. This gives the width of the rectangular channel which would be equivalent to the actual round-topped channel. For impulse noise, the equivalent channel width is determined by dividing the area under the amplitude response curve by the height of the curve at resonance.

A further simplification of the ratios, Z_f and Z_i, may be effected by expressing them in terms of the deviation ratio, μ. Thus,

$$Z_f = K_f \mu \text{ (fluctuation noise)} \tag{10}$$

$$Z_i = K_i \mu \text{ (impulse noise)} \tag{11}$$

$$\text{where } K_f = F_{fi}/2F_d \text{ (fluctuation noise)} \tag{10a}$$

$$\text{and } K_i = F_{fi}/2F_d \text{ (impulse noise)} \tag{11a}$$

in which F_d = maximum applied frequency deviation.

The factors K_f and K_i express the ratio between the equivalent band widths of the intermediate-frequency channel and the total plus and minus peak frequency deviation of the frequency-modulation system. In other words they specify how far out on the intermediate-frequency selectivity curve the frequency deviation may be carried.

In order to determine the formula for the carrier-noise ratio in the frequency-modulation intermediate-frequency channel for the case of fluctuation noise, (4) and (10) may be substituted in (8) which gives:

$$\frac{C_o}{N_o} = \frac{0.0067\sqrt{W}\,ahf\,D_h^{n-2}}{N\,D^n\sqrt{K_f\mu}} \text{ (fluctuation noise, peak values)} \tag{12}$$

Likewise, for impulse noise (4) and (11) may be substituted in (9) which gives:

$$\frac{C_o}{N_o} = \frac{0.0067\sqrt{W}\,ahf\,D_h^{n-2}}{K_i\,\mu\,D^n\,N} \text{ (impulse noise, peak values)} \tag{13}$$

(12) and (13) give the carrier-noise ratio in the intermediate-frequency channel of the frequency-modulation receiver for the two types of noise. Since it is known that the improvement threshold

occurs when these carrier-noise ratios are equal to four and zero decibels, respectively, the distance at which the improvement threshold will occur for a given set of transmission conditions may be determined by equating (12) to 1.585 (four decibels) and (13) to unity (zero decibels), and solving for the distance. Thus,

$$D_i = \left(\frac{0.0042 \sqrt{W}\, ahf\, D_h^{n-2}}{N \sqrt{K_f \mu}} \right)^{1/n} \quad \text{(fluctuation noise)} \qquad (14)$$

$$D_i = \left(\frac{0.0067 \sqrt{W}\, ahf\, D_h^{n-2}}{N K_i \mu} \right)^{1/n} \quad \text{(impulse noise)} \qquad (15)$$

in which D_i indicates the distance at which the improvement threshold occurs.

A study of (14) and (15) shows the effect of a variation of the transmission conditions. For both types of noise the distance to the improvement threshold is directly proportional to the $1/2n$ power of the watts radiated. Hence, for the higher radiation frequencies, where the exponent "n" is large, the improvement threshold distance increases more slowly with increase in power. For fluctuation noise, the improvement threshold distance is inversely proportional to the $1/2n$ power of the product of the factor K_f and the deviation ratio. Thus, as the deviation ratio is increased, the improvement threshold distance decreases and decreases at a less rapid rate when it is beyond the horizon and for the higher radiation frequencies where the exponent "n" is larger. The importance of making the factor K_f as small as possible, by carrying the frequency deviation out on the selectivity curve as far as possible, is also indicated. When the noise is impulse noise (formula 15) the improvement threshold distance is inversely proportional to the "n"th root of the product of the factor K_i and the deviation ratio. Consequently, with this type of noise, the improvement threshold distance decreases more rapidly as these factors are increased.

The distances given by (14) and (15) may be substituted in (5) and (6) to find the frequency-modulation signal-noise ratio existing at the improvement threshold as follows:

$$S_f/N_f \ (at\ D_i) = 1.585 \sqrt{3}\, \mu \sqrt{K_f \mu} = 2.74 \sqrt{K_f}\, \mu^{3/2} \ \text{(fluctuation noise)} \qquad (16)$$

$$S_f/N_f \ (at\ D_i) = 2 \mu K_i \mu = 2 K_i \mu^2 \ \text{(impulse noise)} \qquad (17)$$

Effect of Pre-Emphasizing the Higher Modulation Frequencies

As has been pointed out for the case of amplitude modulation[5], the use of pre-emphasis circuit at the transmitter and a de-emphasis circuit at the receiver produces an overall gain in signal-noise ratio. This gain depends upon the fact that the higher modulation frequencies of voice and program material are of such a small amplitude that their accentuation does not increase the peak voltage of the applied modulating wave as much as the restoring circuit at the receiver reduces the noise. Thus, when the pre-emphasis circuit is inserted at the transmitter, the peak voltage of the modulating wave may increase somewhat so that the modulation level must be lowered, but this loss at the transmitter is overshadowed by the gain at the receiver. Consequently there are two quantities which must be evaluated—the loss at the transmitter and the gain at the receiver.

In order to determine the loss introduced by the pre-emphasis circuit at the transmitter, the following experimental observations were made: The equivalent of a two-string oscillograph was arranged by using two electronically switched amplifiers to feed the vertical plates of an oscilloscope. The outputs of the amplifiers were common, with one of the separate inputs being fed by program material through a pre-emphasis circuit. The gain of the two amplifiers was equalized at a low modulation frequency where the pre-emphasis circuit was not effective. The two amplifiers were alternately switched on by means of a 60-cycle square-wave form and the oscilloscope sweep circuit synchronized with the 60 cycles. The resulting pattern on the oscilloscope screen consisted of two segments of sweep one of which was actuated directly from the program material and the other through the pre-emphasis circuit. Hence, an accurate method of comparing the peak voltages of the two waves could be provided by inserting an attenuator in the pre-emphasized circuit and setting for equal peak voltages indicated by the two traces on the oscilloscope.

The observations made with the electronic-switch oscillograph covered all types of program material and were made for two audio-band widths of 5 and 12 kilocycles. The results of the observations indicated that in general the insertion of the pre-emphasis circuit increases the peak voltage of the level about 2.5 decibels for both the 5 and 12-kilocycle band widths. On certain material such as guitar, harmonica, and piano solos, the level is raised about 4.5 decibels total, but the occasion of such rises is rather infrequent and their duration very short. Hence, a permanent attenuation of 2.5 decibels might be inserted and the volume-limiting equipment relied upon to take care of the occasional higher peaks.

The action of the de-emphasis circuit in reducing the noise at the receiver is somewhat greater for the case of frequency modulation than it is for amplitude modulation. The reason for this is shown in Figures 4 and 5 which show the noise spectrums obtained in the output of the amplitude and frequency-modulation receivers, respectively, with and without the use of a de-emphasis circuit such as would be used with a pre-emphasizing circuit in accordance with R.M.A. Television Transmission Standard M9-218 at the transmitter. In the case of amplitude modulation, the normally evenly distributed noise is concentrated at the lower modulation frequencies. In the case of frequency modulation, the triangular noise spectrum is changed to a spectrum

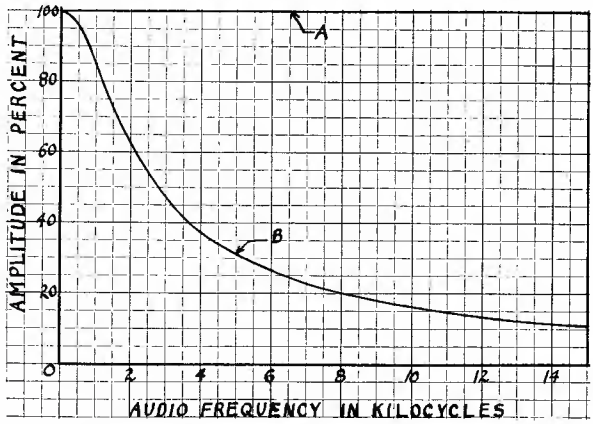


Fig. 4—Amplitude-modulation audio noise spectrums.
Curve A = no de-emphasis used.
Curve B = de-emphasis in accordance with R.M.A. Television Transmission Standard M9-218.

which is practically flat except for the falling off at the lower modulation frequencies.

In determining the relative figures of merit for the noise spectrums of Figures 4 and 5, or, in other words, the noise gains produced by the de-emphasis circuit, there are two determinations which are of importance. The first is the objective comparison which has to do with the relative strengths of the noise as would be measured on a meter. The second is the subjective comparison which takes into consideration the manner of utilizing the signal in the presence of the noise. For program or voice reception, the subjective comparison would be determined by a listening test.

The objective comparison of the noise spectrums of Figures 4 and 5 may be calculated for fluctuation noise by comparing the squared-ordinates areas of the spectrums. Such a comparison gives the ratio

of the energies passed by the two spectrums; the root-mean-square voltage ratio is the square-root of this area ratio. The corresponding comparison for impulse noise may be calculated by comparing the areas of the spectrums. The ratio of the areas of the spectrums gives the peak voltage ratio directly for impulse noise. These areas may be obtained by integration or they may be plotted and a planimeter used. The curves of Figure 6 show the results of such a determination by the use of integration. These curves give the signal-noise ratio gain effected by the de-emphasis circuit alone. To obtain the overall gain due to pre-emphasis, the transmitter loss of 2.5 decibels must be subtracted from the values indicated by the curves.

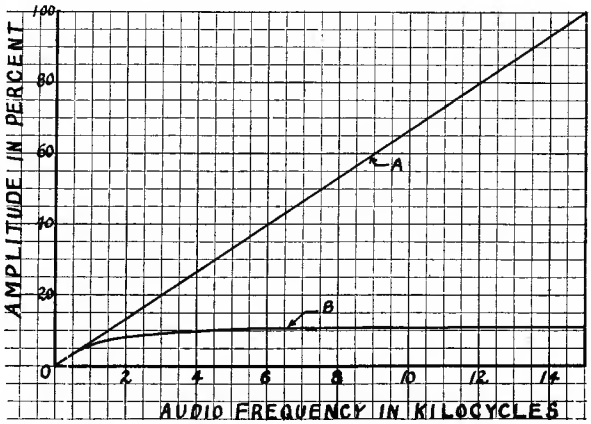


Fig. 5—Frequency-modulation audio noise spectrums.
Curve A = no de-emphasis used.
Curve B = de-emphasis in accordance with R.M.A. Television Transmission Standard M9-218.

In order to obtain the subjective gain effected by the de-emphasis circuit, listening tests were conducted in which the de-emphasis circuit was switched in and out and the annoyance effect of different types of noise compared. Both fluctuation noise and impulse noise were obtained from amplitude- and frequency-modulation receivers to mix with program to produce a signal-noise ratio. The de-emphasis circuit was then switched in and out of the noise while an attenuator was varied to balance the annoyance effect. Audio band widths of 5 and 12 kilocycles were used. The averaged observations of two observers indicated the rather unexpected result that for program and music reception, the subjective effect is practically the same as the objective effect. That is to say that the objective gains as portrayed by Figure 6 may also be taken as the subjective gains that would be realized in a

listening test and the overall gain is obtained by subtracting the transmitter loss of 2.5 decibels from the values obtained from Figure 6.

For speech reception, where the noises were balanced for equal intelligibility, it was found that there was little or no gain effected by the de-emphasis circuit. For the case of amplitude-modulation noise, the use of pre-emphasis would apparently entail a slight loss in intelligibility when the 2.5 decibel transmitter loss was subtracted. With frequency-modulation noise, an intelligibility gain of a few decibels would be realized.

In connection with the listening tests another interesting observa-

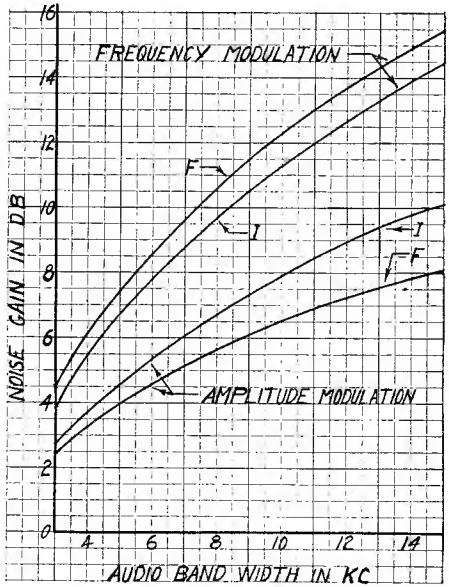


Fig. 6—Calculated noise gains effected by the de-emphasis circuit.
Curves F = fluctuation noise.
Curves I = impulse noise.

tion was made regarding the relative annoying effects of the triangular frequency-modulation noise spectrum and the rectangular amplitude-modulation noise spectrum using tube hiss as a noise source. In this test the noises were first balanced with a meter and then they were mixed with speech and the relative levels readjusted until they impaired the intelligibility of the speech the same amount. The averaged observations of three observers showed that about 8 decibels more noise could be tolerated with the triangular frequency-modulation noise than with the rectangular amplitude-modulation noise.

EXAMPLE

The curves of Figures 7, 8, 9, and 10 have been calculated for the following assumed transmission conditions for a high-fidelity broadcast system:

Transmitting antenna height	= 800 feet.
Receiving antenna height	= 30 feet.
Audio channel	= 15 kilocycles.
Frequency	= 42 megacycles.
Maximum frequency deviations	= 20 and 75 kilocycles.

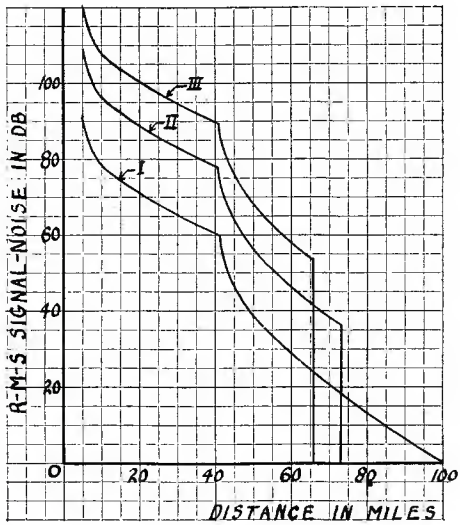


Fig. 7—R-M-S signal-noise ratio versus distance. Fluctuation noise = 1 peak microvolt per meter.
Curve I = amplitude modulation, 500 watts radiated.
Curve II = frequency modulation, 1000 watts radiated, maximum frequency deviation = 20 kilocycles.
Curve III = frequency modulation, 1000 watts radiated, maximum frequency deviation = 75 kilocycles.

Powers of one and 100 kilowatts radiated were used for frequency modulation and these powers were halved for the corresponding amplitude-modulation calculations. This two-to-one power gain effected at the frequency-modulation transmitter was taken instead of the usual four-to-one factor since it represents the gain that would be effected by a frequency-modulation system over the most efficient amplitude-modulation system which is the high-level type of modulation system. If the modulator tubes were paralleled with the final amplifier tubes

in such a system, the increase in power would be two-to-one. The usual factor of four-to-one assumes the more inefficient types of low-level modulation.

The values of signal-noise ratio obtained for all of the curves were corrected by subtracting one-half the maximum fading range as portrayed by Figure 2. Formula (4) was used to calculate the average signal-noise ratio for amplitude modulation and the fading correction was applied to obtain the minimum signal-noise ratio. It was assumed that the transmission conditions of the fading correction curve were near enough to the conditions of this example to allow this direct correction without interpolation.

Where the propagation curves of the corresponding amplitude-modulation system were to be determined as well as those of the frequency-modulation systems, which was the case in this example, the simplest procedure was to draw the amplitude-modulation curves and then construct the frequency-modulation curves which follow lines parallel to the amplitude-modulation curve, but at higher signal-noise ratio levels for distances within the improvement threshold distances. To do this the average amplitude-modulation signal-noise ratios for fluctuation noise were calculated from (4). These were corrected for fading and the pre-emphasis gain of 5.6 decibels was added. For the low-deviation frequency-modulation system, the fluctuation noise gain is equal to 1.73 times the deviation ratio, μ ($\mu = 20/15 = 1.33$) or 7.2 decibels. The use of pre-emphasis adds another 7.4 decibels to the frequency-modulation gain as compared to the amplitude-modulation system with pre-emphasis since the total gain due to pre-emphasis on the frequency-modulation system is 13 decibels. The transmitter power gain adds another 3 decibels. Hence the curve for frequency modulation with a deviation of 20 kilocycles is $7.2 + 7.4 + 3 = 17.6$ decibels higher than the corresponding amplitude-modulation curve for the region within the threshold of improvement distance of the frequency-modulation system. The frequency-modulation system with a 75-kilocycle deviation has a larger deviation ratio ($\mu = 75/15 = 5$) so that its total gain over the amplitude-modulation system with pre-emphasis is $18.7 + 7.4 + 3 = 29.1$ decibels.

For the case of impulse noise, the frequency-modulation gains are equal to twice the deviation ratio and the pre-emphasis gains are also different. For amplitude modulation, the pre-emphasis gain is 7.5 decibels and for frequency modulation it is 4.5 decibels more or a total of 12 decibels. Hence, the gain of the frequency-modulation system using a 20-kilocycle deviation is $8.5 + 4.5 + 3 = 16$ decibels. The corresponding gain for the system using a deviation of 75 kilocycles is $20 + 4.5 + 3 = 27.5$ decibels.

In order to determine the signal-noise ratios at the improvement threshold distances, the gains due to pre-emphasis must be added to the signal-noise ratios calculated from (16) and (17). The pre-emphasis gains to be added in this case are the total gains of 13 and 12 decibels for fluctuation and impulse noise respectively. Before formulas (16) and (17) can be applied, the factors K_f and K_i must be evaluated. These factors are the ratios between the equivalent band widths of the frequency-modulation receiver intermediate-frequency channel and the total plus and minus frequency deviation. It is obvious that, in order to obtain the greatest distance to the improvement threshold, the equivalent band width must be made as small as the frequency deviation of the system will allow. The limitations encountered are the introduction of harmonic distortion and the introduction of amplitude modulation due to the frequency variation exceeding the flat-topped portion of the selectivity characteristic. In practice the limiter tends to take care of this departure from a flat-topped selectivity characteristic. Some rather preliminary measurements have indicated that the deviation may extend out to about 2.5 decibels down (down to 75 per cent) on the sides of the selectivity curve without producing harmonics which are too high for high-fidelity reception.

A study of typical intermediate-frequency equivalent band widths has shown that the band width 2.5 decibels down is about 90 per cent of the equivalent band width for fluctuation noise and about 78 per cent of that for impulse noise. The factors K_f and K_i are equal to the reciprocals of these percentages or 1.1 and 1.3, respectively.

The above evaluations of K_f and K_i are based on somewhat incomplete distortion measurements and assume no guard band to take care of tuning drift in the case of the frequency-modulation systems. Further work is undoubtedly necessary in these respects, but it is believed that this evaluation will serve the purposes of this paper.

With these evaluations of K_f and K_i, formulas (16) and (17) simplify to:

$$S_f/N_f \text{ (at } D_i) = 2.9\ \mu^{3/2} \text{ (fluctuation)} \qquad (16a)$$

$$S_f/N_f \text{ (at } D_i) = 2.6\ \mu^2 \quad \text{(impulse)} \qquad (17a)$$

Applying (16a) to the system with a deviation of 20 kilocycles and adding the gain due to pre-emphasis of 13 decibels gives 25.9 decibels for the peak signal-noise ratio at the improvement threshold distance. The same calculation for the case of a 75-kilocycle deviation gives 43.2 decibels. For impulse noise (17a) is used and a pre-emphasis gain of 12 decibels is added to give ratios of 25.2 and 48.3 decibels for the two values of frequency deviation.

The curves of Figures 7 and 8 for fluctuation noise assume a thermionic agitation and tube-hiss noise level equivalent to one peak microvolt per meter (0.224 r-m-s microvolts per meter). This corresponds to an r-m-s noise voltage of 0.52 microvolts in series with a dummy antenna at the input terminals of the amplitude-modulation receiver. Such a noise voltage is about that which would be obtained with good design using an 1852 radio-frequency amplifier tube. All of the ratios for these curves were converted from peak to r-m-s ratios to correspond to the general practice in considering this type of noise. This was done by adding 10 decibels to the ratios. The figure of 10

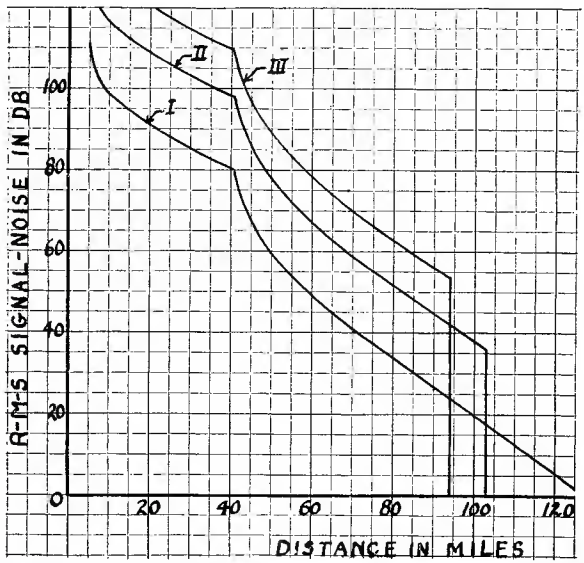


Fig. 8—Same as Fig. 7, but with amplitude modulation radiated power = 50 kilqwatts, frequency-modulation radiated power = 100 kilowatts.

decibels is obtained from the data of the author's previous paper[2]. In that paper the crest factor of fluctuation noise is evaluated at 13 decibels. Subtracting the 3 decibel crest factor of the signal gives what might be termed the crest factor of the signal-noise ratio as 10 decibels.

The curves of Figures 9 and 10 for impulse noise assume a noise field intensity of 100 peak microvolts per meter for an effective band width of 30 kilocycles (audio band width = 15 kilocycles). This intensity is about that which would be received with horizontal polarization from the ignition system of the average automobile at a location about 125 feet from the road over which the automobile travels. The curves for this type of noise have been plotted with peak signal-noise ratios

since root-mean-square values have little significance due to the very high and variable crest factor.

At distances beyond the improvement threshold for impulse noise, the curves are plotted for the condition of full modulation in which the improvement threshold manifests itself. Under this condition, the signal-noise ratio is limited to a value which is equal to the deviation ratio when pre-emphasis is not being used. It will be remembered that this ratio is what might be termed the "silencing" signal-noise ratio since the noise tends to punch holes in the signal, but when the noise pulses have a relatively slow rate of recurrence, this condition is quite

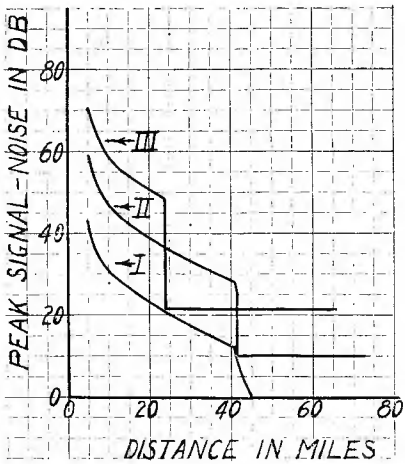


Fig. 9—Peak signal-noise ratio versus distance. Impulse noise = 100 peak microvolts per meter.
Curve I = amplitude modulation, 500 watts radiated.
Curve II = frequency modulation, 1000 watts radiated, maximum frequency deviation = 20 kilocycles.
Curve III = frequency modulation, 1000 watts radiated, maximum frequency deviation = 75 kilocycles.

tolerable. On the other hand, if the pulses have a high rate of recurrence, the noise tends to smother the signal and service is limited to the improvement threshold distance as in the case of fluctuation noise. When pre-emphasis is being used, a gain is added to this limited signal-noise ratio which is equal to the pre-emphasis gain of 7.5 decibels for amplitude-modulation noise. The pre-emphasis gain for amplitude modulation noise is used in this case since the noise in the silencing condition has lost its triangular spectrum characteristic of the frequency-modulation noise received above the improvement threshold. Thus, for the system with a 20-kilocycle deviation, the signal-noise ratio in the silencing condition beyond the improvement threshold dis-

tance is 2.5 + 7.5 = 10 decibels. For the system with a 75-kilocycle deviation it is 14 + 7.5 = 21.5 decibels.

At distances beyond the improvement threshold for impulse noise when the modulation is in the idle condition, the signal-noise ratios are somewhat higher than those shown by the curves due to the synchronization effect pointed out by V. D. Landon. This results in a considerable reduction of the annoyance effect of the noise. Hence, the curves may be taken as somewhat pessimistic, but satisfactory for comparison purposes.

Discussion

It is apparent from the curves that the maximum distance is served if the maximum frequency deviation is such that the minimum

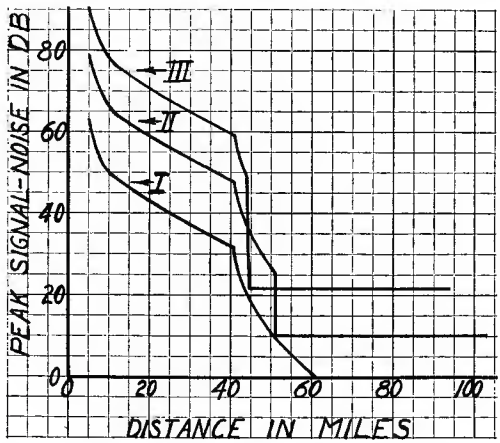


Fig. 10—Same as Figure 9, but with amplitude-modulation radiated power = 50 kilowatts, frequency-modulation radiated power = 100 kilowatts.

tolerable signal-noise ratio exists at the improvement threshold distance. Hence, the choice of the optimum deviation hinges on the definition of the minimum tolerable signal-noise ratio. If the figure of 30 decibels root-mean-square for fluctuation noise is taken as acceptable (this figure has appeared in the literature[6,7] as a commercially satisfactory signal-noise ratio), it is seen from Figures 7 and 8 that a maximum deviation of 20 kilocycles is more than adequate for all distances out to the improvement threshold distance since the lowest signal-noise ratio, which occurs at the improvement threshold distance for this type of noise, is 36 decibels. A deviation of less than 20 kilocycles is probably inadvisable since this would prevent the possible future transmission of an audio fidelity of zero to 20 kilocycles.

It is seen from Figures 7 and 8 for fluctuation noise that at the distance corresponding to the improvement threshold for a deviation of 75 kilocycles, a broadcasting service using a deviation of 20 kilocycles yields a signal-noise ratio of 42 decibels. For the same distance, a service using a 75-kilocycle deviation yields a signal-noise ratio of 53 decibels. At this point it is apparent that a 20-kilocycle deviation produces a signal-noise ratio which is comfortably above what would be considered commercially satisfactory, with an attendant conservation of band width in the available portion of the frequency spectrum. However, if it were assumed that a signal-noise ratio of 53 decibels is necessary, this ratio may be obtained with a deviation of 20 kilocycles at a somewhat shorter distance from the transmitter. For the conditions of the curves of Figures 7 and 8 this distance would be 20 per cent shorter. On the other hand, a 20-kilocycle deviation is capable of furnishing what would be considered better than acceptable service out to a distance which is about 10 per cent greater than the maximum service range for a 75-kilocycle deviation.

When the noise is impulse noise as portrayed by Figures 9 and 10, it can be seen that the general shape of the curves are similar to those for fluctuation noise. If the frequency of recurrence happens to be high so that the noise is continuous, the silencing properties of the frequency modulation cannot be taken advantage of. Instead, the noise, if stronger than the carrier, will depress the signal so as to smother it. For this type of noise a low-deviation service will have a greater range than a high-deviation service by an amount that is considerably larger than the corresponding case for fluctuation noise. This is especially true where the transmitter power is low enough to cause the improvement threshold to occur within the horizon as is the case with the conditions of Figure 9.

If the impulse noise happens to be automobile ignition where the rate of recurrence of the pulses is rather infrequent and the time duration of the impulses short, the noise-silencing properties of the frequency-modulation system may be taken advantage of when the noise is stronger than the carrier. With this type of noise, the effectiveness of the silencing action increases as the deviation is increased. However, for this type of noise, even a system with a low deviation produces a noise output which has a rather low annoyance value.

It is apparent that the service obtainable where fluctuation noise predominates, may be predicted with a fair degree of accuracy. However, owing to the highly variable character and distribution of impulse noise, predictions are difficult regarding this type of noise. Hence, it is felt that it is highly desirable to conduct field tests to further

compare the results of different deviations under actual service conditions.

Conclusions

From the above considerations, it seems apparent that if fluctuation noise is the primary limitation, a frequency-modulation broadcasting system using a deviation of 20 kilocycles will produce a signal-noise ratio greater than the values normally considered acceptable, for all distances out to the distance at which the improvement threshold occurs. If the limitation is impulse noise a similar relationship exists, but it is felt that further field work is desirable to study the results under actual service conditions.

The effect of increasing the deviation of the system is to reduce the distance at which the threshold effect is realized, but at the same time the signal-noise ratio at distances equal to or less than that distance is improved.

The interests of band width conservation are, of course, best served by a choice of the lowest value of deviation that will yield an acceptable signal-noise ratio out to the threshold distance. It would also seem that the same choice will yield the desired service at the lowest cost.

In order to evaluate carefully these effects under a large number of actual service conditions, a series of field tests are being undertaken, using receivers designed for optimum performance at each of the different values of deviation under consideration.

Acknowledgment

The guidance and helpful suggestions of Mr. H. O. Peterson, and the assistance of Mr. R. E. Schock in the experimental work of this paper, are gratefully acknowledged.

References

[1] Murray G. Crosby, Frequency-Modulation Propagation Characteristics, *Proc. I.R.E.*, Vol. 24, No. 6; June (1936).

[2] Murray G. Crosby, Frequency-Modulation Noise Characteristics, *Proc. I.R.E.*, Vol. 25, No. 4; April (1937).

[3] H. H. Beverage, Some Notes on Ultra-High-Frequency Propagation, *RCA Review*, Vol. 1, No. 3, January (1937).

[4] K. G. MacLean and G. S. Wickizer, Notes on the Random Fading of 50-Megacycle Signals Over Nonoptical Paths, *Proc. I.R.E.*, Vol. 27, No. 8; August (1939).

[5] In an unpublished engineering report written in July (1933) and entitled: "Tests of a System for Reducing Interfering Noises in Radio Transmission", Messrs. I. Wolff, G. L. Beers, and L. F. Jones of RCA Manufacturing Company describe tests which indicated improvements of from 8 to 10 decibels in an amplitude-modulation program system.

[6] I. R. Weir, Comparative Field Tests of Frequency-Modulation and Amplitude-Modulation Transmitters, *Proc. The Radio Club of America*, Vol. 16, No. 2; July (1939).

[7] C. V. Aggers, Dudley E. Foster, and C. S. Young, Instruments and Methods of Measuring Radio Noise, *A.I.E.E. Technical Paper 39-144*, Preprinted for A.I.E.E. Pacific Coast Convention, San Francisco, Calif., June 26-30, 1939.

IMPULSE NOISE IN F-M RECEPTION*†

By

VERNON D. LANDON‡

RCA Manufacturing Co., Inc.

Summary—*An investigation into the nature of impulsive noise in frequency modulation receivers, making extensive use of oscillographic records. The relative effects of the plate-volt and grid bias types of limiter on impulsive noise are compared and recommendations made to insure maximum noise reduction.*

IMPULSIVE noise in frequency-modulation receivers has already been studied and reported by M. G. Crosby[1] and others. The purpose of the present article is to carry out the study in more detail. Certain theoretical considerations are presented and oscillographic confirmation shown.

When a shock, such as a noise pulse, is applied to the input of a radio receiver, a wave train or pulsation passes through the amplifier. The output of the final stage of the amplifier is a waive train which is characteristic of that particular receiver. The frequency of the wave train is the frequency of the center of the pass band, and the time duration is $1/f_c$ where f_c is one-half of the bandwidth. Typical response waveforms are shown in Figures 1A, 1B, and 1C. These oscillograms were taken on the 5 Mc intermediate frequency amplifier of a frequency modulation receiver having 200 kc bandwidth. In the cases illustrated the time duration is about $\frac{1}{100{,}000}$ or 10 microseconds. Thus, there are about 50 cycles present in the main portion of the wave train. The shape of the envelope of this wave train is a function of the shape of the selectivity curve. There is a tendency for the voltage envelope to rise to a peak, drop to zero, and then reappear in the form of a secondary lobe with the phase of the radio frequency wave reversed. The size of the secondary lobe is a function of the shape of the nose of the selectivity curve. A slightly double peaked selectivity curve results in a marked secondary lobe as shown in

---

* Decimal Classification: R630.3.

† Reprinted from *Electronics*, February, 1941.

‡ Now with the Research Department, RCA Laboratories Division, Princeton, N. J.

[1] M. G. Crosby, "Frequency-Modulation Noise Characteristics", *Proc. I.R.E.*, Vol. 25, April 1937, No. 6, p. 472.

Figure 1A. The phase of the r-f wave reverses at the valley between the lobes. A flat-topped selectivity curve gives a smaller lobe as shown at *B*. The lobe disappears as shown at *C* if the selectivity curve is round nosed. The valley between the main lobe and the secondary lobe goes to zero as in *A* if the selectivity curve is symmetrical. If one peak of the selectivity curve is higher than the other, the valley between the lobes does not go to zero as shown at *D*.

When a wave train such as that shown in Figure 1A is fed to a discriminator of the type usually used in frequency-modulation receivers, the output has the form shown in Figure 1G. If the circuit

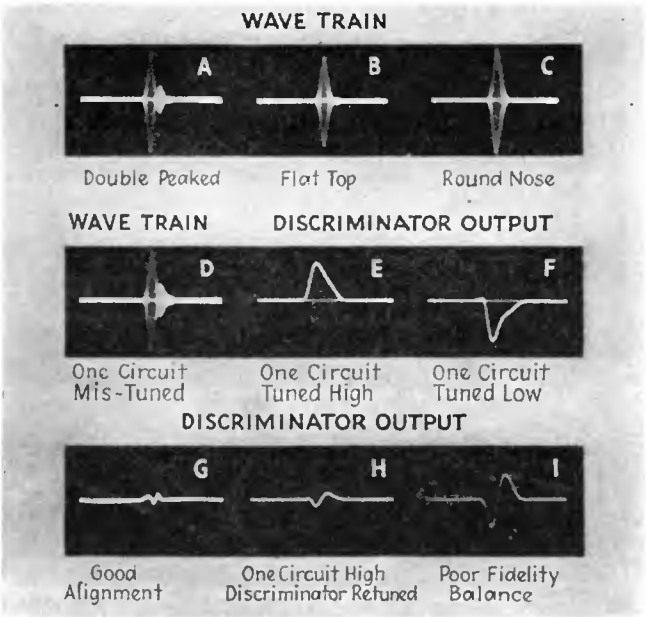


Fig. 1—Oscillograms of noise wave train and output of discriminator for various conditions of tuning.

were perfectly balanced, the output would be zero for this condition. The residual wiggles are due to slight imperfections in the balance.

If one of the circuits of the receiver is detuned so as to give a selectivity curve with the low-frequency peak higher than the other, producing a wave train as in Figure 1D, the discriminator output has the form shown at *E*. If the circuit is tuned to the other side making the high-frequency peak greater, the discriminator output changes to that shown at *F*.

Similar curves to *E* and *F* are obtained by mistuning the secondary of the discriminator one way or the other. Thus, it appears that the

output of the discriminator goes plus or minus on the oscillogram according to whether the selectivity curve has more area above or below the frequency of resonance of the discriminator secondary. Oscillogram *H* of Figure 1 shows the result of having the low-frequency peak exceed the high-frequency peak and of attempting to compensate by discriminator tuning. The indication is that the frequency of the pulse varies slightly with time for this condition.

The output network of the discriminator consists of two separate diode output circuits connected in series. The fidelity of these two circuits should be the same for a good balance, but this is not usually

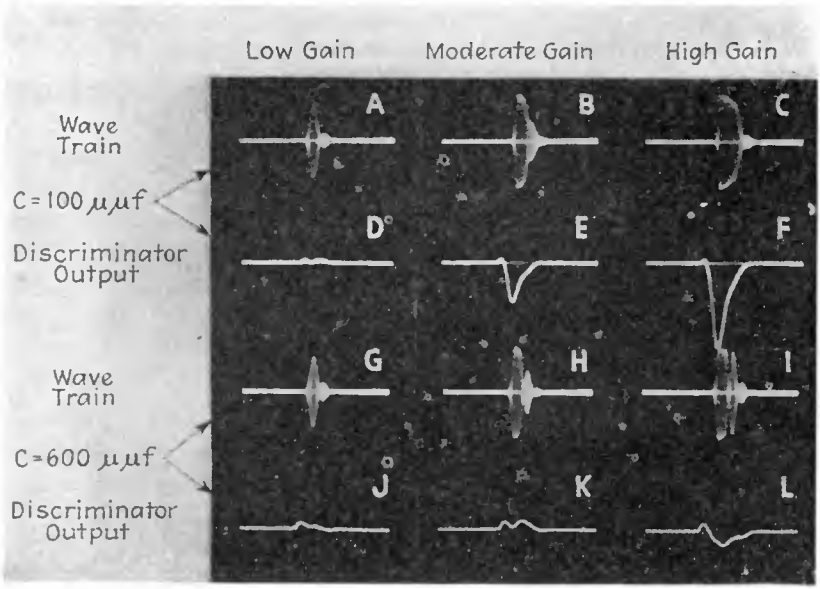


Fig. 2—Oscillograms of noise wave train and output of discriminator showing detuning due to grid current at higher amplitudes.

obtained. The output circuit is grounded on one side instead of in the center. This usually results in more capacitance across one side than across the other. This can be compensated for by adding capacitance to the other side. The effect of a capacitance unbalance of 10 micromicrofarads across 100,000 ohms is shown at *I* in Figure 1.

Change of Tuning with Signnal Strength

An unfortunate effect is the change in tuning with signal strength. This is illustrated in Figure 2. The oscillograms of this figure were taken in the output circuit of a limiter. Limiter action was obtained because of grid current and because of the low plate voltage employed.

At *A* the input level was too low for limiting to take place. The narrow neck between lobes indicates accurate alignment. The oscillograms at *B* and *C* were taken at progressively higher signal levels. The deep valley between lobes is missing indicating that certain circuits were detuned. The corresponding oscillograms of discriminator output are shown at *D*, *E* and *F*. The fact that *E* and *F* go negative shows that certain circuits are tuned to a lower frequency at the higher signal level. The circuit which is detuned most is the grid circuit of the limiter. This is proved by the next six oscillograms. These repeat the conditions of the preceding six except that the grid of the limiter is now fed by a tuned circuit having 600 micromicrofarads instead of 100 micromicrofarads for a tuning condenser. The detuning at high signal level is correspondingly less as illustrated by the deep valley between lobes at *H* and *I* and the smaller negative deflection at *K* and *L*.

The detuning caused by grid current can also be minimized by other methods. If the affected transformer is made broader than the preceding stages the detuning will not affect the overall performance. It also helps if the *Q* of the primary and the *Q* of the secondary are equal and if the coupling is less than critical in the affected transformer.

Noise and Signal Applied Together

When signals are being received, all the foregoing comments will apply only if the noise exceeds the carrier by about ten times or more. It is more usual for the noise to be less than the carrier or about equal to it. Under these conditions the noise output is a function of the reaction between the noise wave train and the carrier wave. The noise may come through purely as amplitude modulation, as frequency modulation, or as both, depending on the relative phase angle between the noise wave and the carrier wave.

To demonstrate these facts with an oscillograph it was necessary to build some special equipment. The difficulty with ordinary equipment is to synchronize the impulse generator with the signal generator so that the relative phase stays constant. Synchronization is obviously almost impossible with separate sources. To get around this difficulty the two waves were obtained from the same source. The source chosen was a 10,000 cps oscillator which was fed into two channels. In one channel the frequency was multiplied through a succession of stages up to the desired frequency of 5 megacycles. In the other channel the 10,000 cps wave was used to generate very narrow unidirectional pulses at the rate of 10,000 per second. The impulses

and the 5-Mc signal were applied simultaneously to the input of a 5-Mc amplifier having a 200-kc bandwidth. Since the carrier wave and the impulses had a common source, the wave trains due to the impulses always had the same phase relative to the carrier. However, the relative phase was adjustable by a slight change in the tuning of one of the circuits in the frequency multiplier.

Some of the oscillograms taken with this equipment are shown in Figure 3. In this figure the vertical column of oscillograms on the left

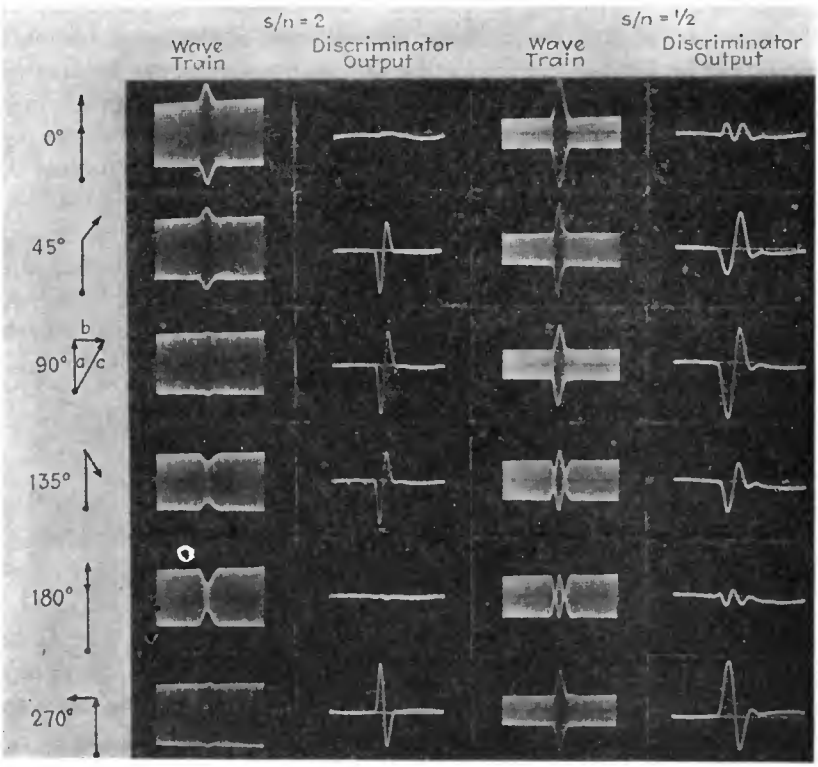


Fig. 3—Right, oscillograms of noise and discriminator output for various phase angles between the noise impulse and signal at two ratios of signal-to-noise with the carrier centered in the pass band.

illustrates the resultant wave form when the noise wave and the carrier wave are added with various phase angles. The second column illustrates the output of the discriminator for the same conditions. It can be seen that the zero and 180-degree phase angles produce the maximum amount of amplitude modulation and the minimum output from the discriminator. The 90-degree and 270-degree phase angles produce minimum amplitude modulation and maximum output from

the discriminator. The intermediate phase angles produce both types of modulation.

The fact that the 90-degree phase angle produces the equivalent of frequency modulation can be seen by referring to the vector diagram on the left. The vector *a* represents the carrier wave, the vector *b* represents the noise wave at the instant of its peak value and the vector *c* represents the resultant when *a* and *b* are added. For the first half of the duration of the noise wave the vector *b* is growing. For the second half it is diminishing. During the time interval when *b* is growing, the phase angle between *a* and *c* is growing. Since frequency is the rate of change of angle, the frequency of the vector *c* is low during the growth of *b* and high while *b* is decreasing. If *b* is much smaller than *a* so that the angle is equal to its sine, then the frequency deviation is proportional to the rate of change of *b*.

In an amplitude-modulation receiver the output is a maximum for a 0-degree or 180-degree phase angle and the output wave form follows the envelope of the noise wave train. In a frequency-modulation receiver the output is a maximum at a 90-degree or 270-degree phase angle, and the output waveform is the first derivative of the envelope of the noise wave train. It should be noted that the output of the discriminator for 270 degrees is inverted compared to that for 90 degrees.

Columns 3 and 4 in Figure 3 repeat the same tests except that a different signal-to-noise ratio was used. For columns 1 and 2 the signal amplitude was twice that of the noise. For columns 3 and 4 the signal had half the noise amplitude. It can be seen that with the stronger relative value of noise, the noise shows up as a peak even for the 90-degree and 135-degree phase angles. For 180 degrees the noise bucks out the carrier completely and in addition there is a lobe in which the r-f voltage is 180 degrees out of phase with the carrier. In spite of these differences the output of the discriminator has very much the same waveform for one signal noise ratio as for the other. This may be due to the inability of the audio circuit to follow the higher-frequency components of the true wave form. The fidelity would probably have to be flat to several hundred kilocycles to follow the output waveform accurately.

Effectiveness of Limiter

Oscillograms of this nature can be used to test the effectiveness of various types of limiters as demonstrated by Figure 4. In this figure the first column of oscillograms is for a signal-to-noise ratio of 2 and shows the output of a plate voltage limiter. This type of limiter obtains the limiting action by using a very low plate voltage. The r-f plate

voltage swing is limited to something less than the d-c plate voltage employed.

The first oscillogram of the left-hand column is for 0 degrees phase angle with a moderate amount of gain so that only partial limiting is obtained. For the second oscillogram the gain was increased so that limiter action was almost perfect. The third and fourth repeat the conditions except that the phase angle between signal and noise is 180 degrees. This shows that the plate voltage limiter is a rather good limiter.

The second column of oscillograms is for a grid-leak limiter. This type of limiter obtains the limiting action by allowing the tube to bias

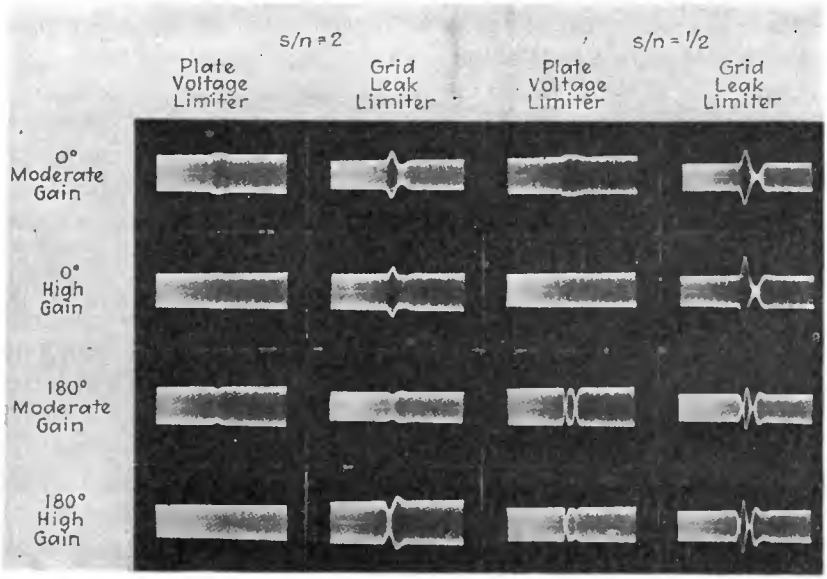


Fig. 4—Oscillograms demonstrating the relative effectiveness of plate voltage and grid leak limiters.

itself off as the signal increases in amplitude. It is obviously important to keep the circuit response as fast as possible. For these oscillograms 100,000 ohms and 20 micromicrofarads were used. The other conditions for the second column are the same as for the first. These tests seem to indicate that the grid-leak limiter is not as satisfactory as the plate voltage type. The grid condenser apparently cannot charge up rapidly enough to follow the pulsation. However, it does build up some extra bias during the pulsation, and after it is over, the carrier is attenuated for a short time until the extra bias leaks off.

The third and fourth columns in Figure 4 repeat the conditions of

the first and second except that the noise amplitude is twice that of the signal. For a phase angle of 0 degrees the plate voltage limiter still does a good job, but the defects of the grid-leak limiter are magnified. For a phase of 180 degrees even the plate-voltage limiter fails as it obviously must since the phase of the resultant wave is reversed during the central portion of the noise pulse.

Noise with Carrier Detuned

As a matter of fact, these oscillograms are not as important as they seem at first sight. As was shown in Figure 3, amplitude modulation does not come through the discriminator when the noise pulse and the carrier are synchronous. Thus a limiter does no good under conditions of no modulation if the signal carrier is tuned to the center of the pass band.

Actually, the limiter action may make the noise worse, as was shown in Figure 2 in which grid current was shown to detune the corresponding tuned circuit. This is also illustrated in the bottom oscillogram of the third column of Figure 4. It will be noted that the valley at the end of the lobe is not as deep as it should be. This is a sign of detuning due to grid current.

The only time a limiter helps to reduce the noise is when the carrier is not in the center of the pass band because of mistuning or modulation. If the carrier wave is tuned to the edge of the pass band, there is a beating action between the noise wave train and the carrier wave. The frequency of the beat note is f_c where f_c is one-half the bandwidth of the receiver. Thus, the beat note goes through one cycle in a period of time equal to $1/f_c$. However, the duration of the pulse itself is also $1/f_c$ (this is accurately true in a sharp cutoff filter and roughly true in any filter depending on what per cent of peak amplitude is called cutoff). Thus, the beat note goes through one complete cycle in the duration of the noise wave train. This is illustrated in Figure 5. In the first oscillogram of the first column, the noise wave starts out 180 degrees out of phase with the carrier wave. As the noise wave grows, it gradually changes its phase relative to the carrier so that at the peak of the pulse it is in phase. The relative phase continues to change for the duration of the pulse so that at the end of the pulse it is again 180 degrees. Since the relative phase assumes all possible values in the duration of each pulse it is difficult to assign a value to it. However, the relative phase at the peak of the pulse has a definite meaning and it is on this basis that the various oscillograms of the first column are labelled. It can be seen that as the relative phase is changed, the valley of the beat note travels progressively across the noise pulse. The corre-

sponding output of the discriminator is shown in the second column.

For the detuned condition characteristic of the oscillograms of Figure 5, it can be seen that the greatest noise output occurs for the 180-degree position. This can be partially explained simply on the basis of amplitude modulation. The carrier alone is producing a d-c output corresponding to peak frequency deviation on the discriminator curve. The reduced amplitude during the pulse reduces the d-c output voltage momentarily.

The third and fourth columns of Figure 5 repeat the conditions of

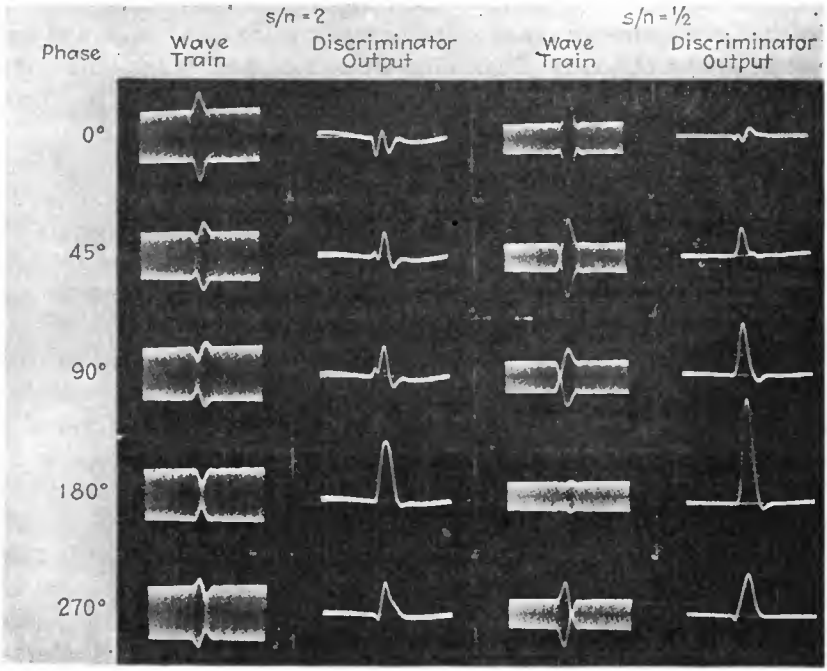


Fig. 5—Oscillograms of noise and discriminator output with carrier tuned to edge of the pass band.

the first and second columns except that the signal-to-noise ratio is changed from 2 to ½. A curious effect is obtained in the third column for 180-degree phase angle. For this condition the amplitude-modulation component disappears almost completely. In spite of the fact, the noise output is a maximum for this phase angle. The reason for this is shown in Figure 6. The vector a represents the carrier wave and the vectors b_1, b_2, b_3 and b_4 represent the successive values assumed by the noise pulse. The vector c represents the resultant wave (shown in one position only). Obviously c makes one complete revolution around the origin. In other words the resultant wave picks up one extra cycle in

the duration of the pulse. The angular velocity of the vector c is twice that of b at the peak value of b. Thus, at the peak of the pulse the resultant wave has an instantaneous frequency lying at the opposite edge of the pass band from the signal carrier wave. This results in maximum noise output. If similar circle diagrams are drawn for the other phase angles, they account for the results quite well.

Effectiveness of Limiter with Signal Detuned

Obviously a limiter can do practically no good for the set of conditions illustrated by Figure 6, since there is almost no amplitude modulation present. It is also true that the limiter does little good at any phase angle if the noise is stronger than the signal. For cases where the signal is stronger than the noise, and the signal is tuned to the edge

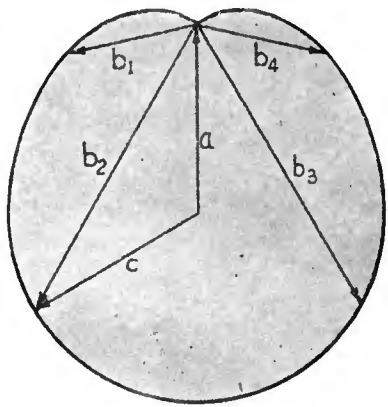


Fig. 6—Vector diagram corresponding to Fig. 5. Signal-to-noise ratio is 0.5. Relative phase is 180 degrees at peak of noise.

of the pass band (columns 1 and 2, Figure 5) the limiter can do a certain amount of good, as is illustrated in Figure 7. Column 1 of this figure is for a signal noise ratio of 2, with the signal tuned to the high-frequency side of the pass band. This is a duplication of column 2 of Figure 5 except that the signal is tuned to the opposite side of the pass band.

Column 2 was taken under the same set of conditions as column 1, except that a plate voltage limiter was used. The oscillograms of column 2 represent an appreciable noise reduction over those of column 1. Strangely enough, column 3 which is for a grid-leak limiter, appears to be equally good.

Perhaps it is not quite self-evident that column 2 does represent a

noise reduction over column 1. This can be made clear by referring to Figure 8.

Frequency Spectrum of Certain Pulsations

In Figure 8 several forms of pulsation are shown with the corresponding energy distribution curves. The first pulsation is called "unit

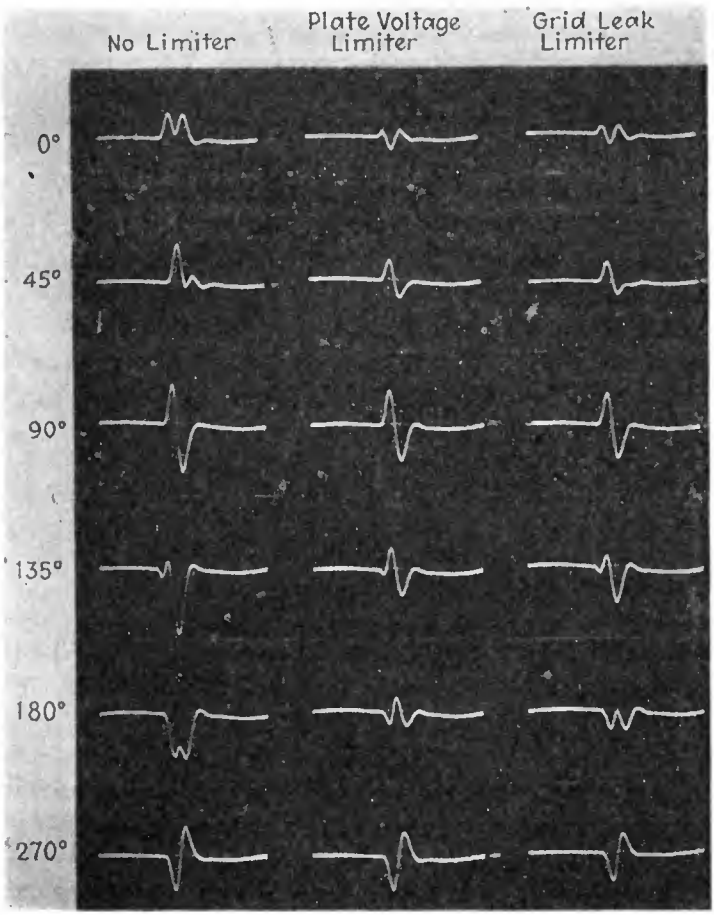


Fig. 7—Oscillograms illustrating noise reduction by plate voltage and grid leak limiters.

step." The corresponding distribution curve has its amplitude inversely proportional to the frequency.

If the first derivative of the unit step is taken the result is called unit impulse. It is supposed to have infinite amplitude and zero time

duration but it encloses unit area. Unit impulse has the same amount of energy at all frequencies.

If a unit impulse is applied to an ideal low pass filter, the theoretical output is $\frac{\sin 2\pi f_c t}{t}$, which is the pulsation illustrated. The energy distribution is uniform out to the cutoff frequency.

If the first derivative of this pulsation is taken, the result is the pulsation illustrated at the bottom of the figure. The energy content of this wave is directly proportional to the frequency up to cutoff.

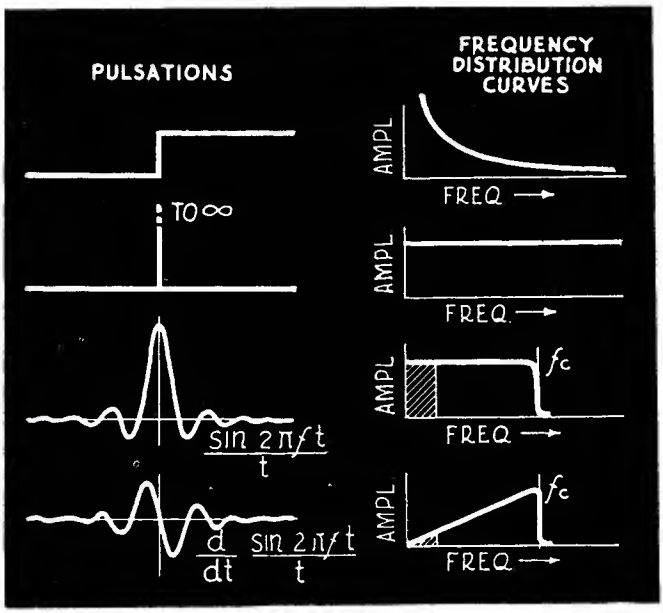


Fig. 8—Various types of pulsations and the corresponding curves of relative amplitude vs. frequency.

It is apparent that many of the oscillograms representing the discriminator output are similar to one or the other of these pulsations. Now the bandwidth of a frequency-modulation receiver is 200 kc, so that from a modulation standpoint the cutoff is 100 kc. The energy content in the audio spectrum is that represented by the shaded area in each case. It is obvious that much less energy is present in the audio spectrum of the pulsation illustrated at the bottom of the figure. One way of looking at it is that the low-frequency components of the wave tend to cancel out if the area in the pulse above the axis is equal to the area below the axis.

Theory of Improvement Due to Limiter

Referring again to Figure 7, there are several oscillograms in column 1 which do not have equal area above and below the axis. This inequality is almost eliminated in columns 2 and 3, indicating a reduction in the low-frequency energy content. The equalization of area is to be expected from a theoretical standpoint. If the noise amplitude is less than the carrier amplitude, the resultant wave can neither gain nor lose a cycle. If no cycles are gained nor lost, then the frequency must average the same as the carrier. Thus if the limiter takes out amplitude variations in the resultant wave, the output of the discriminator must have equal area above and below the axis. Thus, a limiter does reduce the amount of noise if the carrier is detuned and if the noise is weaker than the signal. When the signal is accurately tuned in, the improvement due to the limiter takes place only on noise pulses which occur while the carrier is deviated from center by modulation.

Recommendations for Improved Performance

From the data provided by all these oscillograms, certain recommendations can be made for improving the performance of frequency-modulation receivers:

1. If the best noise reduction is to be obtained, the selectivity curve must be symmetrical and the discriminator must be accurately centered. More specifically, the overall curve through the discriminator must be symmetrical. The best noise reduction cannot be obtained if the selectivity curve is chair shaped, even if a limiter is used.

2. This symmetry must be maintained for all possible signal strengths. If automatic volume control is employed, consideration must be given to the fact that the input capacitance of the tubes varies with bias. A large enough tuning condenser must be used on the grid circuits to make the change in tuning negligible. If automatic volume control is not used, account must be taken of the change in tuning due to grid current. Large tuning condensers should be used for any tuned circuits feeding grids that are apt to draw grid current. The degree to which symmetry has been maintained in the presence of grid current (or limitation of any kind) can best be shown by a series of oscillograms of the type of Figure 2. The desired information cannot be obtained by means of a selectivity curve because any lack of symmetry is masked by the limiter action. The apparent selectivity curve taken point by point might appear symmetrical because of the action of the limiter in flattening the top of the curve. Under the same conditions, impulsive noise might come through at a frequency deviating from the

center of the pass band, the deviation being caused by grid current which detunes one circuit during high amplitude impulses.

From further tests, not shown here, the grid-leak type of limiter seems to be particularly good in avoiding the detuning due to grid current. In fact, it seems to be the most desirable type of limiter tested, in spite of its apparent inability to remove the amplitude modulation. If Figure 4 is re-examined keeping Figure 8 in mind, it can be seen that the grid-leak limiter does remove the low-frequency components from the envelope of the wave. This type limiter may also be used as a source of automatic volume control voltage. It may be used in combination with the plate voltage limiter to obtain still better limiter action.

3. In the output circuits of the discriminator the two sides of the circuit must have the same fidelity characteristics. A good fidelity balance is seldom obtained in practice because the output circuit is usually grounded on one side instead of the center. As a result, there is more capacity across the grounded side than across the high side. The extra capacity can be compensated for, as shown in Figure 1.

4. From the standpoint of noise reduction the limiter has little value, providing the circuit constants are properly balanced. The limiter helps to reduce the noise only when the instantaneous frequency of the carrier is off center. If the signal is accurately tuned in, this means that the improvement takes place only on noise that occurs during a modulation peak. Noise which occurs only during modulation peaks is very likely to be masked by the modulation. However, the desirability of a limiter is affected by other considerations than noise reduction. For example, the limiter does reduce the distortion if the selectivity curve is not flat over the 200-kc frequency band. This may make it desirable to retain the limiter. The grid-leak type of limiter seems to be almost as effective for noise reduction as the plate-voltage limiter in spite of its demonstrated inability to remove the amplitude modulation.

A more accurate idea of the degree of usefulness of a limiter can be obtained by referring to Table I. The upper table indicates the degree of noise reduction obtainable for relatively weak noise pulses. It indicates that the use of a limiter, and the symmetry of the selectivity curve are unimportant, providing the instantaneous frequency of the carrier wave corresponds to the frequency of zero output at the center of the discriminator curve. It also indicates that if the frequency of the carrier wave has a slightly different value, the limiter is of importance, but a symmetrical selectivity curve is still not necessary.

The lower table is for relatively strong noise pulses. It indicates that for this condition the limiter is useful only when the overall selectivity curve has poor symmetry. That portion of the noise pulse which is greatly in excess of the carrier level may be removed about equally well by the limiter or by the symmetry of the selectivity curve.

Table I

Noise Much Less Than Signal Circuit Conditions	Signal Centered **	Signal Slightly Detuned
Limiter In, Symmetry Perfect	Good*	Good
Limiter In, Symmetry Poor	Good	Good
Limiter Out, Symmetry Perfect	Good	Poor
Limiter Out, Symmetry Poor	Good	Poor

Noise Much Stronger Than Signal Circuit Conditions	Signal Centered **	Signal Slightly Detuned
Limiter In, Symmetry Perfect	Good*	Fair
Limiter In, Symmetry Poor	Fair	Fair
Limiter Out, Symmetry Perfect	Good	Fair
Limiter Out, Symmetry Poor	Poor	Poor

* Relative terms apply throughout one table, but not from one table to another.
** Signal centered for zero discriminator output.

In the table it is assumed that the detuning due to grid current is negligible. If suitable precautions are not taken so that grid current does cause detuning, then the use of the limiter may make the noise worse.

Operating tests with commercial frequency-modulation receivers have shown that ignition noise from passing cars is often a limiting factor in reception. It is believed that careful attention to the points outlined will materially reduce noise of this kind.

The author wishes to acknowledge with thanks the help of Messrs. R. A. Felmley, H. L. Daniels and W. P. Bollinger who assisted with the experimental work leading to this paper.

INTERMEDIATE-FREQUENCY VALUES FOR FREQUENCY-MODULATED-WAVE RECEIVERS*†

By

Dudley E. Foster and John A. Rankin

RCA License Laboratory, New York, N. Y.

Summary—The selection of an intermediate frequency for a superheterodyne receiver involves consideration of the signal frequency, the tuning range, the pass-band width, the minimizing of spurious responses, regeneration stability and frequency stability. The consideration of the frequency-modulation broadcast band of 42 to 50 megacycles illustrates the manner in which these several factors affect the choice of intermediate frequency.

Likelihood of Spurious Responses in Frequency-Modulation Receivers

FREQUENCY-MODULATION receivers are subject to spurious responses as are amplitude-modulation receivers, and because of the probable frequency-modulation-transmitter locations, the spurious responses are likely to be worse than on the present broadcast band. The superheterodyne type of receiver is used almost universally today because of its great advantage in sensitivity and selectivity, but it is subject to spurious responses. The selectivity requirements in the 42- to 50-megacycle frequency-modulation band are so severe that the use of the superheterodyne is even more necessary there than in the standard broadcast band.

Stations may be assigned to alternate channels in the frequency-modulation band in a given locality. Since the channels are 200 kilocycles wide this permits assignments 400 kilocycles apart. At the mean frequency-modulation band frequency of 46 megacycles a separation of 400 kilocycles is the same per cent separation as 10 kilocycles at 1150 kilocycles so that selectivity becomes a prime consideration.

Since the service area of a frequency-modulation transmitter is essentially limited by the horizon, transmitter antennas will have as great a height as possible. In large cities this means that the tendency will be to place the transmitter antenna on the tallest building available. Since tall buildings are generally in the center of the city, it is likely

* Decimal classification: R630.1 × R361.111.
† Reprinted from *Proc. I.R.E.*, October, 1941.

that many frequency-modulation transmitters will be located in close proximity to each other near the center of population of the area. With transmitters so located, high field intensities will occur at many receiving locations, a condition which favors generation of spurious responses.

Types of Spurious Responses

Spurious responses may be of several types, some of which are unique to superheterodyne receivers, others exist in any type of receiver. The types of spurious responses which may exist in frequency-modulation receivers are 1. image response, 2. direct intermediate-frequency response, 3. response from two stations separated by the intermediate frequency, 4. combination of signal and oscillator harmonics, 5. half-intermediate-frequency image response, 6. cross modulation, and 7. intermediate-frequency harmonic response.

Cross modulation, whether within the receiver or of the external type, does not depend upon the type of receiver. The other forms of spurious responses occur only in superheterodynes. Therefore the choice of intermediate frequency does not influence this category of spurious response.

Image Response

Image response in frequency-modulation receivers is similar in nature to that in amplitude-modulation receivers. Since the band extent is only 8 megacycles (42- to 43-megacycle educational and 43- to 50-megacycle commercial broadcasting) and intermediate-frequency over 4 megacycles will prevent image response from frequency-modulation stations. Consideration must also be given to image response from other radio services, but except for television transmitters and possibly amateur and police transmitters, the location will make image response unlikely. Furthermore, the greater frequency separation of such transmissions will aid the radio-frequency attenuation of the image-frequency response.

Direct Intermediate-Frequency Response

Transmission of signals having the same frequency as the intermediate frequency through the radio-frequency system is not believed to be a serious factor in frequency-modulation receivers. Virtually any intermediate frequency chosen will be so far different from the radio-frequency tune frequencies, that attenuation at the first-detector input will be considerably better than in the case of the 455 kilocycles on the standard broadcast band. Nevertheless, all other things being equal,

choice of a frequency to minimize direct intermediate-frequency response is desirable.

Response from Two Stations Separated by the Intermediate Frequency

In this type of response one of the signals acts at the first detector as the local oscillator for the other signal, the detector output being the intermediate frequency. This type of response is likely to be serious on frequency modulation when a low intermediate-frequency value is used, because stations may be allocated every 400 kilocycles in a given territory, and the per cent frequency separation is small with respect to the signal frequency. This type of response is particularly troublesome, because in localities where two strong signals exist separated by the intermediate frequency, they will be heard throughout the tuning range of the receiver.

Combination of Signal and Oscillator Harmonics

Most oscillators have appreciable harmonic content, so that if conditions exist under which signal harmonics occur in the first detector, spurious responses will result. The most common condition for signal-harmonic production is when the signal exceeds the bias. Many frequency-modulation receivers have not used automatic volume control, depending upon the limiter to maintain a uniform signal at the second detector. Under such conditions, particularly with a radio-frequency amplifier present, relatively low-intensity signals will generate harmonics. Even with automatic volume control, harmonics may be generated by an undesired signal of high intensity when the receiver is tuned to a weaker desired signal. Whenever the difference frequency between the harmonics of signal and oscillator is equal to the intermediate frequency, a spurious response will occur.

Half Intermediate-Frequency Image

This type of response in which the interfering signal differs from the oscillator by half the intermediate-frequency value, depends upon production of the second harmonic of this difference frequency in the converter plate circuit. The second-harmonic generation in the converter is usually sufficiently small so that this type of response is not serious despite the fact that the radio-frequency separation of interfering and desired signals necessary for its production is small.

Intermediate-Frequency Harmonic Response

Response at frequencies which are harmonics of the intermediate frequency may be generated in the converter, but, unless an unusually high intermediate-frequency value is used, only high harmonic orders can fall within the frequency-modulation band. As the amplitude of the harmonics decreases rapidly with harmonic order, little difficulty is to be expected in practice from this type response.

Influence of Intermediate Frequency on Spurious Responses

The value of intermediate frequency chosen has a large influence on the number of spurious responses produced. It is evident that, with a frequency range of 8 megacycles, an intermediate frequency higher than 4 megacycles will eliminate image response due to frequency-modulation stations. Similarly an intermediate frequency higher than 8 megacycles will eliminate responses from two stations separated by the intermediate frequency and an intermediate frequency higher than 16 megacycles will eliminate the half-intermediate-frequency image.

The effect of intermediate frequency on combinations of signal and oscillator harmonics is not as obvious, and detailed analysis is required to determine them.

Any combination of signal and oscillator harmonics which results in the chosen intermediate frequency will produce a spurious response. It is assumed that the receiver can be tuned to any frequency between 42 and 50 megacycles. Also only signals between 42 and 50 megacycles are considered. The analysis of the number of spurious responses produced is made without regard to the radio-frequency selectivity. That is, for any signal between 42 and 50 megacycles any tune point in the same range may be chosen. It is, therefore, necessary to consider every possible tune frequency for every signal frequency in the range.

In determining the number of spurious responses, consideration of the relative magnitudes is reserved for later discussion. The lowest frequency channel is 42.1 megacycles and the highest frequency 49.9 megacycles, a total of 40 in all. It is assumed that the total number of spurious responses from any cause is additive. That is, if for any given tune frequency there are three different signal frequencies which can cause a spurious response, three spurious responses are counted for that tune frequency. Thus, for some values of intermediate frequency, while there are 40 frequency-modulation channels, the number of spurious responses is appreciably higher than 40. This form of analysis is more illuminating than indicating only the actual channels on which interference can occur without regard to the frequencies

causing the interference. This is so because spurious responses will be caused mainly by strong signals, and a condition which will permit interference from any one of three undesired signals will be three times as liable to interference as one where there is possibility of spurious rsponse from only one undesired signal. In the latter case the undesired signal might be too weak to cause a spurious response, whereas in the former case, if any one of the three undesired signals is strong enough a spurious response will occur. It is further assumed that tuning is continuous between 42 and 50 megacycles, but that signals are confined to definite channels.

Let $S =$ signal frequency

$O =$ oscillator frequency

$I =$ intermediate frequency

$T =$ tune frequency

$M =$ signal-harmonic order

$N =$ oscillator-harmonic order

Then, for spurious response to occur,

$$MS \pm I = NO$$

$$T = O \pm I$$

$$MS \pm I = NT \pm NI$$

$$\frac{M}{N} S \pm I \left(1 \pm \frac{1}{N}\right) = T$$

$$\frac{M}{N} S + I \left(1 \pm \frac{1}{N}\right) = T$$

if the oscillator is lower than the tune frequency.

$$\frac{M}{N} S - I \left(1 \pm \frac{1}{N}\right) = T$$

if the oscillator is higher than the tune frequency.

The method of analysis consists in assuming values of signal and oscillator harmonics, that is, M and N, and then for each value of intermediate frequency determining the number of signal-frequency channels between 42 and 50 megacycles which will satisfy the expres-

sion for all tune frequencies between 42 and 50 megacycles. It is apparent that consideration should be given to the oscillator higher than the tune frequency separately from the condition for oscillator lower than the tune frequency, since the oscillator cannot change from one side to the other during the tuning process.

To illustrate, suppose a receiver has an intermediate frequency of 4.0 megacycles and is tuned to 49 megacycles; the oscillator is then at 45 megacycles and its second harmonic is 90 megacycles. If the second harmonic of the signal is at either 86 or 94 megacycles, the resultant intermediate frequency is 4.0 megacycles and a spurious response occurs. That is, the fundamental of the signal may be either 43 or 47 megacycles.

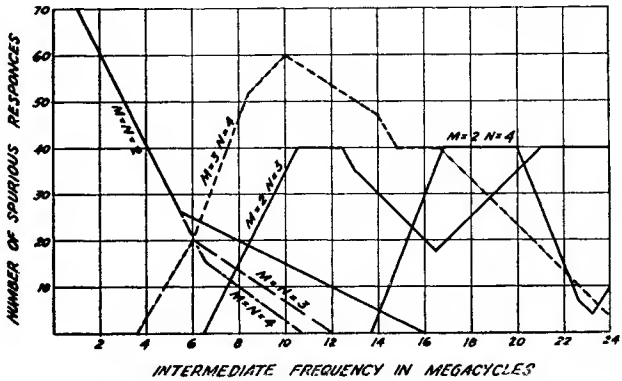


Fig. 1—Spurious responses caused by harmonic combinations of signal and oscillator versus intermediate frequency.
Signals: from 40 channels between 42 and 50 megacycles.
Tuning: from 42 to 50 megacycles.
M = signal-harmonic order.
N = oscillator-harmonic order.
Oscillator lower than tune frequency.

From an examination of the above expressions it may be seen that variation of spurious responses with intermediate frequency is a linear function, so that a plot of the variation of spurious responses with intermediate frequency will consist of a series of straight lines.

A plot of the number of spurious responses as a function of intermediate frequency for the case where the oscillator frequency is lower than the tune frequency is shown in Figure 1 and for the case where the oscillator frequency is higher than the tune frequency in Figure 2.

An examination of the expressions for spurious responses as a function of intermediate frequency, along with the plots of Figures 1 and 2, reveals that for the same harmonic order of signal and oscillator frequency, the number of spurious responses as a function of inter-

mediate frequency is the same for the oscillator frequency higher or lower than the tune frequency. Further, the spurious responses due to second harmonics of signal and oscillator go to zero at an intermediate frequency of 16 megacycles and do not recur at any higher intermediate frequency. The spurious responses due to the third harmonic of signal and oscillator drop out at an intermediate frequency of 12 megacycles, while the spurious responses due to the fourth harmonic of signal and oscillator drop out at 10.6 megacycles.

When different harmonic orders of signal and oscillator are involved, the number of spurious responses is different when the oscillator frequency is higher than the tune frequency than that for the oscillator frequency lower than the tune frequency.

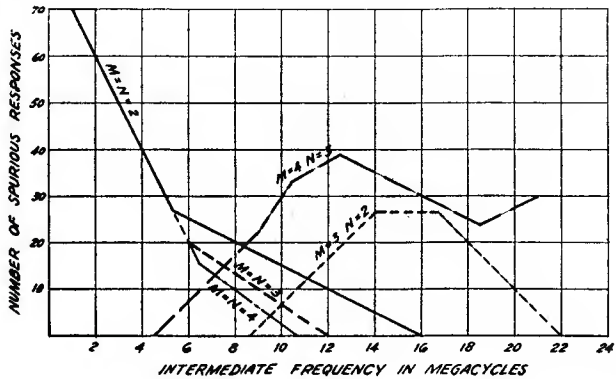


Fig. 2—Spurious responses caused by harmonic combinations of signal and oscillator versus intermediate frequency.
Signals: from 40 channels between 42 and 50 megacycles.
Tuning: from 42 to 50 megacycles.
M = signal-harmonic order.
N = oscillator-harmonic order.
Oscillator higher than tune frequency.

When the oscillator frequency is lower than the tune frequency, as shown in Figure 1, no spurious responses occur for any value of intermediate frequency if the oscillator-harmonic order is lower than the signal-harmonic order. When the oscillator frequency is higher than the tune frequency, as shown in Figure 2, no spurious responses occur for any value of intermediate frequency if the oscillator-harmonic order is higher than the signal-harmonic order.

A plot of the number of spurious responses due to the image and to two stations separated by the intermediate frequency is shown in Figure 3.

Note that these curves consider stations on each channel as having the ability to produce image responses, responses from two stations

separated by the intermediate frequency, and spurious responses caused by oscillator and signal harmonics. In actual practice, stations will not be so allocated in any given location, but the curves are valid in determining the best intermediate frequency to be used to reduce the number of responses, regardless of how many stations are assigned in any particular location, provided of course the stations are uniformly located in the band.

The relative severity of spurious responses is difficult to evaluate exactly without data as to the magnitude of signal-harmonic generation in the converter, which in turn depends upon the signal amplitude. However, an estimate may be made of the effect of oscillator harmonics and instructive deductions made therefrom.

In a typical converter of the pentagrid type, the amplitude of oscillator harmonic has been found to be approximately proportional to

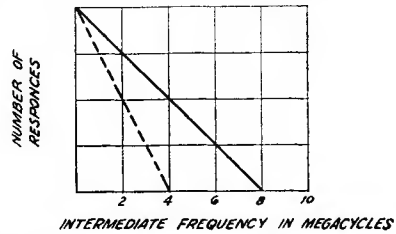


Fig. 3—Interference caused by the image response and by two stations separated by the intermediate frequency versus intermediate frequency.
Signals: from 40 channels between 42 and 50 megacycles.
Tuning: from 42 to 50 megacycles.
Oscillator either higher or lower than tune frequency.
..... image.
——— two stations separated by the intermediate frequency.

$1/N^2$, where N is the order of the harmonic. The conversion gain is not constant with respect to oscillator amplitude but reaches a maximum, usually a broad maximum (the conversion gain is substantially constant over an appreciable oscillator range) at some value of oscillator voltage. For values of oscillator voltage below this maximum the conversion is, to a first approximation, proportional to the oscillator voltage. Since the oscillator amplitude is $1/N^2$ and the conversion proportional to amplitude, the conversion is approximately proportional to $1/N^2$ if the converter is operated so that the fundamental does not exceed the optimum. If the fundamental somewhat exceeds the optimum conversion point, there will be no appreciable decrease in conversion from the optimum value. But under these conditions the second harmonic may produce as much conversion gain as the fundamental. Consequently to minimize harmonic generation it is desirable to limit the oscillator injection to a value only sufficient to develop maximum

conversion. With the oscillator injection limited to such value, the probability of a given signal causing a spurious response is inversely proportional to the square of the oscillator-harmonic order involved. This factor may be applied to the calculations for signal- and oscillator-harmonic combinations.

Considering the image response; this is due to the fundamental of the signal and the oscillator so should not be decreased by any factor as are the harmonic combinations.

In applying weighting to the case of two signals separated by the intermediate frequency, the oscillator is not involved but the strength of the spurious response is proportional to the strength of the signals involved. This response will in general be less than that for the image. Accordingly a factor of 0.5 was applied to the two signals separated by the intermediate frequency, giving it a weight between the image and harmonic combinations due to the second harmonic of the oscillator.

When weighting factors are applied we can no longer say that the ordinates represent the number of spurious response signals. The ordinates in this case represent the probability of spurious responses for various intermediate-frequency values in a receiver having no selectivity ahead of the converter, with the probability for an intermediate frequency of 1 megacycle considered to be 100. The variation of likelihood of interference is shown in Figure 4. This figure combines the weighted effect of image, two signals separated by the intermediate frequency, and oscillator-signal harmonic combinations up to the fourth of both signal and oscillator.

It illustrates that the probability of interference decreases rapidly with increasing intermediate frequency at first and then more slowly, and that over the entire range of intermediate frequency there is little choice between oscillator lower and oscillator higher conditions. It does not indicate definitely any best intermediate frequency but does show that a high intermediate frequency is better. While, as has been stated, direct intermediate-frequency interference is not believed to be a major factor, and as the image ratio for any intermediate frequency over 8 megacycles should be adequately high for any receiver having more than one tuned circuit ahead of the converter, nevertheless it is well to minimize spurious response from these sources where no other disadvantage accrues therefrom.

In considering spurious responses, one characteristic of frequency modulation, not heretofore mentioned, should be borne in mind, namely, that such responses when they occur on the frequency of a desired signal do not cause an audible whistle as they do on amplitude modulation but appear essentially as cross talk.

The types of signals, other than frequency-modulation stations, likely to cause interference are those where the transmitter may be located in a populous district. These are amateurs, television, and police signals mainly and possibly international broadcasting.

If image response from other frequency-modulation stations is to be eliminated an intermediate frequency of at least 4 megacycles has been seen to be necessary. To minimize direct intermediate-frequency interference, a frequency range of some 200 kilocycles free from types

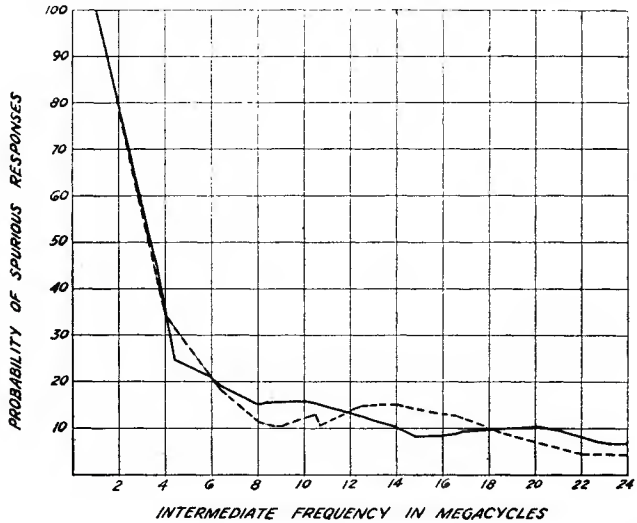


Fig. 4—Probability of spurious responses due to image, two signals separated by the intermediate-frequency and harmonic combinations of signal and oscillator for different intermediate frequencies.
Image weighting factor = 1.0.
Two stations separated by the intermediate-frequency weighting factor = 0.5.
Spurious responses due to *Nth* harmonic of oscillator weighting factor = $1/N^2$.
Probability of spurious responses for 1.0-megacycle intermediate-frequency taken as 100.
——— oscillator lower.
..... oscillator higher.

of services likely to cause such interference is required. The lowest such range above 4 megacycles is at 4.3 megacycles to which are allocated government, general communication, and coastal-harbor transmitters. Another intermediate frequency in that general range which appears relatively free from direct intermediate-frequency interference is at 5.38 megacycles, where are allocated government, fixed, and general communication channels.

A frequency of 8.25 megacycles has been widely used as the inter-

mediate frequency for the sound channel of television receivers, which suggests its use for frequency modulation in order to minimize component types. Examination of frequency allocations shows that 8.26 megacycles is somewhat better than 8.25 megacycles over a 200-kilocycle width. And that to 8.26 megacycles are allocated ship-telegraph and government services, neither likely to cause interference. On the score of image, however, with the oscillator higher, both the 5-meter amateur and the second television bands fall within the image range. With the oscillator lower, we find 10-meter amateur, police, and government frequencies. It would appear therefore that this frequency, particularly with the oscillator lower, will be satisfactory in receivers having enough selectivity preceding the converter to insure good image ratio.

In receivers in the low-price classes, where selectivity ahead of the converter is not high, it would seem desirable to operate with the oscillator lower than the tune frequency to avoid the television channels, and to use an intermediate frequency that would not permit interference from the amateur bands. This requires an intermediate frequency of over 11 megacycles. At 11.45 megacycles we find government, fixed, and aviation allocations, and for the image, fixed, government, and broadcast stations, so that this frequency is comparatively free from likelihood of spurious responses.

It is not likely that frequencies appreciably higher than about 14 or 15 megacycles will be useful, because of decreasing stability and gain limitation due to tube and circuit capacitances.

Influence of Intermediate Frequency on Selectivity and Stability

The value of intermediate frequency does not affect the selectivity if the Q of the intermediate-frequency circuits is varied proportional to the intermediate frequency, thus for the same band width, the Q for 8.26 megacycles should be 1.9 times as great as the Q for 4.3 megacycles intermediate frequency. The necessary Q values for 200-kilocycle channels are readily obtainable for any practical intermediate frequency so selectivity considerations do not influence the choice of an intermediate-frequency value.

Considerations of both frequency stability and regeneration stability are involved in the choice of an intermediate frequency. Both are poorer with a high intermediate frequency than with a low one. In an amplifier of N stages the over-all stable gain with regard to regeneration varies inversely as the $N/2$ power of the intermediate frequency. It is more difficult to analyze frequency stability quantitatively because of the several diverse design factors involved but in general

the frequency drift is proportional to the intermediate frequency. This influence of intermediate-frequency value on frequency and regeneration stability is the principal difficulty involved in the use of a high intermediate frequency and must be weighted against the advantage of high intermediate frequency in reducing spurious responses. Design expedients, such as impedance distribution or neutralization may be used to minimize the regeneration stability influence and expedients such as cathode degeneration used to minimize frequency-stability problems.

Conclusions

It has been shown that the use of an intermediate-frequency value higher than 4 megacycles materially reduces the probability of spurious responses, but at the same time may entail some sacrifice in gain and frequency stability.

It would appear therefore, that for receivers designed to operate on weak signals, where spurious responses are unlikely, a relatively low intermediate frequency of 4.3 megacycles or 5.38 megacycles is preferable. However for a considerable part of the service area, where signal intensities are high and spurious responses likely, high sensitivity is not required and an intermediate frequency of 8.26 megacycles or even 11.45 megacycles has much to recommend it. Where a receiver is designed to operate over a wide range of signal intensities, it may be preferable to use a relatively high intermediate-frequency value to minimize spurious responses, using impedance distribution or neutralization principles, and if necessary an additional stage to obtain the required high gain with stability.

A FREQUENCY-DIVIDING LOCKED-IN OSCILLATOR FREQUENCY-MODULATION RECEIVER*†

By

G. L. Beers

Assistant Director of Engineering in Charge of Advanced Development,
RCA Victor Division, Camden, N. J.

Summary—*A new type of frequency-modulation receiving system is described in which a continuously operating local oscillator is frequency-modulated by the received signal. In an embodiment of the system which is described, the oscillator is locked in with the received signal at one fifth the intermediate frequency. With this 5:1 relationship between the intermediate frequency and the oscillator frequency, an equivalent reduction in the frequency variations of the local oscillator is obtained. Received signal-frequency variations of ±75 kilocycles are reproduced as ±15-kilocycle variations in the oscillator frequency. The frequency-modulated signal derived from the oscillator is applied to a discriminator which is designed for this reduced range of frequencies.*

The oscillator is designed to lock in only with frequency variations which occur within the desired-signal channel. The oscillator is, therefore, prevented from following the frequency variations of a signal on an adjacent channel. A substantial improvement in selectivity is thus obtained.

The voltage required to lock in the oscillator with a weak signal is approximately one twentieth of the voltage applied to the discriminator. Since this voltage gain is obtained at a different and lower frequency than the intermediate frequency, the stability of the receiver from the standpoint of over-all feedback is materially improved.

Other performance advantages and the factors affecting the operation of the system are discussed.

FREQUENCY-MODULATION broadcasting is still in its infancy in terms of a nation-wide entertainment service. Until a large number of high-powered frequency-modulation broadcast stations are operating on a commercial basis, the major technical problems which are involved in the design of frequency-modulation receivers will not be fully appreciated. However, the experience which has already been gained from frequency-modulation broadcasting has indicated some of the problems which must be given serious consideration.

Probably the most difficult requirement to be met is that of obtaining adequate adjacent-channel selectivity. This problem was emphasized by a report on "Blanketing of High-Frequency Broadcast Stations" issued in 1941 by the Federal Communications Commission. High sensitivity is necessary in a frequency-modulation receiver to

* Decimal classification: R361.111.
† Reprinted from *Proc. I.R.E.*, December, 1944.

insure maximum performance. This requirement makes it difficult to provide the desired over-all stability without excessive shielding and other circuit complications. This problem has already been the subject of a great deal of engineering investigation and one of the solutions which has been proposed is the use of the double heterodyne type of superheterodyne receiver. A new approach to a solution of these problems is provided by a frequency-dividing locked-in oscillator frequency-modulation receiving system which has been developed.[1] It is the purpose of this paper to describe the new receiving system and to indicate some of the factors which affect its operation.

Description of System

Basically the operation of the system depends on producing, in the receiver, a local signal which is frequency-modulated by the received signal. The local signal is provided by a continuously operating oscillator. The received signal, after it has been amplified by conventional radio-frequency and intermediate-frequency amplifiers, is applied to the oscillator in such a way as to cause its frequency to change in accordance with the frequency variations of the received signal. In the particular applications of the system to be described in this paper, the oscillator is locked in with the received signal at one fifth the intermediate frequency. With this 5:1 relationship between the intermediate frequency and the oscillator frequency an equivalent reduction in the frequency variations of the local oscillator is obtained. Received-signal frequency variations of ± 75 kilocycles are reproduced as ± 15-kilocycle variations in the oscillator frequency. It should be noted that the locked-in oscillator operating at one fifth the intermediate frequency reduces the frequency deviation corresponding to any modulation frequency but does not change the modulation frequency. The frequency-modulated signal derived from the oscillator is applied to a discriminator which is designed for this reduced range of frequencies.

The output voltage of the oscillator is independent of the strength of a received signal, in fact, the same voltage is applied to the discriminator when no signal is being received as when the receiver is tuned to a near-by transmitter. This feature makes it unnecessary to employ the conventional arrangements for minimizing amplitude variations in the received signal.

The adjacent-channel selectivity of a conventional frequency-modulation receiver is determined by the selectivity characteristics of the radio-frequency and intermediate-frequency circuits. If these circuits

[1] United States Patent No. 2,356,201, filed February 12, 1942.

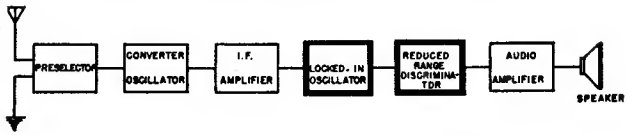


Fig. 1—Block diagram.

do not provide sufficient selectivity, a local transmitter on a channel adjacent to the desired signal may produce, at the discriminator, a substantially greater voltage than is obtained from the desired station. Under these conditions the desired program will not be heard. In the new receiving system a novel principle is used to provide additional adjacent-channel selectivity. The oscillator is designed to "lock-in" only with frequency variations which occur within the desired-signal channel. The oscillator is therefore prevented from following the frequency variations of a signal on an adjacent channel. A substantial improvement in selectivity is thus obtained by electronic means.

The "locked-in" oscillator arrangement which is used provides, under weak signal conditions, a voltage step up of approximately 20. In other words, the voltage required to lock in the oscillator with a weak signal is approximately one twentieth of the voltage applied to the discriminator. Since this voltage gain is obtained at a lower frequency than the intermediate-frequency, the stability of the receiver from the standpoint of over-all feedback is materially improved. This improvement is secured without the disadvantage of the additional image responses which are obtained with the double-heterodyne type of superheterodyne receiver.

One receiver arrangement is shown in Figure 1. In this diagram the units which are heavily outlined are those which are peculiar to the new system.

Description of the Locked-In Oscillator

The locked-in oscillator circuit diagram is shown in Figure 2. The

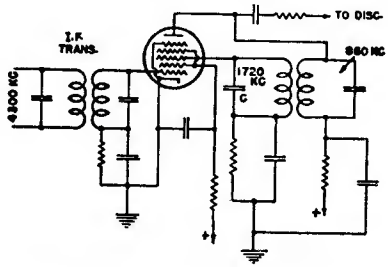


Fig. 2—The locked-in oscillator.

tube generally used in this circuit has been an A-5581, an experimental converter tube, which is similar to the 6SA7 but has a higher mutual conductance. The oscillator tuned circuit is connected to the plate of the tube and the feedback coil is connected to the No. 3 grid. This grid is operated with self-bias. The received signal is applied to the No. 1 grid of the tube through a 4300-kilocycle intermediate-frequency transformer. The No. 1 grid is likewise operated with self-bias.

Description of Discriminator

One type of discriminator that can be used with the locked-in oscillator is shown in Figure 3. This circuit has a pair of diodes connected with their load resistors in opposition so the discriminator is balanced at the center frequency. One diode has a tuned circuit in series with it and the other has a tuned circuit across it. The discriminator is connected across the tank circuit of the locked-in oscillator

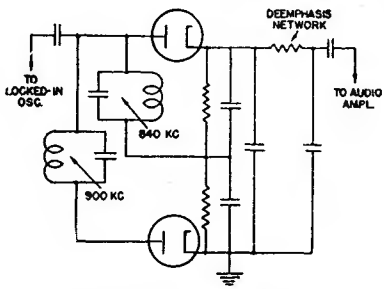


Fig. 3—Discriminator.

through the coupling capacitor shown in the diagram. The audio-frequency output from the discriminator is fed through a de-emphasis network to the audio amplifier.

Why the Oscillator Locks In with the Received Signal

Theoretical and experimental evidence indicates that the locked-in oscillator circuit operates in accordance with the following theory.

As previously stated, the oscillator is designed to lock in at one fifth the intermediate frequency. With an intermediate frequency of 4300 kilocycles the oscillator tank circuit is tuned to 860 kilocycles. When no signal is being received the tube will function as a normal oscillator. The amplitude of the oscillation in a feedback oscillator is determined by the curvature of the $E_g - I_p$ characteristic and is usually so great that the grid voltage swings well into the curved parts of the tube characteristic during the cycle. This means that a distorted output

current is produced in the plate circuit, having component frequencies 2ω, 3ω, 4ω, $\cdots$ where ω is the natural frequency of the tuned plate circuit. These harmonics are applied to the No. 3 grid because of the regenerative coupling. Furthermore, the No. 3 grid operates with self-bias and draws grid current during the positive swings of voltage. The grid-current pulses also contain the harmonics of ω.

Suppose now that the signal voltage of frequency 5ω (4300 kilocycles) is applied to the No. 1 grid. Since the tube is a nonlinear device and operates as a converter, combination frequencies will be produced equal to $\pm 5r\omega \pm s\omega$ where $r, s = 0, 1, 2, 3, \cdots$. Since the plate circuit is tuned to a frequency ω (860 kilocycles), the only frequencies which will be amplified are those of frequency ω; the others will be by-passed effectively. If $r = 1$, then $s = 4$ or 6 will give the frequency ω. This means that either the fourth or the sixth harmonics of the oscillator will beat with the incoming signal, having the frequency 5ω, to give the frequency ω.

This added 860-kilocycle component of the plate current caused by

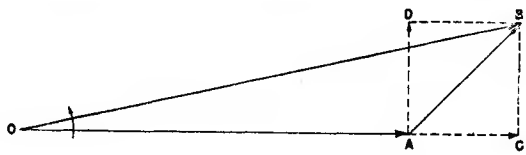


Fig. 4—Vector diagram.

the harmonics of the oscillator beating with the incoming signal is in phase with the 860-kilocycle current in the oscillating plate circuit. The circuit becomes stable in this condition and the injected current will "lock in" the incoming 4300-kilocycle signal with the 860-kilocycle current in the plate circuit. Since the injected current has the same phase and frequency as the normal current, it is merely equivalent to an increased output from the tube.

Now suppose that the frequency of the incoming signal is increased somewhat. The effect of the fourth harmonic will be to inject a current of slightly greater frequency than 860 kilocycles into the tank circuit. The sixth harmonic will also cause an injected current of slightly less than 860 kilocycles; this will be considered later. Assume for the moment that the oscillator is *not* locked in. In Figure 4, OA is a vector rotating 860,000 times per second and represents the normal current in the oscillating tank circuit. Let AB be the injected current of frequency slightly greater than 860 kilocycles, from the fourth harmonic of the oscillator voltage beating with the incoming signal voltage. This vector will rotate slightly faster than 860,000 times per second

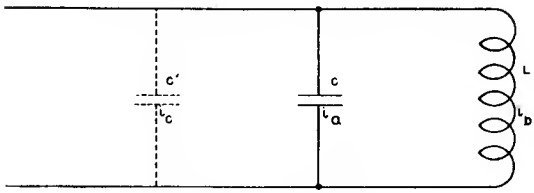


Fig. 5—Effect on circuit.

and thus will have an angular velocity relative to *OA* equal to the difference of the two angular velocities.

Now consider the instantaneous condition shown in Figure 4. The injected current *AB* has a component *AC* in phase with *OA* and another component *AD*, 90 degrees out of phase with respect to *OA*. Let this resultant current *OB* be applied to a tuned circuit *LC* as shown in Figure 5. Since the *LC* circuit is tuned to 860 kilocycles, it will be at resonance with respect to the current *OC* which is also 860 kilocycles and equals $i_a + i_b$. The quadrature current *AD* is a leading current at the instant shown by Figure 4, and the result is the same as though an additional condenser *C'* is in the circuit. The effect is to *decrease* the natural frequency of the tuned circuit.

Now consider the condition at a later instant as shown by Figure 6. Since the vector *AB* is rotating with respect to *OA* it has now rotated to the new position as shown. The injected current *AB* now has an in-phase component *AC* as before, but the component *AD* is now lagging instead of leading. If this current *OB* is now impressed on the circuit of Figure 5, the lagging component *AD* will cancel part of the leading current through *C* and this will be equivalent to reducing the capacitance *C* since the circuit is now drawing a smaller leading current. This will raise the resonant frequency of the tuned circuit.

It is now evident that the circuit of Figure 2 behaves like a reactance tube and swings the frequency of the tuned circuit back and forth. It is easy to see that if the frequency of the incoming signal is approximately five times that of the tuned circuit, a point will be reached when the frequency of the tuned circuit becomes exactly one

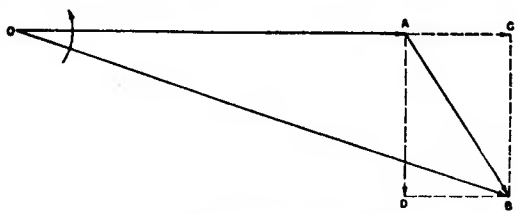


Fig. 6—Vector diagram.

fifth of the incoming signal frequency. When this happens the oscillator will "lock in" with the incoming signal. This means that the amplitude and phase of the plate current now remain fixed with respect to the incoming signal; vector *AB* now makes a constant angle *CAB* with *OA*.

If the incoming signal is exactly five times the frequency of the tuned-plate circuit, the vector *AB* will be in phase with *OA*. As the incoming signal frequency is decreased, the vector *AB* rotates to some position such as that shown in Figure 4. A further decrease in frequency will rotate the vector until it is 90 degrees out of phase with respect to *OA*. Since this position gives the maximum amount of quadrature current it corresponds to the maximum amount the oscillator frequency can be pulled over, and thus gives the lower limit of the lock-in range.

If the incoming-signal frequency becomes greater than five times the plate-circuit frequency, the conditions will be similar except that the vector *AB* will be lagging as shown by Figure 6 instead of leading. The upper limit of the lock-in range is reached when the injected current lags by 90 degrees. The lagging current tends to reduce the effective capacitance of the circuit and thus raises the frequency.

When the sixth and fourth harmonics are both present simultaneously, it can be shown that the result is a single injected current of variable amplitude and phase. This causes the frequency of the tuned circuit to swing back and forth in accordance with these variations; the process is very similar to that already explained when the fourth harmonic only is present. Usually, the fourth and sixth harmonics will be of unequal amplitude and the effect of the weaker one is to produce relatively small variations in the other.

Lock-In Range Requirements

As previously stated by restricting the lock-in range of the oscillator to frequency variations in the desired channel a material improvement in selectivity can be obtained. On the other hand, it is necessary that the lock-in range be adequate to follow the frequency variations of the received signal and in addition provide for receiver mistuning and frequency drift in the transmitter and receiver.

The effect of the fourth and sixth harmonics in controlling the lock-in range of the oscillator has been previously discussed. The amount of fourth and sixth harmonics on the No. 3 grid of the oscillator is limited, and this limits the lock-in range. When the deviation exceeds the lock-in range the oscillator breaks out and starts back toward the center frequency since it is no longer controlled. The

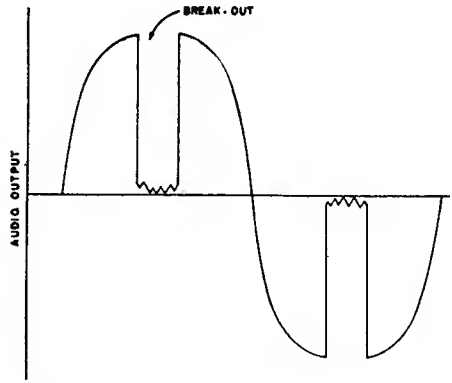


Fig. 7—Breakout characteristic.

oscillator may then suddenly jump to a series of different frequency ratios such as, · · · 36/7, 41/8, 46/9 · · · 5/1, · · · , 44/9, 39/8, 34/7, · · · for short intervals. The lock-in range for each of those ratios is very small, and the oscillator breaks out between them. The result can be a distorted output as shown by Figures 7 or 8. It is, therefore, necessary to provide adequate lock-in range in order to prevent this distortion.

Factors Which Determine The Lock-in Range

The lock-in range of the oscillator depends upon several factors which will now be discussed.

Effect of Discriminator

When a discriminator is connected to the oscillator, it changes the impedance relations of the tank circuit and increases the lock-in range.

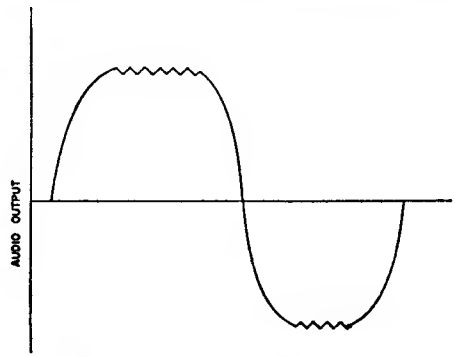


Fig. 8—Breakout characteristic.

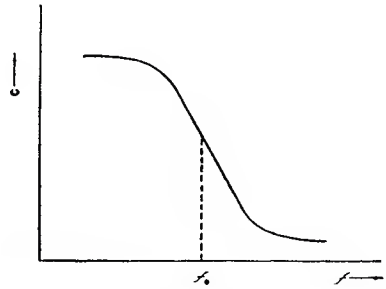


Fig. 9—Input capacitance of discriminator.

The equivalent input capacitance of the discriminator circuit shown in Figure 3 decreases rapidly with frequency near the center frequency of the oscillator. Figure 9 shows how this capacitance falls off near the center frequency f_0. If the oscillator tank circuit is to be kept in tune over the operating frequency range, the tank circuit capacitance should decrease with increasing frequency as shown by Figure 10. The slope of this curve is determined by the L/C ratio of the tank circuit.

The discriminator input capacitance characteristic can be designed to provide an apparent capacitance change with frequency nearly to match the requirements for tuning the oscillator.

In Figure 11 the solid line represents the falling input capacitance of the discriminator and the dashed line is the variation of capacitance required to keep the oscillator in tune as the frequency is varied. If the two curves have approximately the same slope at the center frequency f_0, the lock-in range will be greatly increased since only a small amount of reactive current will shift the oscillator frequency a considerable amount.

Effect of Signal Voltage

If the No. 1 grid is operated with self-bias so that the operating

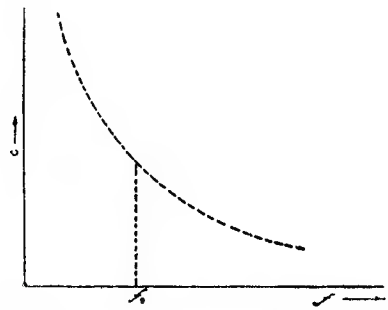


Fig. 10—Oscillator-tuning capacitance.

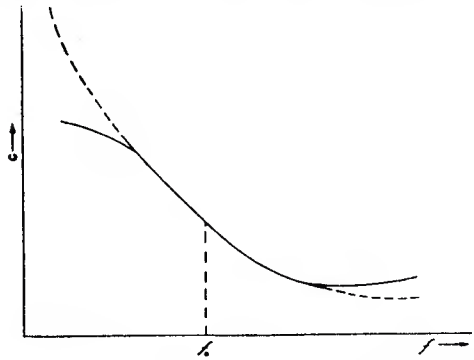


Fig. 11—Matching the discriminator to the oscillator.

bias is approximately equal to the peak amplitude of the applied signal voltage, the lock-in range will be as shown by Figure 12.

For small applied voltages the lock-in range increases rapidly from zero with increasing signal voltage until it reaches a maximum, and it then decreases slowly with further increase in voltage as shown by the dashed line.

In practice, the screen and plate resistors can be chosen to correct this falling off of the lock-in range with increased input. This compensation will give the constant lock-in range beyond the knee of the curve as shown by the solid line.

Effect of Tube Constants

The lock-in range depends upon the amount of quadrature current that can be developed by the tube. This means that the tube should have a fairly high zero-bias plate current and a fairly high mutual conductance from the No. 1 grid to plate. This assures large pulses of plate current which produce the required reactive current. The

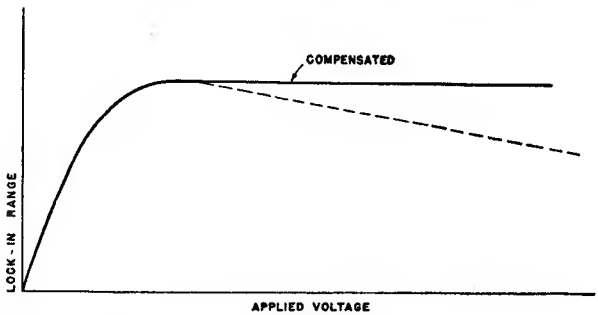


Fig. 12—Effect of signal voltage on lock-in range.

experimental A-5581 tube has been found to meet these requirements. This tube is similar to the 6SA7 but provides increased peak current and increased mutual conductance.

Effect of Intermediate-Frequency Selectivity on Lock-In Range

The primary effect of intermediate-frequency selectivity is to attenuate the voltage on the No. 1 grid as the signal frequency moves down the side of the selectivity curve. Naturally the oscillator cannot lock in if the incoming signal voltage becomes too small. This means that the bandwidth of the intermediate-frequency amplifier will affect the lock-in range. Figure 12 shows the variation of lock-in range with input voltage. The range falls off very rapidly when the applied voltage falls below the knee of the curve. The amplifier should be designed so it is broad enough to assure sufficient voltage to lock in the oscillator at the maximum frequency swings encountered and also to provide for drift and mistuning.

Effect of Oscillator Frequency

The lock-in range is in general inversely proportional to the oscil-

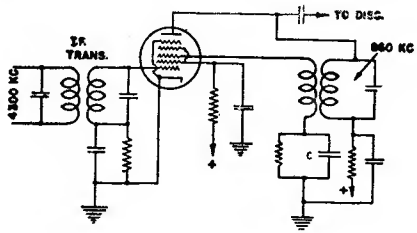


Fig. 13—Tuned-feedback coil.

lator tank circuit *C*. An increase in the intermediate frequency will result in an increase in the lock-in range only when *C* is correspondingly reduced.

Effect of Feedback Winding

The lock-in range will increase somewhat with increased mutual inductance from the tank coil to the feedback winding. Fairly tight coupling should be used for increased range.

A method which can be used to increase the lock-in range is to tune the feedback winding to the second harmonic of the oscillator as shown by Figure 13. Capacitor *C* is chosen to tune the grid circuit to 1720 kilocycles. This builds up the second harmonic, which in turn causes an increase in the fourth and sixth harmonics because of the nonlinearity of the tube. The result is an increase in the lock-in range.

Noise-Reduction Characteristics

It has been previously stated that the lock-in oscillator arrangement can be designed to increase materially the adjacent-channel selectivity of a receiver. This improvement is obtained by restricting the lock-in range of the oscillator so that it will follow only the frequency variations which occur within the desired channel. This restricted lock-in range is of interest also from the standpoint of the noise-reducing properties of the receiver.

In conventional frequency-modulation receivers the discriminator is designed so that the linear portion of its response characteristic is adequate to accommodate the frequency variations of received signals, with due allowance for mistuning both by the user and that resulting from frequency drift of the heterodyne oscillator. The curved portions of the discriminator characteristic which extend beyond the linear region just referred to provide an additional frequency range in which noise components with wide frequency variations are converted into

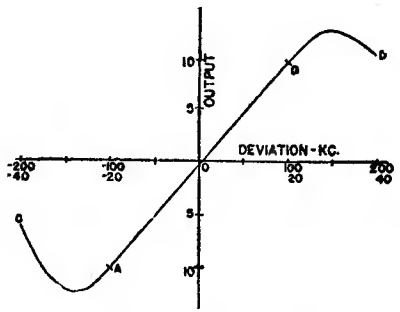


Fig. 14—Discriminator characteristic.

amplitude variations. Figure 14 shows a typical discriminator characteristic in which *AB* is the linear region within which the frequency variations of received signals are converted into amplitude variations. The sections of the characteristic designated *CA* and *BD* are the portions which are not useful in the reception of desired signals because of the curvature, but which are effective in converting frequency-modulation noise components into amplitude variations. The upper figures, indicating deviation, correspond to the discriminator characteristic in a conventional receiver, while the lower figures are the frequency values for the locked-in oscillator discriminator. The restricted frequency range of the oscillator in the locked-in oscillator type of receiver can be used to limit the portion of the discriminator characteristic, which is utilized in converting the frequency variations of both

the received signal and noise components into amplitude variations, to the linear region *AB*.

Another characteristic of the locked-in oscillator which may be used to advantage in minimizing the effects of noise is the ability to prevent the oscillator from following the frequency variations corresponding to superaudible noise components. This is accomplished in the oscillator arrangement shown in Figure 2 by the proper choice of circuit constants.

Modified-Circuit Arrangement

A modification of the frequency-dividing frequency-modulation receiver has been developed by which its ability to select between desired signals or noise is further extended. Figure 15 is a block diagram of this modification.

The locked-in oscillator used in this arrangement is likewise designed to operate at one fifth of the intermediate frequency. The normal lock-in range of the oscillator, however, is restricted to only 20 to 35 per cent of the frequency-variation range required for received signals. This very restricted lock-in range is extended by means of a

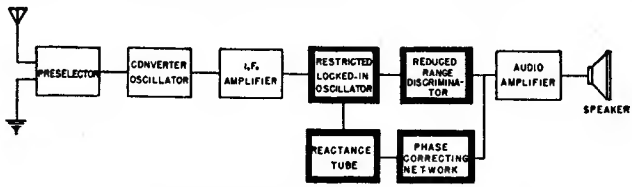


Fig. 15—Modified-circuit arrangement.

reactance-tube arrangement so that the oscillator will follow the maximum frequency variations of received signals. The audio-frequency potential developed at the discriminator-rectifier combination is applied through a phase-correcting network to the reactance tube in the proper phase and magnitude to cause the reactance tube to shift the oscillator resonant frequency so that at any instant its frequency is such that the limited lock-in range will permit it to lock in with the received signal. The amplitude of the control potential applied to the reactance tube is normally kept slightly below the value which would shift the oscillator to the correct frequency, assuming that the oscillator had no lock-in range. In other words, for 100 per cent modulation the reactance tube shifts the oscillator frequency by slightly less than $\pm$ 15 kilocycles.

Let us consider the merits of this arrangement in connection with noise impulses and adjacent-channel selectivity. Superaudible fre-

quency-modulation noise components applied to the input circuit of the locked-in oscillator may appear in the oscillator output circuit. The phase-correcting network, however, may be designed so that these components either are not fed back to the reactance tube at all or are not fed back in such phase and amplitude as to permit the oscillator to follow them. In other words, the receiving system is provided with a circuit which is responsive only to small frequency variations and this restricted-response range is moved back and forth at a rate which follows the desired modulation of received signals but is not moved back and forth at a rate which will follow superaudible noise impulses which may be present with the received signals.

The effect of the reactance-tube arrangement on adjacent-channel selectivity is also of interest. This can best be understood by reference to the discriminator-rectifier—voltage-frequency response characteristic shown in Figure 14. As the output potential of the discriminator-rectifier and hence the potential applied to the reactance tube varies over the useful portion of the discriminator characteristic (the linear portion of the characteristic between the points *A* and *B*) the effect of the reactance tube is to shift the oscillator frequency in the same direction as the frequency changes which give rise to the demodulator potentials. If, on the other hand, we assume that a signal on the adjacent channel could reach the discriminator circuits and produce potentials caused by frequency variations over the side of the discriminator characteristic as indicated by the portion *A-C* of the curve, the phase of the potentials applied to the reactance tube would be such that the effect of the reactance tube on the oscillator would be to reverse the direction of the oscillator-frequency change. That is, the reactance tube cannot shift the oscillator frequency so that it will lock in with the signal on an adjacent channel because the circuit elements are so designed that if the frequency of the oscillator were to change beyond the useful range of the discriminator and toward the adjacent channel, the phase and magnitude of the potential applied to the reactance tube would shift in such a manner that the oscillator frequency would be shifted away from the adjacent channel frequencies.

Experimental Results

As a part of an experimental investigation of the new receiving system, work was carried on with two identical commercial receivers. One was modified by incorporating the locked-in oscillator and reduced-range discriminator, shown in Figure 16, in place of the two-tube cascade limiter and the discriminator used in the original construction. The other receiver was used for comparative tests in the laboratory and

field. This procedure was repeated with two identical laboratory receivers constructed along conventional lines.

It should be noted that the locked-in oscillator circuit shown in Figure 16 is representative of the receiving system illustrated by the block diagram in Figure 1. This arrangement was used in preference to the modification illustrated by the block diagram in Figure 15 because it was less complicated and, therefore, considered more suitable for commercial receivers.

With the arrangement shown in Figure 16 an intermediate-frequency signal of about 1 volt on the No. 1 grid of the oscillator tube was required to provide the desired lock-in range of approximately ± 110 kilocycles. The frequency range in excess of the ± 75 kilocycles required for the normal modulation of a received signal is provided to take care of mistuning by the user, frequency drift of the heterodyne oscillator, and over-modulation at the transmitter. The oscillator voltage developed at the discriminator was between 20 and 30 volts. From

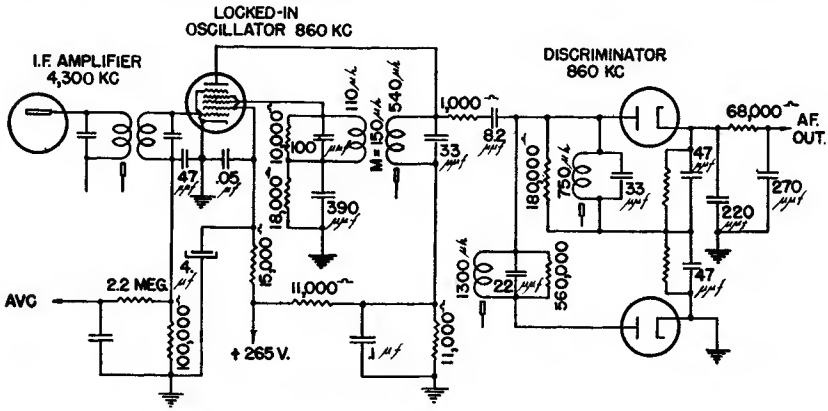


Fig. 16—Locked-in oscillator and discriminator.

the foregoing, it is apparent that the receiver should be sufficiently sensitive to produce 1 volt on the No. 1 grid of the oscillator to provide satisfactory reception of a desired signal.

Improvement in Selectivity

The results of selectivity measurements, made by the two-signal method, are shown in Figure 17. In these tests, the receivers were tuned to a desired signal of 100 microvolts, with 400-cycle modulation and a deviation of ± 25 kilocycles. An interfering signal, modulated with 1000 cycles, and a deviation of ± 25 kilocycles, was adjusted in signal strength and frequency to give an interference output 30 decibels below the 400-cycle output. A considerable improvement in selectivity,

especially for the entire adjacent channel, is shown with the receiver employing the frequency-dividing locked-in oscillator system. It should be noted that with an increase in interfering signal, a point of oscillator breakout may always be reached. The level of interfering signal at which breakout occurs is higher than the —30-decibel interference level. The improvement in adjacent-channel selectivity, shown by these curves, is equivalent to the addition of two intermediate-frequency stages in the receiver.

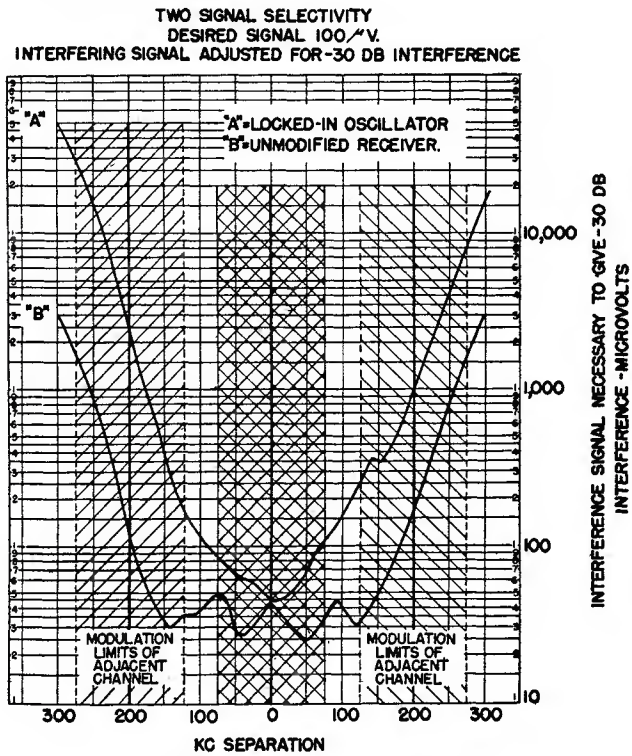


Fig. 17—Selectivity curves.

Impulse Noise Interference

Oscilloscopic investigations of the effects of impulse interference with both modulated and unmodulated signals were made with the four receivers. The results indicated a general superiority in noise reduction for the frequency-dividing locked-in oscillator system.

Field Tests

Field tests showed the receivers using the new receiving system

to be considerably more selective with respect to adjacent-channel interference than conventional commercial receivers. More distortion was, however, encountered when the locked-in oscillator receivers were tuned so that the signal was received at the edges of the receiver-response characteristic than was obtained with the conventional units. This is due to the oscillator breakout characteristic and the fact that the voltage at the discriminator remains fixed irrespective of the signal applied to the oscillator. In general, it can be stated that an increase in distortion, when tuned to one side of a desired signal, goes hand in hand with increased adjacent-channel selectivity in any type of radio receiver. Some observers felt that this effect assisted in properly tuning the receiver.

Observations with respect to noise reduction substantiated the laboratory measurements which previously have been discussed.

Modified-Circuit Arrangement

An experimental receiver was also constructed incorporating the modified arrangement illustrated by Figure 15. Although the tests on this receiver were not so extensive as those on the receivers in which the Figure 16 arrangement was used, they did indicate that the modified circuit possessed superior noise reducing and adjacent-channel selectivity characteristics.

Conclusions

A novel method of receiving frequency-modulated signals has been investigated both theoretically and experimentally. The investigation indicates that the system has the following advantages:

1. By restricting the lock-in range of the oscillator to follow only the frequency variations which occur within the desired-signal channel, a material improvement in selectivity is obtained.
2. An equivalent voltage step-up is secured at a different and lower frequency than the intermediate-frequency and a corresponding improvement in freedom from over-all feedback is secured.
3. A constant voltage is applied to the discriminator irrespective of the strength of a received signal, and arrangements for minimizing amplitude variations in a received signal are, therefore, not required.
4. The frequency-dividing locked-in oscillator receiving system provides a means for incorporating, in a frequency-modulation receiver, a type of selectivity which can be used to discriminate between the desired-signal modulation and frequency-modulation-noise components.

The following characteristics should also be considered in an evaluation of the system:

1. Adequate receiver gain ahead of the locked-in oscillator must be provided if distortion of the weaker signals (due to the oscillator falling out of step), is to be prevented.
2. When the receiver is tuned through a signal, more noticeable distortion occurs at the edges of the receiver response characteristic than is obtained with a corresponding conventional receiver.

ACKNOWLEDGMENT

The writer wishes to acknowledge the valuable assistance of Messrs. M. S. Corrington, G. L. Grundmann, W. R. Koch, and W. F. Sands during the development of the frequency-dividing locked-in oscillator frequency-modulation receiver.

FREQUENCY-MODULATION DISTORTION CAUSED BY MULTIPATH TRANSMISSION*†

By

MURLAN S. CORRINGTON

Home Instrument Department, RCA Victor Division,
Camden, N. J.

Summary—When a frequency-modulated wave is received by more than one path, so that two or more of the voltages which are induced in the antenna have nearly the same amplitude, considerable distortion can result. Large objects, such as hills or high buildings, can reflect and absorb the waves and thus cause interference. Two interfering waves will be in phase part of the time and out of phase part of the time during modulation. This causes amplitude modulation on the resultant carrier, and a sharp irregularity in the instantaneous frequency, corresponding to each hole in the carrier caused by the interference.

The distortion can often be greatly reduced by detuning the receiver or discriminator until the hole in the carrier coincides with the zero-balance point of the discriminator. Directional antennas are helpful when the signals from the desired station are not coming from the same direction.

Formulas are derived for the modulated envelope and for the distortion. A Fourier-series analysis of the distorted audio output makes it possible to calculate the effects of de-emphasis networks and other audio selectivity.

INTRODUCTION

DURING a demonstration of a frequency-modulation receiver in New York in the spring of 1942, difficulty was experienced in locating an antenna in the RCA Building which would produce a signal of quality comparable with that known to be transmitted by station W2XWG. The antenna was located in a room in this steel building on the side away from the transmitter. The distortion was present even on relatively small deviations.

The tests were repeated in an apartment house on 93rd Street, and similar distortion was encountered. Further tests were made later with a second frequency-modulation receiver to determine whether the distortion was caused by a defective receiver or was introduced during the transmission of the signal. Both receivers produced the same type of distorted output. Since intermediate-frequency amplifier selectivity would not cause this effect, it was suggested that it might be caused by multipath reception. During July and August of 1942, further observations on a qualitative basis were made on the eighty-fifth floor of the Empire State Building. Path differences due to reflections were encoun-

* Decimal classification: R113.110XR148.2.
† Reprinted from *Proc. I.R.E.* December, 1945.

tered which were of sufficient magnitude to cause partial or complete cancellation of the waves at one or more frequencies in a 150-kilocycle-wide channel as used by frequency-modulation broadcast stations. A large number of field tests have been made since that time, and they show that this trouble can occur in nearly any location in Manhattan as well as in many of the surrounding towns in northern New Jersey.

In midtown Manhattan there are many steel buildings which cause multipath transmission. If the receiver is inside such a structure, an outside antenna is desirable. If such an antenna is not practicable, it probably will be necessary to use more than one inside receiving antenna in order to obtain full performance on the high-fidelity programs from all the New York frequency-modulation stations. This will be true even if the receiver has the best performance characteristies now commercially available. For a single stationary antenna, distortion will probably be noticed in many locations on at least one of the stations.

Cases have been observed where the movement of a person around in the room near the receiving antenna has changed the relative field strengths enough so that, when the person was in some parts of the room, the reception was satisfactory, and when he was in other parts it was unsatisfactory. A parasitic element such as a metal rod or dipole can sometimes be moved around in a room to produce the same effect. Maximum distortion occurs at the time of cancellation of the two waves. This causes a great reduction in the voltage in the intermediate-frequency channel and it may fall below the threshold of limiting in the receiver.

If the cancellation occurs near one edge of the band, the signal may sound fairly good during the soft passages but it becomes distorted on the loud passages. It will then sound like an overloaded audio-amplifier stage. If the cancellation occurs near the middle of the channel, the entire output may be so distorted that it becomes almost unintelligible; it then sounds somewhat like severe selective fading in amplitude modulation. Since high-order harmonics are produced, the distortion may appear as buzzes, rattles, or swishes, and if it occurs during the playing of a phonograph record, one may think the pickup is not tracking properly. The high frequencies often tend to become irritating due to intermodulation.

Two amateurs[1] have reported distortion of the programs from WMIT when the receiver is in the mountainous region near Asheville, N. C. At times they found that the audio signal was badly distorted to the point of becoming unintelligible, even for strong signals, and that

[1] A. D. Mayo and Charles W. Sumner, "F. M. distortion in mountainous terrain," *QST*, vol. 28, pp. 34-36; March, 1944.

the distortion is a constant phenomenon over periods of weeks in those spots where it occurs. DuMont and Goldsmith[2] found that multipath conditions are common in which two signals of approximately equal strength arrive at a receiving antenna, and therefore this type of distortion with frequency-modulation sound transmission can be expected to occur frequently.

A similar type of distortion had previously been encountered in long-distance transmission. In 1930, Eckersley[3] was working with amplitude-modulation transmitters which had incidental frequency modulation. Because of this frequency shift he found "most appalling distortion" resulting from delayed echoes caused by deflections from the Heaviside layer. He found that it was necessary to use special precautions to keep any frequency shift from getting into the amplitude-modulation transmitters. The distortion when frequency modulation and amplitude modulation were both present was worse than that for amplitude modulation only.

A study of frequency-modulation propagation over long distances was made by Crosby.[4] A transmitter was set up in Bolinas, California, and the transmissions were observed on receivers at the Riverhead, New York, station. He found that there was considerable distortion, and that in some cases a signal which gave fair intelligibility on amplitude modulation was practically unintelligible on frequency modulation. He stated, "The general conclusion derived from the tests and theory is that, on circuits where multipath transmission is encountered, frequency modulation is impracticable." In a later paper,[5] he again reported similar distortion. Other studies[6,7] have shown that reflections from the ground and nearby buildings cause multipath transmission, for both horizontal and vertical polarization, which can introduce distortion. Reflections from airplanes flying overhead sometimes cause interference for short intervals.

The same difficulty is encountered in television reception. The higher frequencies used tend to increase the difficulty, since the phase changes encountered are greater. This causes light and dark bands in

[2] Allen B. DuMont and Thomas T. Goldsmith, Jr., "Television broadcast coverage," *Proc. I.R.E.*, vol 32, pp. 192-205; April, 1944.

[3] T. L. Eckersley, "Frequency modulation and distortion," *Exp. Wireless and the Wireless Eng.*, vol. 7, pp. 482-484; September, 1930.

[4] Murray G. Crosby, "Frequency-modulation propagation characteristics," *Proc. I.R.E.*, vol. 24, pp. 898-913; June, 1936.

[5] Murray G. Crosby, "Observations of frequency-modulation propagation on 26 megacycles," *Proc. I.R.E.*, vol. 29, pp. 398-403; July, 1941.

[6] P. S. Carter and G. S. Wickizer, "Ultra-high-frequency transmission between the RCA building and the Empire State Building in New York City," *Proc. I.R.E.*, vol. 24, pp. 1082-1094; August, 1936.

[7] R. W. George, "A study of ultra-high-frequency wide-band propagation characteristics," *Proc. I.R.E.*, vol. 27, pp. 28-35; January, 1939.

the picture and results in synchronization difficulties when frequency modulation is used on video or synchronizing signals.

Since field tests have demonstrated the possibility of encountering distortion due to multipath propagation even though the distance from the transmitter to the receiver is relatively short, a theoretical and experimental study was made to determine the factors contributing to this type of distortion. It is the purpose of this paper to present the results of this investigation.

Theoretical Considerations

Suppose that one frequency-modulated wave is delayed with respect to the second by a time interval t_0 because of two-path transmission. The equations for the instantaneous voltages of the two waves are

$$e_1 = E_1 \sin \left\{ \omega t + \frac{D}{\mu} \sin 2\pi\mu t \right\} \tag{1}$$

$$e_2 = E_2 \sin \left\{ \omega (t - t_0) + \frac{D}{\mu} \sin 2\pi\mu (t - t_0) \right\} \tag{2}$$

where ω = unmodulated-carrier angular frequency

D = maximum frequency deviation

μ = audio frequency

t_0 = time delay of the second wave with respect to the first

E_1 = amplitude of first wave

E_2 = amplitude of second wave.

By combining these two waves by the parallelogram law, the equation for the resultant voltage becomes, as shown in Appendix I, equation (22),

$$e_1 + e_2 = E_1 \sqrt{1 + x^2 + 2x \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}} \tag{3}$$

$$\sin \left[\omega t + \frac{D}{\mu} \sin 2\pi\mu t - \tan^{-1} \frac{x \sin \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}}{1 + x \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}} \right]$$

where

$$z = 2 \frac{D}{\mu} \sin \pi\mu t_0 \tag{4}$$

and

$$x = \frac{E_2}{E_1}. \tag{5}$$

The first part of this expression (that under the radical sign) gives the variations in amplitude of the resultant carrier, and the argument of the sine function shows the variations in phase of the signal. There is considerable amplitude modulation introduced, and the frequency-modulated signal is also distorted. The resulting carrier envelope will be considered now, and the distortion in the audio output will be discussed later.

Effect of Interference on the Envelope of the Radio-Frequency Wave

The resultant amplitude of the radio-frequency carrier is given by

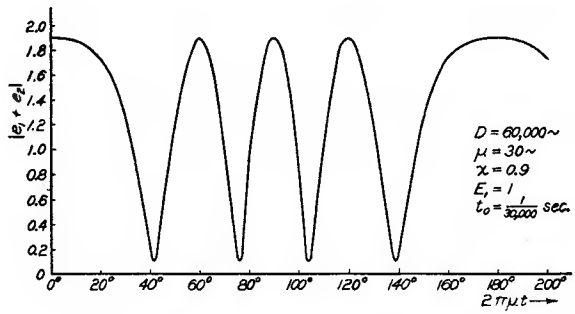


Fig. 1—Carrier envelope.

$$|e_1 + e_2| = E_1 \sqrt{1 + x^2 + 2x \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}}. \tag{6}$$

The amplitude is a function of the ratio of the two signal voltages x, the maximum deviation D, the time delay t_0, and the audio frequency μ. The only effect of the carrier frequency ω is to determine the initial phase angle ωt_0.

Figure 1 shows the carrier envelope for a deviation of 60,000 cycles per second from a mean carrier frequency of 45 megacycles per second. E_1 was assumed to be one volt and E_2 was taken as 0.9 volt, so the voltages are nearly equal. The conditions were chosen so that the two radio-frequency signal voltages were in phase at the undeviated position corresponding to $2\pi\mu t = 90$ degrees. The equation of the envelope becomes

$$|e_1 + e_2| = \sqrt{1.81 + 1.80 \cos \{4\pi \cos (2\pi\mu t - 0.18°)\}}. \tag{7}$$

The path difference corresponding to t_0 is 6.2 miles. Zero degrees correspond to full deviation on one side.

Since the audio signal is a cosine function, it varies slowly at first, then more rapidly as it goes through 90 degrees, and then slows down as it approaches 180 degrees. Figure 1 shows how the spacing of the peaks and dips of the carrier envelope corresponds to this variation. Zero degrees corresponds to the slowest rate of change, and the peak is very wide; at 90 degrees, the variation is most rapid, and the peak is narrow. The amplitude can be represented by the resultant of two rotating vectors R as shown by Figure 2 If E_2 were to rotate at a uniform rate, the resultant R (which corresponds to the instantaneous value of the envelope), would go through identical cycles, and the peaks of Figure 1 would be evenly spaced. When the carrier wave is modulated sinusoidally, however, the vector E_2 does not rotate uniformly, but rocks back and forth, with a sinusoidal variation of the angle θ.

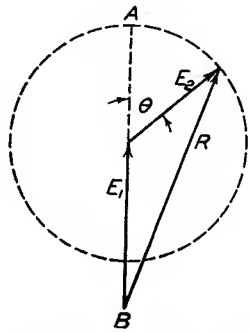


Fig. 2—Resultant of two signals.

Under the conditions represented by Figure 1, the vector E_2 starts with $\theta = 0.18$ degrees and makes two revolutions (4π radians) while the carrier frequency varies from maximum deviation to the mean undeviated frequency. Each time it goes past point B it gives a hole in the carrier envelope, and when it goes through A it gives a peak.

Figure 3 shows what happens when the frequency is increased from 30 cycles per second to 5000 cycles per second. The equation of this envelope is

$$| e_1 + e_2 | = \sqrt{1.81 + 1.80 \cos \{12 \cos (2\pi\mu t - 30°)\}}. \qquad (8)$$

Since $\pi\mu t_0 = \pi/6$ radians $= 30$ degrees, the curve is shifted over approximately 30 degrees from that in Figure 1, corresponding to one half the fraction of the audio cycle by which one wave is lagging behind the other.

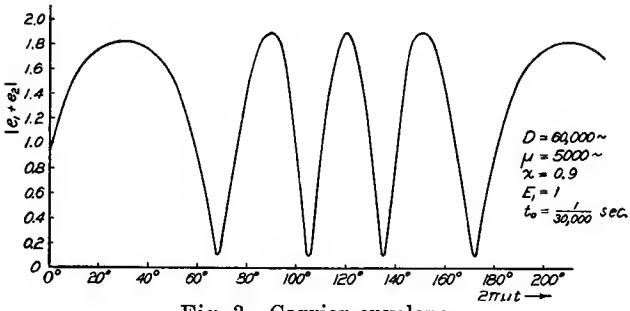


Fig. 3—Carrier envelope.

Another interesting difference is that the two voltages come almost, but not quite, into phase at 30 degrees. This peak is therefore slightly lower than the others. The difference is caused by the fact that the sine of an angle is not linear, but departs from a straight line for larger angles. The maximum angle through which the one vector of Figure 2 rotates with respect to the other is given by

$$z = 2\frac{D}{\mu} - \sin \pi\mu t_0 \text{ radians} \tag{9}$$

and this sine function introduces the nonlinearity. Further changes of the frequency merely move the curve along the axis, but effect slight changes in shape.

Variation of the Envelope with Deviation

Figure 4 shows the effect of changes of the maximum deviation D. The constants were chosen such that the two voltages are always in phase when $2\pi\mu t = 0$ degrees. For a small deviation, say $D = 15{,}000$ cycles per second, the two voltages start to go out of phase as $2\pi\mu t$ increases, but do not go completely out of phase. The maximum angle between the vectors in this case is

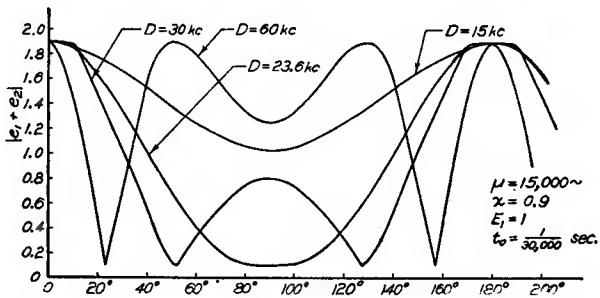


Fig. 4—Variation of carrier envelope with D.

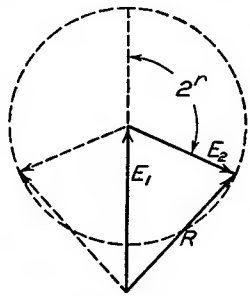


Fig. 5—Variation of resultant.

$$z = 2\frac{D}{\mu}\sin \pi\mu t_0$$
$$= 2\frac{15{,}000}{15{,}000}\sin\frac{\pi\,(15{,}000)}{30{,}000}$$
$$= 2 \text{ radians.}$$

The vector E_2 rotates clockwise 2 radians, then reverses and goes 2 radians counterclockwise, as shown in Figure 5. This process gives the envelope shown in Figure 4.

If the maximum deviation D is increased to 23,600 cycles per second, the two voltages go just out of phase in the mean position. A further increase in D causes them to start to come into phase again, as shown by the curve for $D = 30{,}000$ cycles per second. Further increases in D increase the number of peaks and holes in the envelope.

Variation of the Envelope with Path Difference

A change in the path difference for the two waves will also change the number of peaks and holes. Figure 6 shows the envelopes for dif-

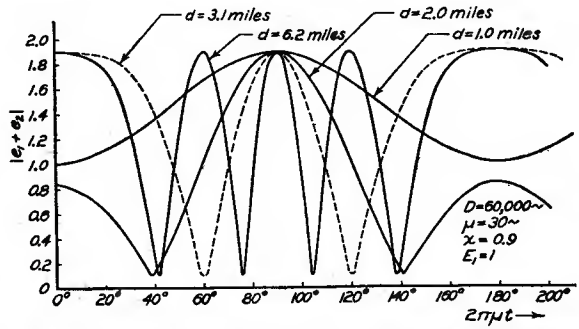


Fig. 6—Variation of carrier envelope with path difference.

ferent path differences from 1.0 mile to 6.2 miles. The curves are very similar to those of Figure 4. The shapes can be changed further by adjusting the initial phase of the two waves.

Variation of the Amplitude with Deviation

When the wave is modulated sinusoidally, the variation of the frequency deviation is not linear, and this causes the unequal spacing of the peaks and dips as shown in Figure 1. If the same set of conditions is used, and the variation of amplitude is plotted against the deviation, the spacing of the peaks and dips will be uniform as shown by Figure 7. A further increase in deviation will cause more peaks, but will not alter the spacing if the time delay is unchanged.

The equation of the amplitude is

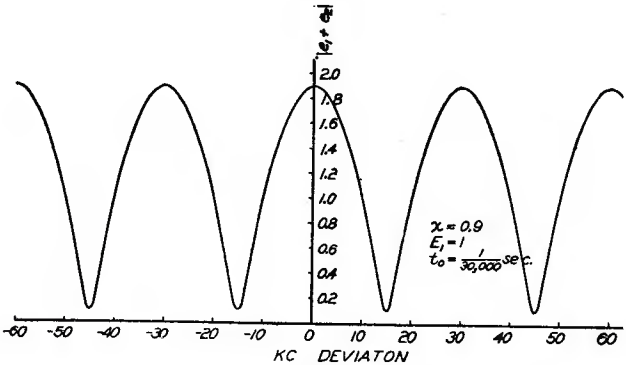


Fig. 7—Variation of amplitude with deviation.

$$| e_1 + e_2 | = \sqrt{1.81 + 1.80 \cos\left(\frac{2\pi D}{30{,}000}\right)} \qquad (10)$$

if the waves start out in phase.

Experimental Study of Carrier Envelope

An artificial time delay was obtained by passing the signal through 10,500 feet of high-frequency cable, and then combining it with a signal from the same source which had traveled only a short distance. The circuit is shown in Figure 8.

The frequency-modulation signal generator was modulated by a sinusoidal audio tone, and the potentiometer P was adjusted until the two voltages on the oscilloscope were nearly equal. The maximum deviation D was adjustable, and the initial phase could be adjusted by shifting slightly the mean carrier frequency. A series of oscillograms was taken that showed the carriers beating together for different

deviations D. It took approximately 30,000 cycles per second deviation to cause the carrier to go from in-phase to out-of-phase.

The oscillograms of Figure 9 show how the number of peaks and holes varies with the deviation D. They are seen to be very similar to those of Figures 1, 3, 4, and 6. In each one, the two voltages were nearly equal.

Figure 10 shows an oscillogram in which one signal is about 3.5 times as strong as the second. The general shape is the same as those of Figure 9, but the holes do not go nearly so deep. The velocity of propagation in the cable was 66 per cent of that in free space; therefore, the effective path difference was 15,900 feet. This corresponds to a time delay of 16.2 microseconds.

Determination of Time Delay

Let the path difference be d and the time delay corresponding to

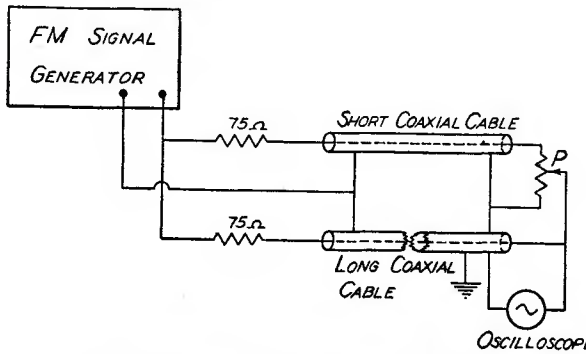


Fig. 8—Circuit for studying interference.

this difference be t_0. The unmodulated carrier frequency is ω. Let λ_1 be the wavelength corresponding to ω, and let λ_2 equal the wavelength corresponding to $\omega + D$, where D is the frequency deviation required to increase by one the number of wavelengths in d.

Then

$$d = N\lambda_1 = (N + 1)\ \lambda_2 \tag{11}$$

and

$$d = N\frac{3 \times 10^8}{\omega} = (N + 1)\frac{3 \times 10^8}{\omega + D}\ \text{meters} \tag{12}$$

where $\omega\lambda_1 = 3 \times 10^8$ meters per second, the free-space velocity of the wave, and $(\omega + D)\ \lambda_2 = 3 \times 10^8$ meters per second. Eliminate N from the second equation

$$d = \frac{d\omega}{\omega + D} + \frac{3 \times 10^8}{\omega + D}$$

and

$$d = \frac{3 \times 10^8}{D} \text{ meters.} \tag{13}$$

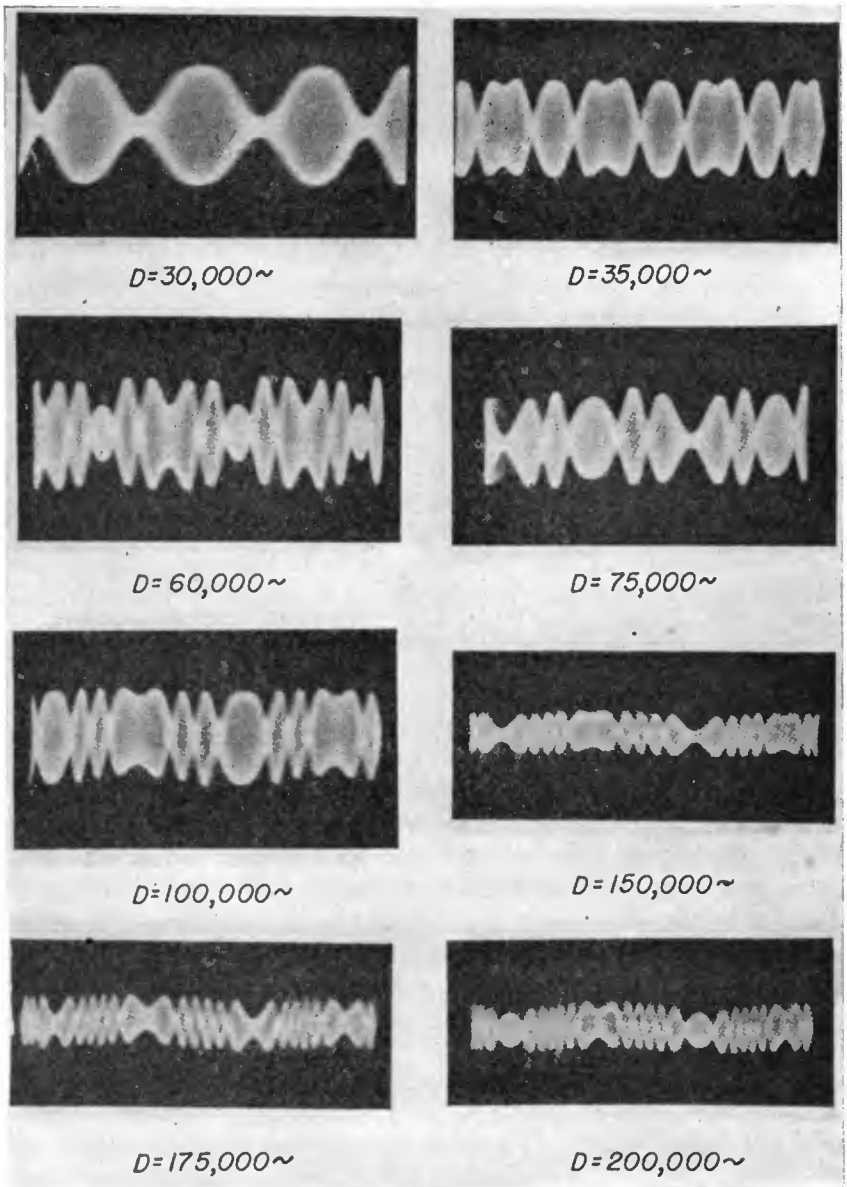


Fig. 9—Oscillograms of carrier envelopes.

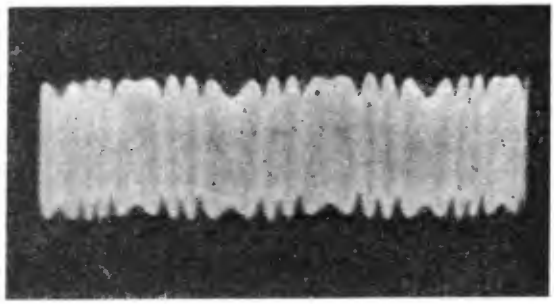


Fig. 10—Oscillogram of carrier envelope.

The time delay

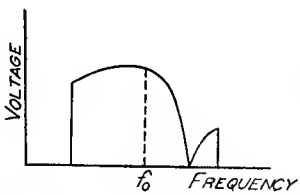
$$t_0 = \frac{d}{3 \times 10^8} = \frac{1}{D} \text{ seconds.} \qquad (14)$$

If D is the frequency shift required to go from out-of-phase, to in-phase, and out-of-phase again, the time delay will be $1/D$.

Field Tests

To study the distortion encountered in New York City, the NBC transmitter at the Empire State Building was operated with 300-cycle pure-tone modulation and a deviation of ± 75 kilocycles. A medium-priced receiver and an antenna were set up in a room next to the transmitter. The antenna could easily be located to give two-path reception, with a time delay of about seven microseconds. Figures 11 and 12 show the carrier envelope and the discriminator output. In Figure 11 the two waves canceled each other near one end of the swing. The resultant amplitude modulation which passed the limiter, and the coincidental phase modulation gave the result shown in Figure 12. One half of the cycle is relatively free from distortion, since the two signals did not go out of phase at any part of that swing. The two irregularities in the audio output correspond to the point of cancellation. This distortion could be eliminated by correct antenna location.

In the RCA Building on the side away from the transmitter, the



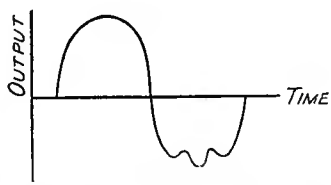


Fig. 11—Carrier envelope. Fig. 12—Discriminator output.

antenna was fairly well shielded from the direct wave from the transmitter, and it received signals which may have been reflected from other buildings. Figure 13 shows several curves of the audio output and the corresponding variation in the carrier voltage during the cycle. They are all for a 300-cycle audio tone and ± 75-kilocycle deviation, for various antenna positions. The audio output should be a pure sine wave, and the intermediate-frequency voltage should show the intermediate-frequency selectivity curve, as shown by the dotted lines. Although there were several irregularities along the straighter portions, the most serious distortion occurs at the ends of the swing. The first intermediate-frequency voltage curve shows that there were several paths of nearly equal voltage. In the second case, the secondary paths were still present, but the relative voltage was much less, as shown by the decreased departure from the selectivity curve. Since the

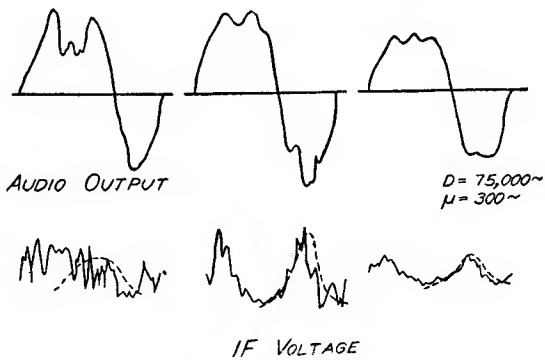


Fig. 13—Distortion for various antenna positions.

signal strength was fairly low, some of the distortion was probably due to the inability of the limiter to remove all the deep holes in the carrier.

Many other field tests in New York City and northern New Jersey have shown this distortion. The time delay is usually about 7 to 12 microseconds. This means that, for two-path transmission, there will be one or two holes in the carrier. Since the limiter tends to fail at these holes, the distortion can often be reduced by detuning the receiver until the hole in the carrier coincides with the discriminator zero-balance frequency. Since the discriminator output is zero at this point, variations in input voltage cause a minimum of distortion. It sometimes happens, when the direct and indirect waves are nearly equal, that severe distortion is encountered in the audio output of the set even though the input signal, as indicated by the voltage developed at the limiter grid, would normally be sufficient to secure limiting.

Effect of Interference on the Audio Output from the Discriminator

As shown in Appendix I, Equation (23), the audio output from the discriminator is proportional to

$$D \cos 2\pi\mu t + \frac{2D \sin \pi\mu t_0 \sin (2\pi\mu t - \pi\mu t_0)}{\dfrac{1/x + \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}}{x + \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}} + 1}. \quad (15)$$

The first term represents the undistorted output, and the second term shows the distortion introduced by the interference. This result was obtained by assuming that a limiter removes the coincidental amplitude modulation from the wave, and that the discriminator output is proportional to the instantaneous frequency.

General Case — Two Voltages Nearly Equal

The distortion term can be expanded in a Fourier series, as shown in Appendix II. The audio output from the discriminator is proportional to

$$\text{output} = D \cos 2\pi\mu t$$

$$- 2\mu \sum_{n=0}^{\infty} (-1)^n C (2n+1, z; x, \omega t_0) (2n+1) \sin (2n+1) \gamma$$

$$- 2\mu \sum_{n=1}^{\infty} (-1)^n S (2n, z; x, \omega t_0) (2n) \sin 2n\gamma \quad (16)$$

where

$\gamma = 2\pi\mu t - \pi\mu t_0$ and the C and S functions are defined as follows:

$$C (m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m (sn) \cos s\theta. \quad (17)$$

$$S (m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m (sn) \sin s\theta. \quad (18)$$

The amplitude of each harmonic can be calculated from these equations by assuming the proper value of n. The C and S functions can be computed from tables of Bessel functions. Figures 14, 15, 16, and 17 show some graphs of these functions.

Figure 18 shows the distortion caused at an audio frequency of 500 cycles per second under the conditions listed. The phase of the two

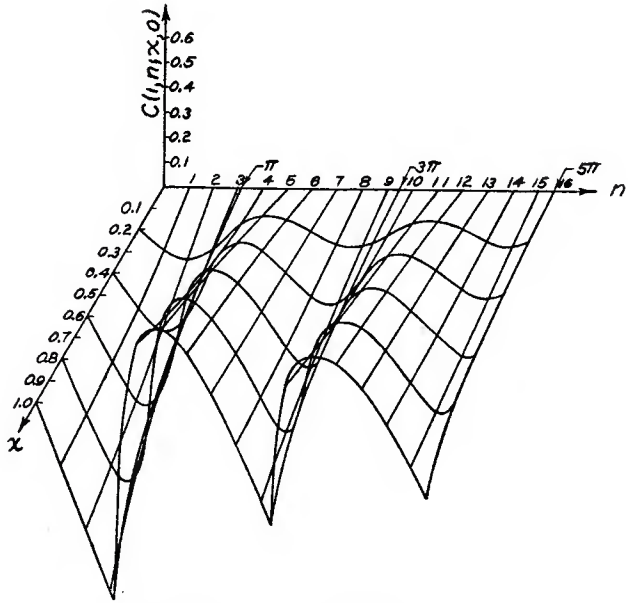


Fig. 14—Graph of $C(1, n; x, 0)$.

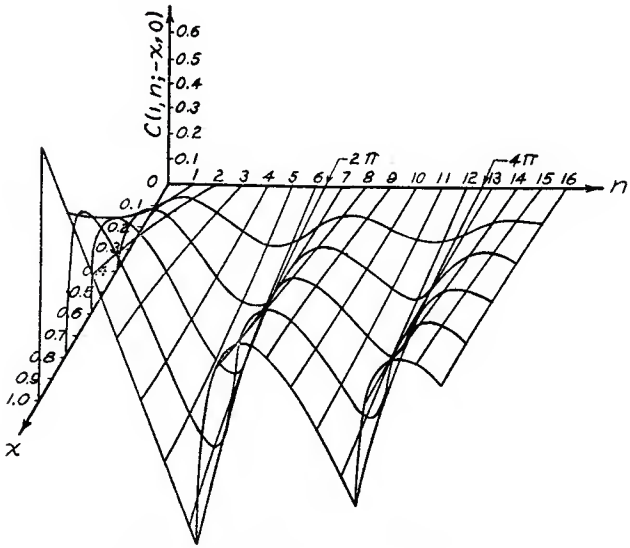


Fig. 15—Graph of $C(1, n; -x, 0)$.

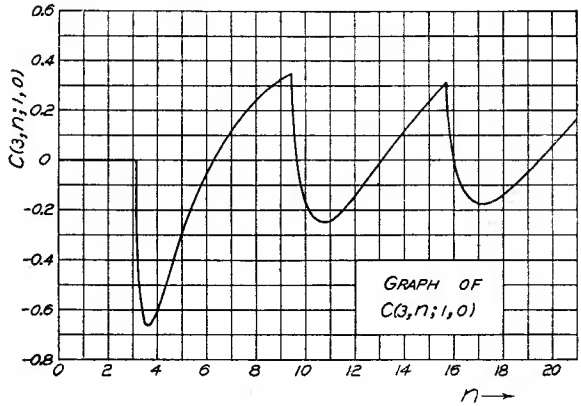


Fig. 16—Graph of $C(3, n; 1, 0)$.

radio-frequency waves was chosen such that they are 180 degrees out of phase at the unmodulated frequency. This gives the maximum of distortion. The first irregularity comes at about 10 degrees, where the vectors are still rotating relatively slowly with respect to each other. By the time the angle $2\pi\mu t = 63$ degrees, the two vectors are rotating rapidly and a sharp dip is created. A similar effect occurs at 93 degrees and 123 degrees. At 180 degrees the vectors are again rotating slowly, and the peak is small. The process is repeated during the second half of the audio cycle.

Figure 19 shows the distorted audio output when the audio frequency is increased to 5000 cycles per second. The value for ωt_0 was chosen to make the two radio-frequency signals start 90 degrees out of phase.

Figure 20 shows the two vectors E_1 and E_2 and the angle θ, which

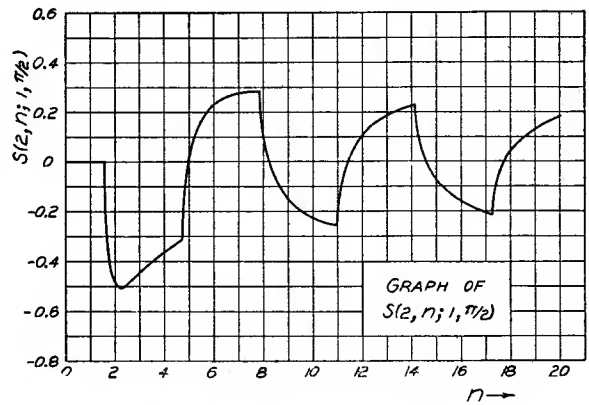


Fig. 17—Graph of $S(2, n; 1, \pi/2)$.

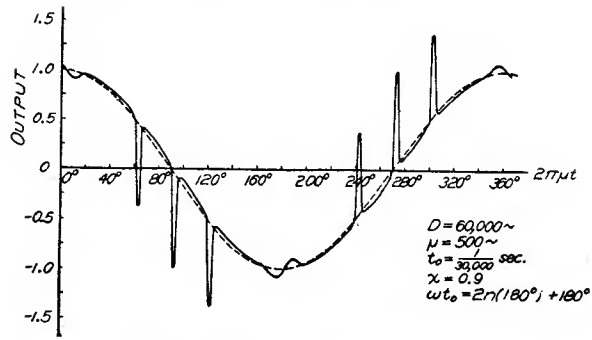


Fig. 18—Distorted audio output.

is the angle between them. The resultant R makes an angle α with E_1. The amount the distorted curve of Figure 19 departs from the pure cosine wave (shown by the dotted line) is proportional to the time rate of change of angle α. It is easy to see from Figure 20 that, as the angle θ nears and goes through 180 degrees, there is a very great change in α for a relatively small change in θ, when E_1 and E_2 are nearly the same length. This means that the first derivative of α with respect to time is very large, and this causes the deep dips and high peaks. The nearer E_2 comes to E_1 in absolute magnitude, the greater the departure from the undistorted curve. In the limit, as $E_1 = E_2$, $\alpha = +90$ degrees as θ approaches 180 degrees from the right and $\alpha = -90$ degrees as θ passes through 180 degrees. This means a 180-degree change in α as θ passes 180 degrees. The derivative of α with respect to time is therefore infinite at that instant, and the curve representing the audio output becomes a sharp pulse.

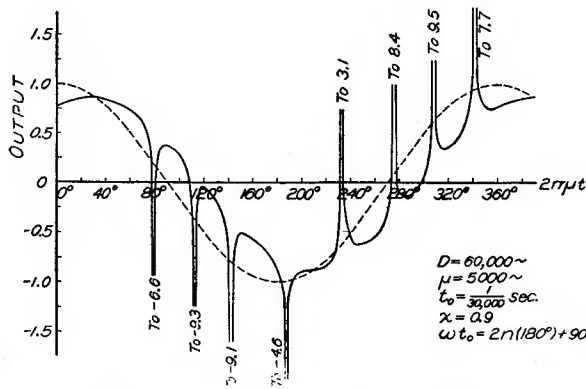


Fig. 19—Distorted audio output.

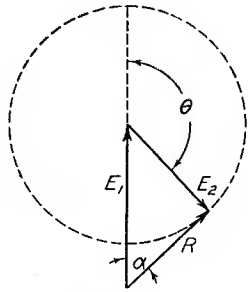


Fig. 20—Variation of α.

This theory assumes an infinite bandwidth and linear phase-shift in the receiver. The very deep and narrow dips and peaks show the presence of harmonics of high order which are in phase at each peak. Since the receiver has a limited bandwidth, many of these harmonics will be filtered out in the audio system, so the audio output will not be

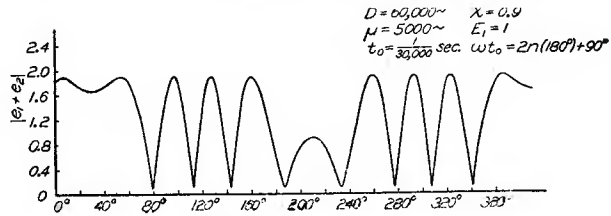


Fig. 21—Carrier envelope.

actually as distorted as shown. The effect of nonlinear phase shift and limited bandwidth is to broaden out and shorten the peaks in the output. The undistorted cosine curve is shown for comparison. The envelope of the carrier for this set of conditions is shown by Figure 21. There is one dip or peak corresponding to each hole in the envelope of the carrier.

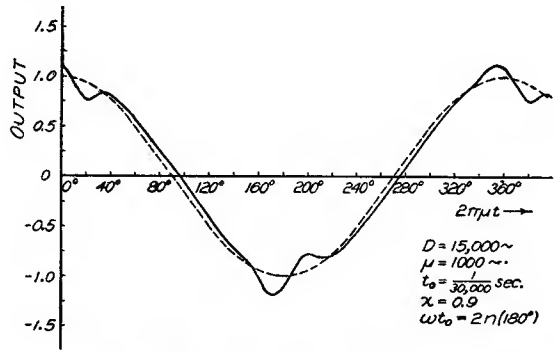


Fig. 22—Distorted audio output.

If the maximum deviation is reduced, such that the two vectors do not go out of phase so often, the distortion will be much less, because the number of peaks will be reduced and the one vector will not rotate so fast relative to the second. Figure 22 shows the output at 1000 cycles per second with a maximum deviation of 15,000 cycles per second. The time delay is 1/30,000 second as before. The phase of the two radio-frequency carriers was chosen such that they were in phase at the undeviated frequency. This assumption also reduces the distortion. Upon reducing the deviation further, or decreasing the time delay of the retarded signal, the irregularities will disappear.

Experimental Study of Distorted Output

The distortion in the audio output from the receiver was studied by using the circuit of Figure 8 with a frequency-modulation receiver connected just ahead of the oscilloscope. The signal was passed through the intermediate-frequency amplifier, limiter, and discriminator, and the distortion was studied with the oscilloscope. Figure 23 shows a series of oscillograms taken with different audio frequencies and deviations.

It is easily seen that high deviations and high audio frequencies cause the most distortion. Reducing either one will decrease the distortion. These curves cannot be compared directly with the theoretical ones previously shown, since the time delay is not the same. The time delay used experimentally was limited to 16.2 microseconds by the available length of cable, while the theoretical curves were based on a time delay of 33.3 microseconds. Naturally, the increased time delay causes the one signal to lag further behind and thus causes greater distortion. Crosby[8] has also shown many oscillograms of this type of distortion, taken over a path from Kansas City to Riverhead, Long Island. These show the characteristic sharp breaks in the audio output and the effect of noise.

CONCLUSION

The calculations and experimental study show that serious distortion of a frequency-modulation signal can be caused by multipath transmission over relatively short distances. This distortion is liable to be encountered in the vicinity of large buildings or other objects which reflect and absorb the waves and thus cause interference. These buildings or other large objects can be near either the transmitter or the receiver and cause such reflections. The two signals will be in phase part of the time and out of phase part of the time during the

[8] See page 401 of footnote reference 5.

audio cycle. If the signals have approximately equal amplitudes, nearly complete cancellation results when they go out of phase. This change in amplitude modulates the resultant carrier since it causes a sudden drop or deep hole in the resultant carrier amplitude, as shown by

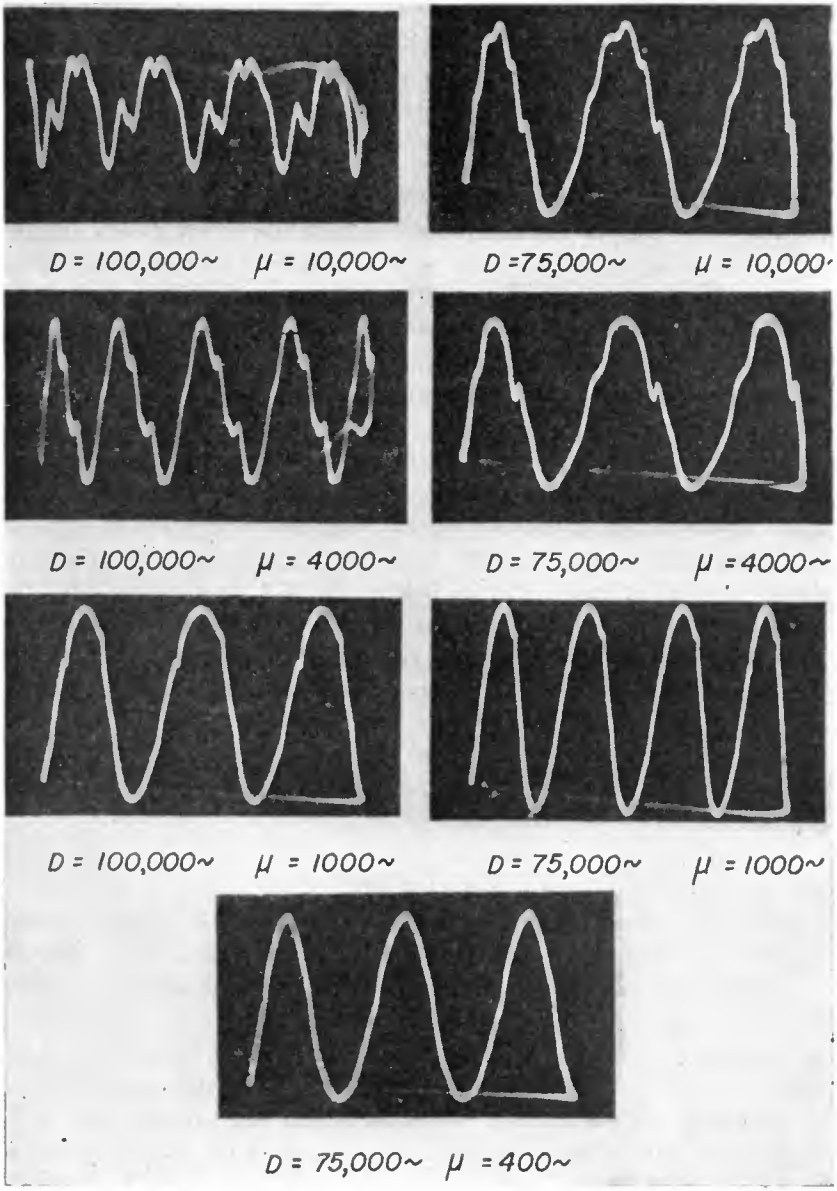


Fig. 23—Oscillograms showing distortion.

Figures 1, 3, 9, etc. If these holes are so deep that the limiter in the conventional receiver cannot hold the output constant, amplitude modulation will be impressed on the discriminator. If the discriminator can respond to these variations in amplitude, the output will be distorted at the part of the cycle which corresponds to the hole in the resultant carrier. If the resultant-carrier amplitude drops below the noise level, the distorted part of the cycle will have noise superimposed on it. Cases have been observed in which the output was undistorted at all parts of the cycle except at one point, and there the output was all noise.

Each time the two nearly-equal carriers go out of phase, there is a very rapid phase shift which causes a peak or dip in the output corresponding to this phase modulation. The worst distortion occurs when the two signals are of nearly equal voltage, and when the two radio-frequency carriers are out of phase when the modulation is removed. The distortion increases with the audio frequency because the higher the audio frequency, the further the retarded wave lags behind, in terms of an audio cycle. An increase in the maximum frequency deviation increases the number of times the two signals go in and out of phase in each audio cycle, and this also increases the distortion. This increase of the distortion as the deviation is increased is in contrast to the reduction of ordinary noise as the bandwidth of the transmission is increased.

The amplitude of each harmonic can be calculated by the formulas given. From these values, the effect of a low-pass audio filter can be determined. The low frequencies are not distorted very much, and the high frequencies will have their distortion reduced by the audio selectivity. This means that the medium frequencies will be the ones which are distorted the most and they are the ones most needed for intelligibility. For shorter path differences the time delay is less, and the distortion is reduced. If the two signals are approximately in phase at the undeviated position, they cannot go entirely out of phase for path differences less than about one mile with the deviations used today. However, if they are out of phase at the undeviated position, distortion will be produced even for small time delays. For small deviations and low audio frequencies the signals do not get far out of phase; thus there is not much distortion.

The distortion due to multipath transmission of frequency-modulated waves contains many high-order harmonics and is, therefore, different from amplitude-modulation distortion caused by the same process. In amplitude modulation the harmonics are all of low order and are not so disagreeable. Frequency-modulation distortion often sounds

like crackles, rattles, swishes, or gurgles and may not be objectionable on low modulation. It sometimes makes one wonder if the transmitter is being operated properly. At times it is noticeable only on the loud passages and may sound somewhat like overloading in an audio-amplifier tube. At other times the distortion may be so bad that the signal is almost unintelligible; it then sounds somewhat like selective fading in amplitude modulation.

It is doubtful whether an increase in selectivity can decrease the distortion. The hole in the resultant carrier and the coincidental phase modulation can occur at any part of the audio cycle, and thus can occur at the carrier frequency as well as any other. This cancellation at the center frequency cannot be eliminated by greater selectivity. An increase in sensitivity usually helps to decrease the distortion, since it helps to maintain sufficient carrier amplitude to operate the limiter at all times. A commercial frequency-modulation receiver designed to reduce multipath distortion should have a sensitivity of at least 10 or 20 microvolts, and for that signal strength should remove all amplitude variations in the signal even at the ends of the swing. The receiver should also remove the holes from the carrier; otherwise, distortion may be noted on some stations that may have much higher signal strengths than others which have no apparent distortion.

A properly oriented directional antenna, such as a dipole, will reduce the distortion if the two signals are coming from different directions, as is likely to be true near large buildings. Such antennas are of limited value when the signals are coming from the same direction. A power-line antenna or built-in loop often gives multipath trouble, and little can be done to reduce it since most such antennas are not ordinarily adjustable, and the voltage pickup is usually less than for a dipole.

In some receiving locations, the antenna should preferably be remote from the receiver so the relative field strengths will not be changed appreciably as the occupants of the room move about. Cases have been observed where the signal was either distorted or undistorted depending upon where one stood in the room. For time delays of approximately 6 or 8 microseconds, the distortion can often be greatly reduced by mistuning the receiver or discriminator until the hole in the carrier coincides with the zero-balance point of the discriminator, since any amplitude modulation that passes the limiter then causes very little audio output. However, this may cause an increase in noise. Under these conditions, a tuning meter is of little value.

Appendix I

Derivation of Equations for Frequency-Modulation Distortion Caused by Multipath Transmission

The analysis of this appendix is similar to that developed by Crosby,[4] and is repeated here for convenience. It will serve as an introduction to the derivation in Appendix II.

Let one frequency-modulated wave be delayed with respect to another by a given time interval t_0. This condition can be caused by a two-path transmission, where one path is longer than the other by a fixed distance. The equations for the instantaneous voltages of the two signals are

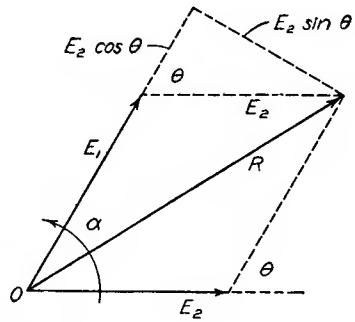


Fig. 24—Composition of two vectors.

$$e_1 = E_1 \sin\left(\omega t + \frac{D}{\mu}\sin 2\pi\mu t\right) \tag{19}$$

$$e_2 = E_2 \sin\left\{\omega\,(t - t_0) + \frac{D}{\mu}\sin 2\pi\mu\,(t - t_0)\right\} \tag{20}$$

where

ω = unmodulated-carrier angular frequency

D = maximum frequency deviation

μ = audio frequency

t_0 = time delay of the second signal with respect to the first.

These two waves can be combined by the parallelogram law. In Figure 24, the law of cosines gives

$$R = \sqrt{E_1^2 + E_2^2 + 2E_1E_2 \cos \theta}.$$

The angle θ between the two vectors E_1 and E_2 equals the difference of the two arguments of the sine functions.

$$\theta = \frac{D}{\mu} \sin 2\pi\mu t - \frac{D}{\mu} \sin 2\pi\mu\ (t - t_0) + \omega t_0.$$

The angle α between R and E_1 is given by

$$\tan \alpha = \frac{E_2 \sin \theta}{E_1 + E_2 \cos \theta} = \frac{x \sin \theta}{1 + x \cos \theta}$$

where $$x = \frac{E_2}{E_1}.$$

The law of cosines gives, therefore

$$e_1 + e_2 = \sqrt{E_1^2 + E_2^2 + 2E_1E_2 \cos \theta} \sin \left(\omega t + \frac{D}{\mu} \sin 2\pi\mu t - \tan^{-1} \frac{x \sin \theta}{1 + x \cos \theta} \right). \quad (21)$$

Since $\sin \alpha - \sin (\alpha - \beta) = + 2 \cos (\alpha - \beta/2) \sin \beta/2$,

$$\theta = 2 \frac{D}{\mu} \sin \pi\mu t_0 \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0$$
$$= z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0$$

where $z = 2D/\mu \sin \pi\mu t_0$.

The equation for the resultant voltage becomes

$$e_1 + e_2 = \sqrt{E_1^2 + E_2^2 + 2E_1E_2 \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}}$$
$$\sin \left[\omega t + \frac{D}{\mu} \sin 2\pi\mu t - \tan^{-1} \frac{x \sin \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}}{1 + x \cos \{z \cos (2\pi\mu t - \pi\mu t_0) + \omega t_0\}} \right]. \quad (22)$$

The amplitude term of this expression shows the variation of the resultant carrier amplitude, and the argument of the sine function

gives the frequency modulation and resulting distortion. When the signal goes through a perfect limiter the amplitude term is suppressed, and the argument of the sine term is effective at the discriminator.

Calculation of Audio Output

The output from the discriminator is proportional to the instantaneous frequency, where the instantaneous frequency[9] is defined by

$$f = \frac{1}{2\pi}\frac{d}{dt}(\text{argument of sine function})$$

$$= \frac{1}{2\pi}\frac{d}{dt}\left[\omega t + \frac{D}{\mu}\sin 2\pi\mu t - \tan^{-1}\frac{x\sin\{z\cos(2\pi\mu t - \pi\mu t_0) + \omega t_0\}}{1 + x\cos\{z\cos(2\pi\mu t - \pi\mu t_0) + \omega t_0\}}\right].$$

Since

$$\frac{d}{dx}\tan^{-1} u = \frac{1}{1+u^2}\frac{du}{dx}$$

and

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}$$

the derivative can be simplified to give

$$f = \frac{\omega}{2\pi} + D\cos 2\pi\mu t + \frac{2D\sin\pi\mu t_0\sin(2\pi\mu t - \pi\mu t_0)}{\dfrac{1/x + \cos\{z\cos(2\pi\mu t - \pi\mu t_0) + \omega t_0\}}{x + \cos\{z\cos(2\pi\mu t - \pi\mu t_0) + \omega t_0\}} + 1}. \quad (23)$$

A balanced discriminator is operated so there is no output at the unmodulated carrier frequency ω, and the output is proportional to the deviation from this frequency. The audio output is therefore proportional to the second and third terms of (23). The third term represents the distortion, and it can be serious.

[9] J. R. Carson, "Notes on the theory of modulation," *Proc. I.R.E.*, vol. 10, pp. 57-64; February, 1922.

Appendix II

Fourier Series Analysis of Distorted Audio Output

The third term of (23) of Appendix I represents the distortion in the audio output from the discriminator. It was obtained by differentiation of the argument of the sine function of (22). Consider this term

$$\frac{d}{dt}\tan^{-1}\frac{x\sin\{z\cos(2\pi\mu t-\pi\mu t_0)+\omega t_0\}}{1+x\cos\{z\cos(2\pi\mu t-\pi\mu t_0)+\omega t_0\}}=\frac{d}{dt}\tan^{-1}\frac{x\sin\theta}{1+x\cos\theta} \tag{24}$$

where $\theta = z\cos(2\pi\mu t-\pi\mu t_0)+\omega t_0$.

Let
$$\alpha=\tan^{-1}\frac{x\sin\theta}{1+x\cos\theta} \tag{25}$$

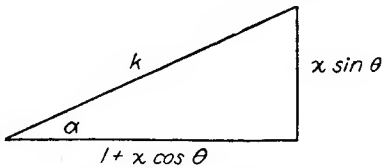


Fig. 25—Determination of k.

so
$$\tan\alpha=\frac{x\sin\theta}{1+x\cos\theta}.$$

From Fig. 25

$$k\sin\alpha=x\sin\theta \tag{26}$$

$$k\cos\alpha=1+x\cos\theta \tag{27}$$

where

$$k=\sqrt{(1+x\cos\theta)^2+(x\sin\theta)^2}$$

$$=\sqrt{1+x^2+2x\cos\theta}. \tag{28}$$

Multiply (26) by i and add (27).

$$1+x\cos\theta+ix\sin\theta=k\,(\cos\alpha+i\sin\alpha).$$

Change these complex numbers to the exponential form,

$$1+xe^{i\theta}=ke^{i\alpha} \tag{29}$$

and take logarithms of both sides,

$$\log (1 + xe^{i\theta}) = \log k + i\alpha. \tag{30}$$

Expand this in a power series by the well-known relation.

$$\log (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad -1 < x \leqq 1 \tag{31}$$

so

$$\log (1 + xe^{i\theta}) = xe^{i\theta} - \frac{x^2}{2} e^{2i\theta} + \frac{x^3}{3} e^{3i\theta} - \frac{x^4}{4} e^{4i\theta} + \cdots$$

$$= x (\cos \theta + i \sin \theta) - \frac{x^2}{2} (\cos 2\theta + i \sin 2\theta)$$

$$+ \frac{x^3}{3} (\cos 3\theta + i \sin 3\theta) - \frac{x^4}{4} (\cos 4\theta + i \sin 4\theta) + \cdots. \tag{32}$$

Equate the imaginary terms of (32), and use (30).

$$\alpha = x \sin \theta - \frac{x^2}{2} \sin 2\theta + \frac{x^3}{3} \sin 3\theta - \frac{x^4}{4} \sin 4\theta + \cdots. \tag{33}$$

Differentiate this expression with respect to time,

$$\frac{d\alpha}{dt} = (x \cos \theta - x^2 \cos 2\theta + x^3 \cos 3\theta - x^4 \cos 4\theta + \cdots) \frac{d\theta}{dt}$$

$$= 2\pi\mu z \sum_{n=1}^{\infty} (-x)^n \sin (2\pi\mu t - \pi\mu t_0)$$

$$\cos \{nz \cos (2\pi\mu t - \pi\mu t_0) + n\omega t_0\}. \tag{34}$$

Two lemmas will now be proved.

Lemma α

$$e^{ix \cos \theta} = \sum_{k=-\infty}^{\infty} i^k J_k (x) e^{ik\theta}.$$

Two expansions in series of Bessel coefficients due to Jacobi[10] are

[10] G. N. Watson, "A treatise on the theory of Bessel functions," The Macmillan Company, New York, N. Y., Second edition, 1944, p. 22.

$$\cos(x\cos\theta) = J_0(x) + 2\sum_{n=1}^{\infty}(-1)^n J_{2n}(x)\cos 2n\theta$$

$$\text{and } \sin(x\cos\theta) = 2\sum_{n=0}^{\infty}(-1)^n J_{2n+1}(x)\cos(2n+1)\theta.$$

Multiply the second equation by i and add the first equation. Change the cosine terms to the complex form. This gives

$$\cos(x\cos\theta) + i\sin(x\cos\theta)$$

$$= J_0(x) + \sum_{n=1}^{\infty}(-1)^n J_{2n}(x)\,e^{2in\theta} + \sum_{n=1}^{\infty}(-1)^n J_{2n}(x)\,e^{-2in\theta}$$

$$+ i\sum_{n=0}^{\infty}(-1)^n J_{2n+1}(x)\,e^{(2n+1)i\theta} + i\sum_{n=0}^{\infty}(-1)^n J_{2n+1}(x)\,e^{-(2n+1)i\theta}.$$

Since $(-1)^n = i^{2n} = i^{-2n}$ and $J_n(x) = (-1)^n J_{-n}(x)$, this proves the lemma.

Lemma β

$$\sin\theta\cos(a\cos\theta + b)$$

$$= 2\cos b\sum_{m=0}^{\infty}(-1)^m\frac{2m+1}{a}J_{2m+1}(a)\sin(2m+1)\theta$$

$$+ 2\sin b\sum_{m=1}^{\infty}(-1)^m\frac{2m}{a}J_{2m}(a)\sin 2m\theta.$$

In complex form, using lemma α

$$\sin\theta\cos(a\cos\theta + b)$$

$$= \frac{1}{4i}\{e^{i\theta} - e^{-i\theta}\}\{e^{i(a\cos\theta+b)} + e^{-i(a\cos\theta+b)}\}$$

$$= \frac{1}{4i}\left\{e^{i(\theta+b)}\sum_{k=-\infty}^{\infty} i^k J_k(a)\,e^{ik\theta} + e^{i(\theta-b)}\sum_{k=-\infty}^{\infty}(-i)^k J_k(a)\,e^{ik\theta}\right.$$

$$\left. - e^{-i(\theta-b)}\sum_{k=-\infty}^{\infty} i^k J_k(a)\,e^{ik\theta} - e^{-i(\theta+b)}\sum_{k=-\infty}^{\infty}(-i)^k J_k(a)\,e^{ik\theta}\right\}$$

$$= \frac{1}{4i}\left\{\sum_{k=-\infty}^{\infty} i^k J_k(a)\,[e^{i(k+1)\theta+ib} - e^{i(k-1)\theta+ib}]\right.$$

$$\left. + \sum_{k=-\infty}^{\infty}(-i)^k J_k(a)\,[e^{i(k+1)\theta+ib} - e^{i(k-1)\theta-ib}]\right\}$$

$$= \frac{1}{4i} \sum_{k=-\infty}^{\infty} \{ i^{k-1} J_{k-1}(a)\, e^{i(k\theta+b)} - i^{k+1} J_{k+1}(a)\, e^{i(k\theta+b)}$$

$$+ (-i)^{k-1} J_{k-1}(a)\, e^{i(k\theta-b)} - (-i)^{k+1} J_{k+1}(a)\, e^{i(k\theta-b)} \}$$

$$= -\frac{1}{4} \sum_{k=-\infty}^{\infty} \{ J_{k-1}(a) + J_{k+1}(a) \} \{ i^k e^{i(k\theta+b)} - (-i)^k e^{i(k\theta-b)} \}$$

$$= -\frac{1}{2} \sum_{k=-\infty}^{\infty} i^k \frac{k}{a} J_k(a) \{ e^{i(k\theta+b)} - (-1)^k e^{i(k\theta-b)} \}$$

$$= -\frac{1}{2} \sum_{k=1}^{\infty} \frac{k}{a} i^k J_k(a) \{ e^{i(k\theta+b)} + (-1)^k e^{-i(k\theta+b)}$$

$$- (-1)^k e^{i(k\theta-b)} - e^{i(k\theta-b)} \}$$

$$= \sum_{m=0}^{\infty} (-1)^m \frac{2m+1}{a} J_{2m+1}(a) \{ \sin [(2m+1)\,\theta + b]$$

$$+ \sin [(2m+1)\,\theta - b] \}$$

$$- \sum_{m=1}^{\infty} (-1)^m \frac{2m}{a} J_{2m}(a) \{ \cos (2m\theta + b) - \cos (2m\theta - b) \}.$$

The identities

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

and $$\cos (x+y) - \cos (x-y) = -2 \sin x \sin y$$

prove the lemma.

An application of lemma β to (34) gives the relation

$$\frac{d\alpha}{dt} = 2\pi\mu z \sum_{n=1}^{\infty} (-x)^n$$

$$\left\{ \cos n\omega t_0 \sum_{m=0}^{\infty} (-1)^m \frac{2\,(2m+1)}{nz} J_{2m+1}(nz) \sin (2m+1)\gamma \right.$$

$$\left. + \sin n\omega t_0 \sum_{m=1}^{\infty} (-1)^m \frac{2\,(2m)}{nz} J_{2m}(nz) \sin 2m\gamma \right\}$$

$$= 2\pi\mu \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} (-1)^m (-x)^n \frac{\cos n\omega t_0}{n}$$

$$2\,(2m+1)\,J_{2m+1}(nz)\,\sin\,(2m+1)\,\gamma$$

$$+ 2\pi\mu \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (-1)^m (-x)^n \frac{\sin n\omega t_0}{n} 2\,(2m)\,J_{2m}(nz)\,\sin 2m\gamma \qquad (35)$$

where $\gamma = 2\pi\mu t - \pi\mu t_0$. The audio output from the discriminator is therefore proportional to

$$\text{output} \propto D \cos 2\pi\mu t$$

$$- 2\mu \sum_{n=0}^{\infty} (2n+1)\,(-1)^n\,C\,(2n+1, z; x, \omega t_0)\,\sin\,(2n+1)\,\gamma$$

$$- 2\mu \sum_{n=1}^{\infty} (2n)\,(-1)^n\,S\,(2n, z; x, \omega t_0)\,\sin 2n\gamma \qquad (36)$$

where the C and S functions are defined as follows:

$$C\,(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m\,(sn)\,\cos s\theta$$

$$S\,(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m\,(sn)\,\sin s\theta.$$

The graphs of Figures 14, 15, 16, and 17 show how these functions vary for the particular values of the parameters which were chosen. Other graphs of the functions can be plotted by using tables of Bessel functions[11] which were prepared for this purpose.

[11] Murlan S. Corrington and William Miehle, "Tables of Bessel functions $J_n(x)$ for large arguments," *Jour. Math. and Phys.* (M.I.T.), vol. 24, pp. 30-50; February, 1945.

INPUT IMPEDANCE OF SEVERAL RECEIVING-TYPE PENTODES AT FM AND TELEVISION FREQUENCIES*

By

F. Mural

Industry Service Laboratory, RCA Laboratories Division,
New York, N. Y.

Summary—*The input impedance of vacuum tubes is an important circuit design consideration at high frequencies. This report includes information on the input impedance of a number of currently available r-f pentodes.*

Measurements are given on the variation of input resistance with frequency, and on the variation of input resistance and capacitance with plate current. Measurements are also given for the compensation of input resistance and capacitance variation with plate current by means of unbypassed cathode resistance.

The frequency range of measurement was chosen roughly to cover the frequency modulation and television transmission assignments as well as the recommended intermediate frequencies of receivers for these services.

Method of Measurement

THE resistance measurements were made by observing the change in Q when the vacuum tube input was shunted across a tuned circuit. For this purpose two standard commercial Q-meters were used to cover the desired frequency range. An important consideration in these measurements, particularly for low values of bias voltage and for high values of measured resistance, was to keep the r-f voltage across the tuned circuit to as low a value as possible. This was done not only to prevent grid current but also to avoid a shift of the tube's operating point. Since this difficulty was encountered primarily at the lower frequencies the Q-meter used for this work was equipped with an external 30-microampere d-c meter connected in series with the Q-indicating meter of the instrument. This meter was calibrated by connecting a standard signal generator with calibrated attenuator across the Q-meter terminals and using the 1-volt point on the Q-meter as reference. This arrangement also extends Q measurements to lower values.

It was also found necessary to place a shield plate, electrically connected to the case of the Q-meter, between the coil of the tuned circuit and the tube being measured. This was done to eliminate the effects

* Decimal Classification: R330 × R630.

of inductive and capacitative coupling between the tube and the tuned circuit.

The capacitance measurements were made by observing the detuning with the Q-meter's calibrated condenser.

In the measurements for varying conditions of transconductance, i.e. bias voltage, the parameter measured was the plate current. This was done because plate current is nearly proportional to transconductance, the variable of interest, and readily measured.

No attempt was made to operate the tubes at plate and screen voltages other than the typical operating conditions given in Table 1.

Table 1

Tube Type	E_b *Volts*	E_{sg} *Volts*	I_b *ma.*	g_m *μmhos*
6AC7	300	150	10	9000
6AG5	250	150	7	5000
6AK5	180	120	7.7	5100
6BA6	250	100	11	4400
6SG7	250	125	11.8	4700
6SH7	250	150	10.8	4900
6SJ7	250	100	3	1650
6SK7	250	100	9.2	2000

Limitations of Data

In considering the data presented in this report several conditioning factors must be borne in mind. Measurements of this type depend on how representative are the tubes selected and on the arrangement of the circuit components.

The tubes were selected from a limited quantity available and do not necessarily represent average values for any tube type or for tubes of any particular manufacture. The fact that input resistance has not in general been standardized and controlled in manufacture permits wide variations for any particular type. A variation of as high as 50 per cent was actually found in a group of tubes of the same type number.

Since the method of measurement used requires certain connections to be interposed between the points of measurement and the actual tube elements, the measurement is to a certain extent affected by these circuit connections. The effect of the inductance in the grid and grid-return lead is discussed in the next section. The by-passing of the plate, screen and grid-return are also known to affect the results. In

general, circuit values and structural features were designed to simulate conditions found in typical circuit applications.

For these reasons the graphically presented data is given in terms of smoothed out curves representing gross phenomena.

Input Resistance

The data is given in terms of input resistance since this is of practical interest to the designer. In the discussion, however, input conductance is also used because, when complex, it is more readily resolved into its component factors.

Figure 1 gives the results of measurements on a number of tubes illustrating the relationship between the input resistance and the frequency. These tubes were operated under the conditions given in Table 1. The circuit connections are shown in Figure 2, where C has the values 4000 and 500 micromicrofarads for the frequency ranges 10-50 and 50-200 megacycles respectively.

The input conductance of a vacuum tube with fixed operating voltages may be considered as varying with frequency according to the expression:

$$g_i = k_1 + k_2 f + k_3 f^2 \tag{1}$$

where the k's are constants for a specific tube and for specific operating voltages. The terms k_1 and $k_2 f$ are primarily dielectric losses due to leakage and hysteresis respectively, and distinguished from $k_3 f^2$ which is electronic. The electronic term, which is of particular interest at high frequencies, is generally resolved into two factors, i.e. the motion of electrons in the cathode-to-plate space and the reaction of the plate current through any residual coupling element between the plate and grid circuits. Although these factors could be analysed by means of electron ballistics and circuit theory an empirical method is indicated as being more practical for engineering purposes.

If k_1 and $k_2 f$ were negligible (1) could be written in terms of input resistance as:

$$R_i = \frac{1}{k_3 f^2} \tag{2}$$

Plotted in logarithmic coordinates (2) would give a straight line with a negative slope of 2. This is shown in Figure 1 for the purpose of comparison with the measured values.

The departure at low frequencies from the theoretical curve is due to the increasing importance of k_1 and $k_2 f$.

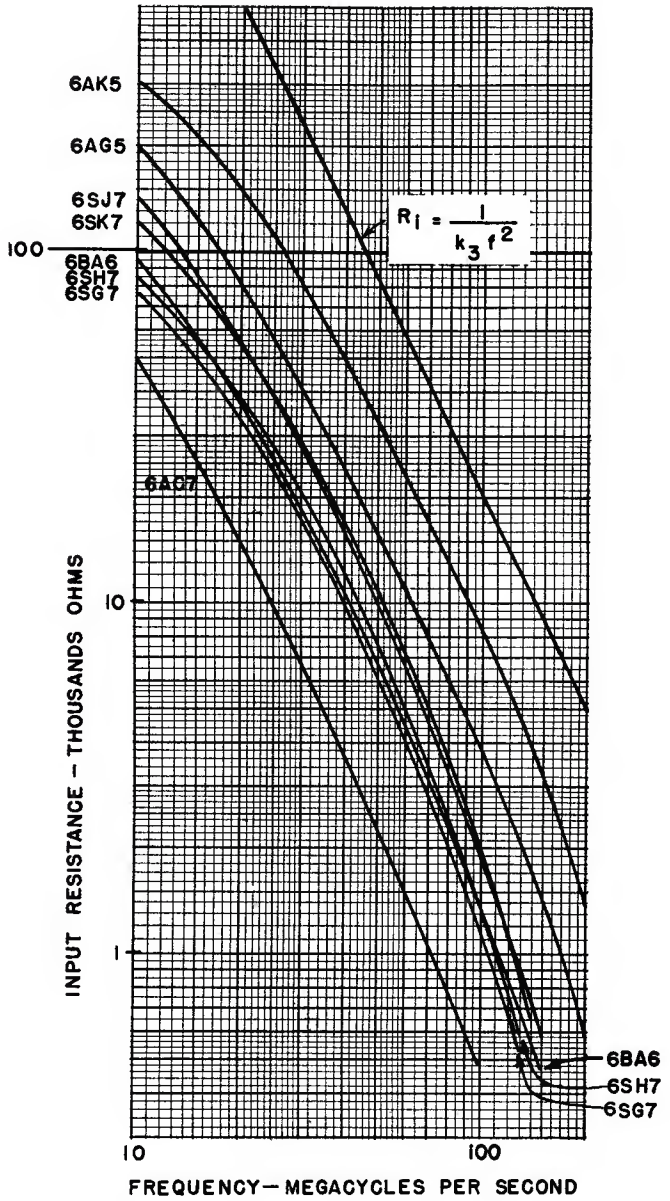


Fig. 1—Input resistance vs. frequency with plate by-passed and tubes operated under conditions given in Table 1.

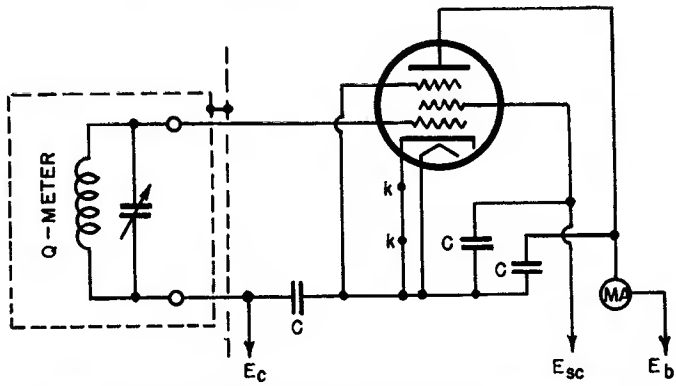


Fig. 2—Circuit for input impedance measurements.

The departure at high frequencies, however, is the result of approaching series resonance of the lead inductance, both internal and external, with the input capacitance of the tube. This condition is approximately indicated in Figure 3. Although the actual input resistance between grid and cathode, R_i', is inaccessible for direct measurement by the method employed, we can assume for the sake of argument that it does not depart from the inverse square law with frequency. The resonant effect then makes the apparent input resistance, R_i, seem to be lower than the corresponding value of R_i'. However, in

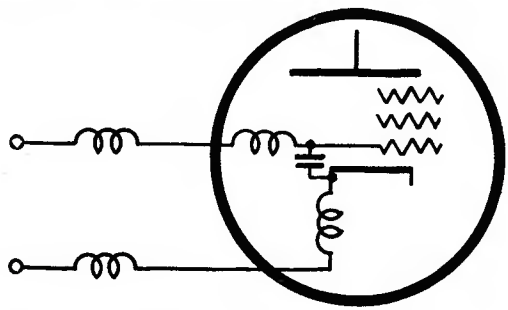


Fig. 3a—Series resonance in tube input circuit.

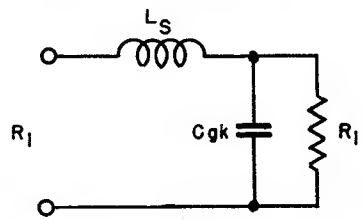


Fig. 3b—Equivalent tube input circuit.

addition to this result, the resonance effect gives a voltage rise between grid and cathode resulting in an increase in the apparent transconductance. This may be considered as a compensating effect with respect to the $R_i g_m$ figure of merit, which is discussed later, and should be taken into account by the designing engineer. It should be noted in

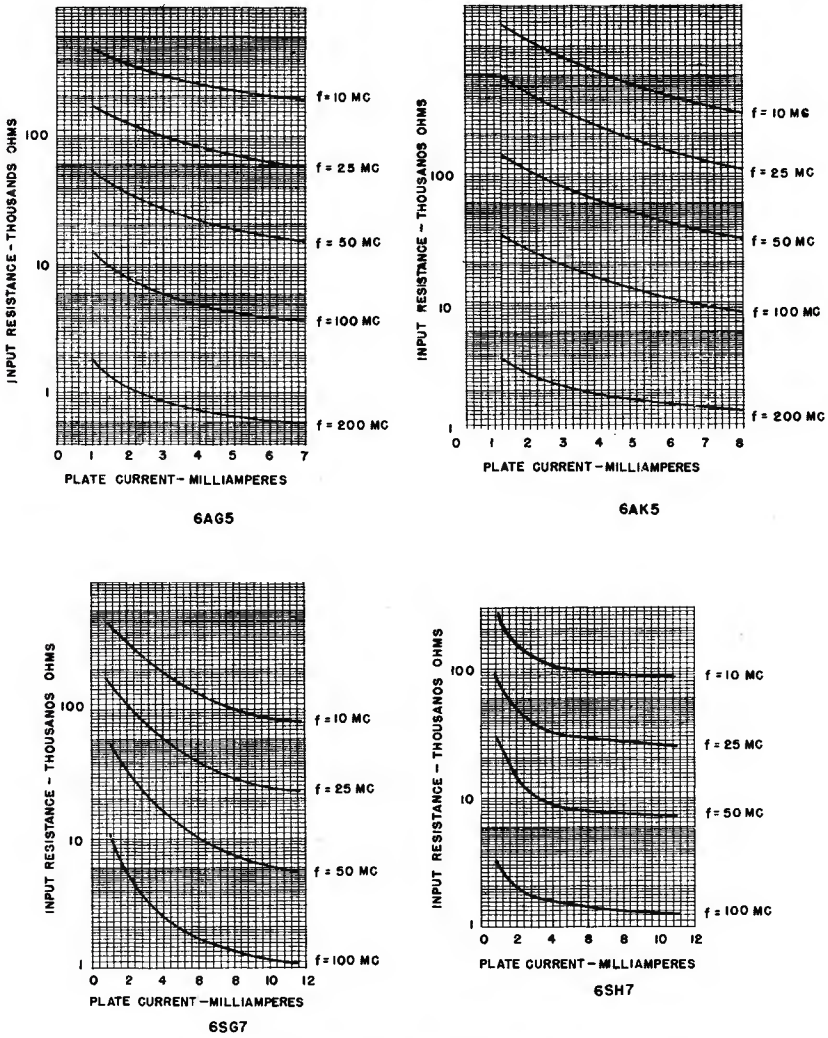


Fig. 4—Input resistance variation with plate current.

Figure 1 that this departure at high frequencies is less for the miniature types 6AK5, 6AG5 and 6BA6.

Figures 4 and 5 show the measurements of variation of input

resistance with plate current for the tubes in Table 1. The plate and screen voltages used were those in the table and the plate current was varied by adjustment of the externally supplied bias voltage as indicated in Figure 2.

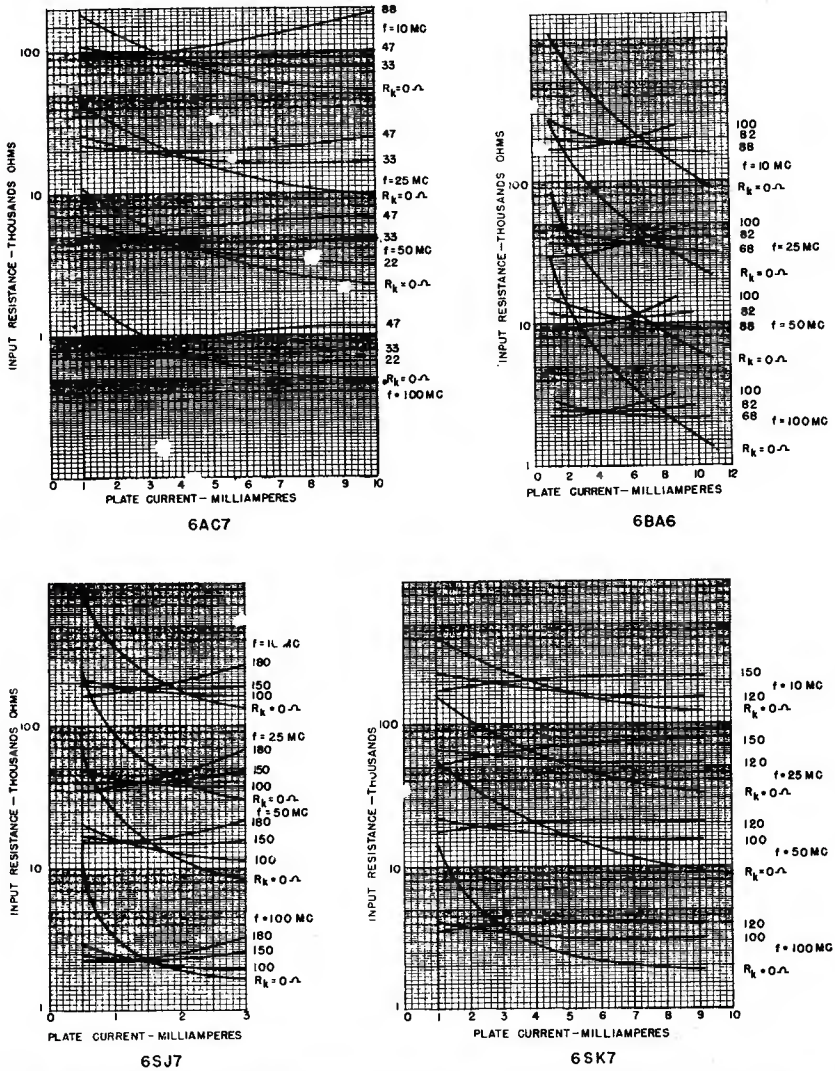


Fig. 5—Input resistance variation with plate current, showing compensating effect of unbypassed cathode resistance.

Figure 6 was obtained from the measurements of Figures 4 and 5 and the published characteristics of the tubes. These curves compare tube performance for the product of input resistance and transcon-

ductance which for most applications is a more useful figure of merit than input resistance alone as this product determines the gain of high frequency stages coupled by high Q single-tuned circuits.

In original design work at high frequencies some effort is always made to adapt the circuits to the tubes being used by obtaining an impedance transformation in the interstage coupling network. The

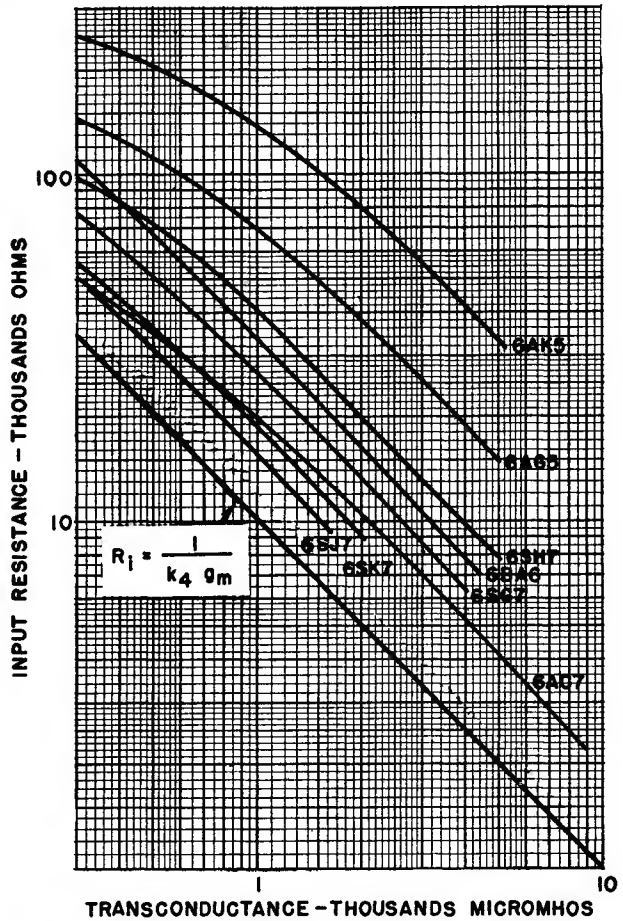


Fig. 6—Input resistance with variable transconductance at 50 megacycles.

gain is then proportional to the product of the square-root of the input resistance and the transconductance and the relative merit of tubes used under these conditions should be judged on this basis.

For a fixed frequency the input conductance may be considered as varying with transconductance according to the expression:

$$g_i = g_o + k_4 g_m \tag{3}$$

where g_o is the residual conductance with zero plate current and k_4 is a constant for a particular tube and a particular frequency. Inspection of the graphical data shows that for these tubes at the frequencies investigated the residual conductance, g_o, in (3) is so small that it can be neglected. This makes possible the expression of the variation of input resistance with transconductance by:

$$R_i = \frac{1}{k_4 g_m} \tag{4}$$

If plotted in logarithmic coordinated (4) would give a straight line with a negative slope of 1.

Compensation of Input Resistance Variation

The variation in input resistance of a vacuum tube may be substantially reduced by means of negative feedback through an unby-passed cathode resistor. This consideration is of particular value in the operation of tubes with varying bias voltage. Also the use of an unby-passed cathode resistor makes possible an increase in the effective input resistance for either fixed or variable bias operation. Figure 5 gives measurements of input resistance with unby-passed cathode resistors for those tubes which had their suppressors brought out in a separate base connection. The cathode resistor was interposed in place of the connection *kk* in Figure 2.

The effect of an unby-passed cathode resistor on the input impedance of a tube is discussed in report LB-450. Neglecting higher order frequency terms and making a number of approximations the input conductance of a tube with unby-passed cathode resistor may be expressed as:

$$g_i' = g_i - g_i R_k g_m - \omega^2 R_k^2 C_k C_{gk} g_m + \omega^2 R_k C_{gk}^2 \tag{5}$$

where g_i is the input conductance without the cathode resistor for the particular frequency and g_m being considered, R_k is the cathode resistance, C_k the cathode-to-ground capacitance and C_{gk} the grid-to-cathode-capacitance. In this expression g_i corresponds to the input resistance measurements with zero cathode resistance. The variation of g_i with transconductance of the tube may be corrected through the action of the second and third terms in (5). The fourth term, which is independent of g_m, adds a constant value of conductance. This cor-

responds to the average input conductance of the compensated tube. Since g_i, for the frequencies investigated, varies almost as the square of the frequency the value of R_k is almost independent of frequency. It may be noted from the measurements in Figure 5 that at lower frequencies the optimum compensating value of R_k runs somewhat higher. This is to be expected since at lower frequencies g_i varies more slowly.

The use of an unby-passed cathode resistor produces a loss in the effective transconductance of a tube. This is expressed by:

$$g_m' = \frac{g_m}{1 + \frac{R_k g_m}{\sqrt{1 + \omega^2 C_k^2 R_k^2}}} \tag{6}$$

where g_m is the nominal transconductance. Since the second term in the denominator is usually small compared to unity the $R_i g_m$ figure of merit is generally improved.

It should be noted that in the case of the 6BA6 tube the resistor for optimum compensation is somewhat larger than the normal cathode bias resistor. This makes it necessary to either operate the tube at a lower than normal value of nominal transconductance or secure the bias from a point in the circuit which is positive with respect to the cathode return point. It is also conceivable, since the departure is not too great, that the optimum compensating resistor could be used by changing the plate and screen voltages to shift the operating point.

Input Capacitance

It was found that for the frequencies and tubes investigated the input capacitance was practically independent of frequency.

The measurements presented in Figure 7 were made at 10 megacycles using the tube operating voltages in Table 1 and the circuit in Figure 1. If transconductance instead of plate current were used for abscissae the curve would be substantially linear so that the input capacitance could be expressed as:

$$C_i = C_o + k_5 g_m \tag{7}$$

which bears similarity to (3) except in that the first term is not negligible and the C's are independent of frequency.

Compensation of Input Capacitance Variation

The variation of input capacitance with transconductance is a disadvantage where the capacitance of the input circuit must remain unchanged under conditions of varying transconductance, i.e. varying grid bias. This is particularly true at the higher frequencies where the tube input capacitance usually forms a larger fraction of the total input circuit capacitance.

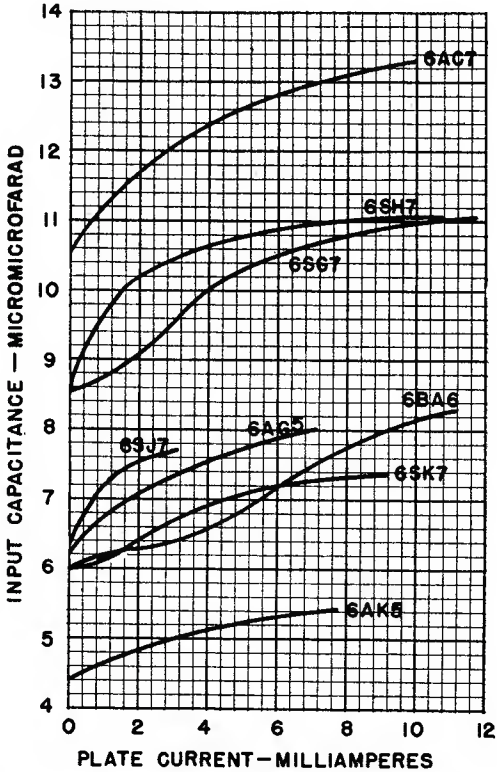


Fig. 7—Input capacitance variation with plate current.

The variation in input capacitance may be substantially reduced by means of an unby-passed cathode resistor. An approximate expression for the input capacitance with unby-passed cathode resistance is:

$$C_i' = C_i - C_{gk}R_k g_m - \omega^2 C_{gk}^2 C_k R_k^2 \qquad (8)$$

where C_i corresponds to (7), i.e. the measurements given in Figure 7. Since the third term is usually small an approximate value for R_k may be calculated to correct for the measured variation in C_i.

The measurements in Figure 8 correspond to those in Figure 5 in that the same values of unby-passed cathode resistance were used. It should be noted that at higher frequencies the third term in (8) becomes more important and contributes the negative shift observed in the measurements. If (8) is compared to (5) it is noted that the coefficient of the second term in each is identical, i.e. $R_k g_m$. Now if the third term in (5) were negligible the optimum compensating resistor for variations in both input resistance and input capacitance would be the same. However, since this is not usually the case, the optimum value of compensating resistance for input capacitance variation is somewhat higher than for input resistance variation.

Effect of Plate Load

The measurements given in this report are for plate by-passed to ground. Consideration of a finite plate load involves circuit design which is beyond the scope of this paper. However, it should be pointed out that the existence of a finite plate load has an effect on the input impedance of a vacuum tube through the coupling action of the grid-to-plate capacitance. The effective input admittance of a tube whose plate load is small in magnitude compared to its plate resistance and to the reactance of its grid-to-plate capacitance is given by:

$$Y_i' = Y_i + j\omega C_{gp} (1 + g_m Z_p) \tag{9}$$

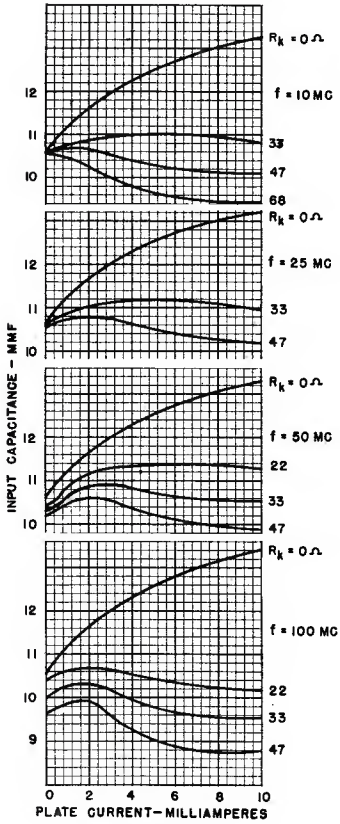
where Y_i is the input admittance between grid and cathode, C_{gp} is the grid-to-plate capacitance and Z_p is the plate load impedance. If Z_p were a pure resistance the effective input resistance would remain unaltered but the effective input capacitance would be increased.

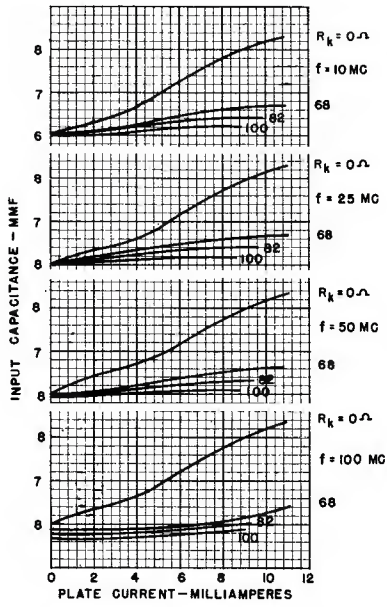
Conclusion

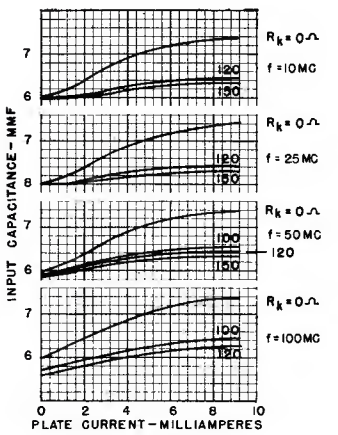
Because of the complexity of analysis the study of input impedance of vacuum tubes has been, for engineering purposes, largely empirical. Also, the compensation of undesired impedance changes has been effected largely through experiment.

The input resistance with zero plate load is practically inversely proportional to the square of the frequency for the frequencies of interest here. The input capacitance is practically independent of frequency.

The variation of input resistance is negative and that of input capacitance is positive with increase in transconductance. Negative feedback through an unby-passed cathode resistor can be used to substantially reduce these variations. In general, the value of cathode







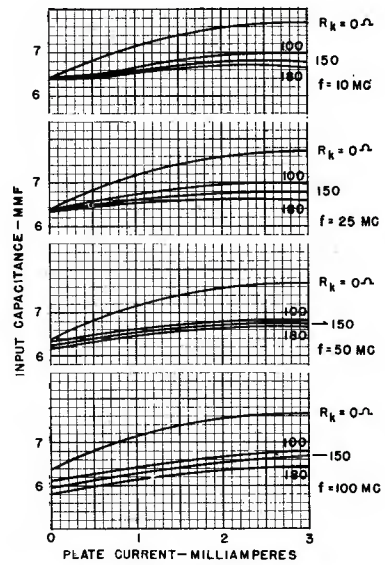


Fig. 8—Input capacitance variation with plate current, showing compensating effect of un-bypassed cathode resistor.

resistance must be higher for input capacitance than for input resistance variation compensation.

At the higher frequencies the designer should take into consideration series resonance of the lead inductance with grid-to-cathode capacitance. Also he should take into account any appreciable reaction due to a finite plate load.

FREQUENCY MODULATION DISTORTION CAUSED BY COMMON- AND ADJACENT-CHANNEL INTERFERENCE*†

By

MURLAN S. CORRINGTON

Home Instruments Department, RCA Victor Division,
Camden, N. J.

Summary—*During frequency-modulated radio broadcasting the signal is liable to be badly distorted whenever multipath transmission occurs or when any other interfering signal is present on the same or an adjacent channel. During hot weather, or before a storm, long-distance reception has been observed from frequency modulation broadcast stations on the 42–50 megacycle band. When such a distant station was in the same channel as a desired station, it sometimes happened that for short intervals the undesired station became stronger than the desired one. When this happened there was a small amount of noise and the programs suddenly changed. This interchange often lasted for several seconds but sometimes was limited to a word or two or a few notes of music.*

Formulas are given for computing the amplitudes of the harmonics and cross-modulation frequencies produced by the interference. These enable the calculation of the effect of a de-emphasis network following the discriminator, of a low-pass audio filter, and of nonlinear phase shift in the amplifiers.

INTRODUCTION

FOR several years it has been evident that frequency-modulated radio broadcasting offers certain advantages in noise reduction when compared with the usual amplitude-modulation systems. Many papers describe and discuss frequency modulation systems and their noise-suppressing properties.[1–8] Extensive field tests showed[9] that

---

* Decimal Classification: R148.2 × R430.

† Reprinted from *RCA REVIEW*, December, 1946.

[1] Edwin H. Armstrong, "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," *Proc. I.R.E.*, Vol. 24, No. 5, pp. 689-740; May, 1936.

[2] Murray G. Crosby, "Frequency Modulation Noise Characteristics," *Proc. I.R.E.*, Vol. 25, No. 4, pp. 472-514; April, 1937.

[3] H. Roder, "Noise in Frequency Modulation," *Electronics*, Vol. 10, No. 5, pp. 22-25, 60, 62, 64; May, 1937.

[4] E. H. Plump, "Störverminderung durch Frequenzmodulation," *Hochfrequenztechnik und Elektroakustik*, Vol. 52, pp. 73-80; September, 1938.

[5] Stanford Goldman, "F-M Noise and Interference," *Electronics*, Vol. 14, No. 8, pp. 37-42; August, 1941.

[6] Harold A. Wheeler, "Common-Channel Interference Between Two Frequency-Modulated Signals," *Proc. I.R.E.*, Vol. 30, No. 1, pp. 34-50; January, 1942.

(See next page for References 7, 8 and 9.)

when frequency modulation was used there was less interference produced by two stations operating at the same frequency than for the corresponding case of amplitude modulation, and that less power was required to cover a given area. It was also found that when the ratio of the carrier voltage to the noise voltage is high, the signal-to-noise ratio improvement due to frequency modulation is considerable. As the interfering noise voltage is increased with respect to the desired carrier-wave voltage, the improved noise suppression is obtained as long as the desired signal is several times as strong as the noise.

When a definite carrier-to-noise voltage ratio is reached (a ratio of 2 or 3 for wide-band frequency modulation) the amount of distortion in the audio output increases rapidly. When the noise voltage exceeds the signal voltage during all parts of the audio cycle, the noise eliminates the desired signal. This means that when frequency modulation is used the signal is either good or bad; there is only a small range for the ratio of carrier voltage to noise voltage that gives a noisy, but tolerable, signal.

Multipath transmission occurs when two or more interfering signals come from the same transmitter, but one is delayed with respect to the others because of a longer transmission path. Considerable distortion has been observed when multipath transmission occurs in frequency-modulated broadcasting and fairly complete discussions of this problem are available.[10–13] If the second wave comes from a different station than the desired wave, the result is common- or adjacent-channel interference according to whether the two carrier frequencies are nearly the same or are separated by the width of one channel.

There is not much information available on the amount of interference to be expected in the new frequency modulation band. The effects to be described were observed on the old 42–50 megacycle band and on the 30–42 megacycle police bands. The frequency of occurrence and the magnitude of these effects will not be known for the new 88–108 megacycle band until a reasonable number of transmitters with normal power and antenna gains are in operation. If such interference does occur, the analysis given here will be applicable.

[7] Herbert J. Reich, "Interference Suppression in A-M and F-M," *Communications*, Vol. 22, No. 8, pp. 7, 16, 19, 20; August, 1942.

[8] Robert N. Johnson, "Interference in F-M Receivers," *Electronics*, Vol. 18, No. 9, pp. 129–131; September, 1945.

[9] I. R. Weir, "Field Tests of Frequency- and Amplitude-Modulation With Ultra-High-Frequency Waves," *Gen. Elec. Rev.*, Vol. 42, Nos. 5 and 6, pp. 188–191, May, 1939; pp. 270–273, June, 1939.

[10] Murray G. Crosby, "Observations of Frequency-Modulation Propagation on 26 Megacycles," *Proc. I.R.E.*, Vol. 29, No. 7, pp. 398–403; July, 1941.

[11] A. D. Mayo and Charles W. Sumner, "F.M. Distortion in Mountainous Terrain," *Q.S.T.*, Vol. 28, No. 3, pp. 34–36; March, 1944.

[12] Murlan S. Corrington, "Frequency-Modulation Distortion Caused by Multipath Transmission," *Proc. I.R.E.*, Vol. 33, No. 12, pp. 878–891; Dec., 1945.

[13] S. T. Meyers, "Nonlinearity in frequency-modulation radio systems due to multipath propagation," *Proc. I.R.E.*, Vol. 34, No. 5, pp. 256–265; May, 1946.

Sometimes during hot weather, or before a storm, long-distance transmission has been observed from frequency modulation broadcast stations. During the summer of 1944, station WSM-FM in Nashville, Tenn. was heard often in Camden, New Jersey. During July it was very strong and free of noise for nine evenings in succession and it was heard several other evenings. Occasionally, long-distance reception from stations in all directions was observed. On July 7, 1944 nearly all the mid-western stations and several from other directions could be received in Camden, for about 2½ hours with a standard commercial receiver and indoors antenna. The following list of stations received was compiled that evening:

Call	*Station*	*Megacycles*
WWZR	Zenith Radio Corp., Chicago	45.1
WGNB	WGN, Inc., Chicago	45.9
WBBM-FM	Columbia Broadcasting System, Chicago	46.7
WDLM	Moody Bible Institute, Chicago	47.5
WSBF	South Bend Tribune, South Bend, Ind.	47.1
WMLL	Evansville on the Air, Evansville, Indiana	44.5
WENA	Evening News Assn., Detroit	44.5
WMFM	The Journal Company, Milwaukee, Wisconsin	45.4
WSM-FM	National Life & Accident Ins. Co., Nashville, Tenn.	44.7
WMIT	Gordon Gray, Winston-Salem, N. C.	44.1
WMTW	Yankee Network, Mt. Washington, N. H.	43.9
W2XMN	Edwin H. Armstrong, New York	43.1
WHNF	Marcus Loew Booking Agency, New York	46.3
WBAM	Bamberger Broadcasting Service, New York	47.1
WABC-FM	Columbia Broadcasting System, New York	46.7
WABF	Metropolitan Television, Inc., New York	47.5
WIP-FM	Pennsylvania Broadcasting Co., Philadelphia	44.9

Some interesting common-channel phenomena were observed. Stations WENA, Detroit, and WMLL, Evansville, were of nearly equal strength. First one, and then the other was received; they changed about every fifteen seconds. There would be a slight amount of noise and the programs would suddenly be interchanged. This continued for about one-half hour. Sometimes the carrier-wave voltage levels dropped below the level at which the limiter in the receiver operated and both programs could be heard simultaneously.

Stations WSBF, South Bend, and WBAM, New York, were also in a common channel. WSBF was stronger and was clear most of the time; WBAM would come in with sudden bursts of a word or two or a bit of music as station WSBF faded rapidly. These bursts occurred at interval of about ten seconds.

Some of the state police frequency-modulation systems have reported serious skip interference on numerous occasions. In Missouri, on the talk-back frequency of 39.78 megacycles, the interfering signals are usually those of the New Jersey State Police and the North Carolina Highway Patrol Cars, although cars of the Ohio State Patrol and those of Rhode Island occasionally cause interference. The signal strengths of the undesired stations are greatest during May, June, and July and range from weak to strong. The strong signals are of sufficient intensity to swamp out the local cars and may be received for an hour or two or for the whole day, from about two hours after sunrise to an hour or so after sunset.

The Florida State Patrol have reported considerable interference on frequency modulation from stations in California, New Jersey, Connecticut, and Massachusetts, and they have made car-to-car contacts with Pittsfield, Massachusetts. The Michigan State Police reported that signals from the Alabama State Patrol stations were received by their patrol cars with signal levels at the input to the receiver as high as 300 microvolts, and these stations in Alabama have taken control of their receivers throughout Michigan for hours at a time.

The Indiana State Police have had their cars blocked out by stations in Virginia and Oklahoma for all cars more than three miles from the transmitter. During the hunt for escaped German war prisoners near Carlisle, Indiana, on June 10th, the interference was so bad they had considerable difficulty maintaining contact with their cars. On June 22nd, during a man-hunt and road blockade following a bank holdup at San Pierce, Indiana, cars were completely blocked out at various times by cars in Virginia and Massachusetts. Further disruption of service was caused many afternoons by the second harmonic of short-wave broadcast stations in Massachusetts and New York.

Recent observations by the Federal Communications Commission show that such bursts or sudden increases in strength of signals received beyond the line of sight occur regularly.[14-15] The long-distance transmission that occurs during such bursts can be interpreted as reflections from media of height comparable to the E layer, but lying at each side of the great-circle plane. It is assumed that when meteors pass through the upper atmosphere, the air is ionized and this causes the bursts.

If a local station is on the same channel as a distant one which is being received in bursts, interference may be expected to occur for intervals as

---

[14] "Measurement of V-H-F Bursts," *Electronics*, Vol. 18, No. 1, p. 105; January, 1945.

[15] K. A. Norton and E. W. Allen, Jr., "Very-High-Frequency and Ultra-High-Frequency Signal Ranges as Limited by Noise and Co-Channel Interference," *Proc. I.R.E.*, Vol. 33, No. 1, p. 58; January, 1945.

long as several seconds. This might even cause the program to change suddenly from one station to the other during these short intervals.

Analysis of Fundamental Case

The most elementary case of frequency modulation interference is that produced when two unmodulated radio-frequency carriers, having nearly the same frequency, are added together. This gives the usual heterodyne envelope as the two voltages beat together. In addition there is a variation in the phase of the resultant which is equivalent to frequency modulation. If the difference in frequency of the two carriers is now varied sinusoidally by changing the frequency of one, keeping the two amplitudes constant, the result is common-channel interference or adjacent-channel interference, depending upon the way the one frequency is varied. It is thus evident that, if the most elementary case is properly analyzed, the frequency modulation interference is merely a generalization of the results.

Heterodyne Envelope

As shown in Appendix I, if two radio-frequency carriers $e_1 \sin \omega t$ and $e_2 \sin (\omega + 2\pi\mu)t$ are added, the heterodyne envelope is given by

$$\text{Envelope} = e_1\sqrt{1 + x^2 + 2x \cos 2\pi\mu t} \tag{1}$$

where

e_1 = amplitude of first carrier
e_2 = amplitude of second carrier
$x = e_2/e_1$
ω = angular frequency of first carrier, radians per second
μ = difference in frequency, cycles per second

This is the voltage that will be obtained if the resultant signal is sent through a linear rectifier and filtered. Figure 1 shows the variation of the envelope over one beat-note cycle as the ratio of the amplitudes of the two signals, x, is changed. For small values of x the envelope is approximately,

$$\text{Envelope} = e_1(1 + x \cos 2\pi\mu t) \qquad x \ll 1 \tag{2}$$

As the ratio x is increased gradually, the higher harmonics increase in amplitude; so the peaks become broader and the hole in the carrier becomes deeper and narrower. In the limit, as $x \rightarrow 1$, the envelope becomes a series of rectified cosine waves, or:

$$\text{Envelope} = 2e_1 \mid \cos \pi\mu t \mid \qquad x = 1 \tag{3}$$

Average Value of Envelope. If the resultant heterodyne voltage is sent through a linear rectifier, the direct-current voltage across the rectifier output increases gradually as x is increased. Figure 2 shows that this

voltage increases 27.3 per cent when x changes from zero to one. As shown in Appendix I, this voltage is given by:

$$\text{Average voltage} = \frac{2(1+x)e_1}{\pi} \, \mathrm{E}\left\{\frac{2\sqrt{x}}{1+x}\right\} \tag{4}$$

where $\mathrm{E}\left\{\frac{2\sqrt{x}}{1+x}\right\}$ is a complete elliptic integral of the second kind with modulus $\frac{2\sqrt{x}}{1+x}$.

Root-Mean-Square Value of Envelope. If a square-law rectifier instead of a linear rectifier is used, the root-mean-square value of the rectified envelope can be read with an average-reading direct-current voltmeter. The root-mean-square voltage will increase more rapidly with x than the average

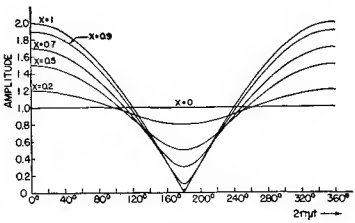


Fig. 1—The heterodyne envelope.

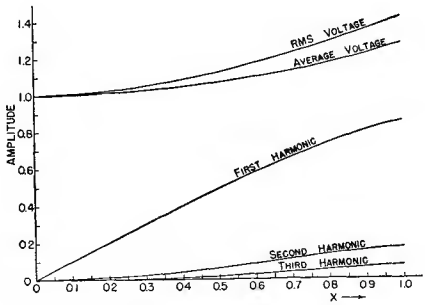


Fig. 2—Harmonic content of the heterodyne envelope.

voltage, as shown by Figure 2. The voltage is given by:

$$\text{Root-mean-square voltage} = e_1\sqrt{1+x^2} \tag{5}$$

and it increases 41.4 per cent when x increases from zero to one.

Fourier-Series Analysis of Envelope. If the heterodyne envelope is rectified with a linear rectifier, and the radio frequency is filtered out, the resultant audio signal (shown by Figure 1) can be expanded in a Fourier series. The coefficients of this series are given in Appendix I and the zero-frequency component is the same as the average value which is shown by Figure 2. The fundamental component increases almost linearly with increasing x to a maximum value of ⅔ of the corresponding direct current voltage. The second harmonic increases slowly until it equals 20 per cent of the fundamental when $x = 1$, and the third harmonic has a maximum value of 8.6 per cent of the fundamental.

Phase-Angle Variations

The two signals $e_1 \sin \omega t$ and $e_2 \sin (\omega + 2\pi\mu)t$ are in phase when $t = 0$. Since the frequency of the second signal is higher than the frequency of the first signal, this means that a vector representing e_2 will rotate with respect to one representing e_1. If e_1 is a vector rotating at ω radians per second, then e_2 will rotate at $\omega + 2\pi\mu$ radians per second.

Figure 3 shows the variation of the phase angle φ which the resultant, R, of e_1 and e_2 makes at any given instant with the vector e_1. When $t = 0$, the two vectors are in phase and $\varphi = 0$. At a later time $2\pi\mu t =$ 90 degrees, so e_2 and e_1 are at right angles and tan $\varphi = e_2/e_1 = x$. When $2\pi\mu t = 180$ degrees, φ is again zero. This process gives the variations in φ shown by Figure 4. The maximum value of φ is equal to $\sin^{-1} x$, as shown

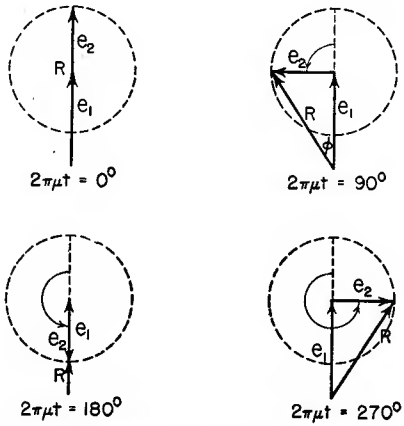


Fig. 3—Variations of the phase angle.

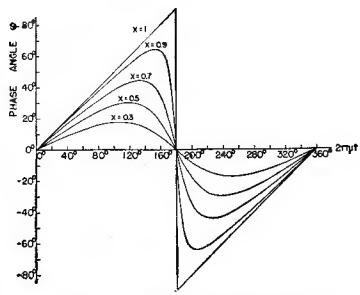


Fig. 4—Phase-angle variations.

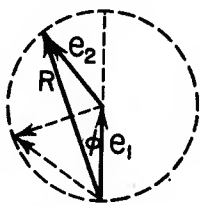


Fig. 6—Variations of φ.

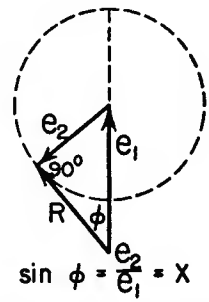


Fig. 5—Maximum value of φ.

by Figure 5. As x approaches one, the angle φ varies more and more rapidly near $2\pi\mu t = 180$ degrees. When $e_2 = e_1$ or $x = 1$, φ increases linearly from zero to 90 degrees as e_2 turns through 180 degrees.

As shown by Figure 6, φ is then an inscribed angle, and since an

inscribed angle is measured by one-half its intercepted arc, φ increases linearly when e_2 turns uniformly. As e_2 approaches cancellation of e_1, R is an infinitesimal vector and $\varphi \rightarrow +90$ degrees. As e_2 swings past cancellation, the direction of R suddenly reverses so $\varphi = -90$ degrees; i. e., there is an instantaneous change of φ equal to 180 degrees. Beyond that point φ increases linearly toward 0 degrees, as shown by Figure 4.

Instantaneous Frequency

The output from a linear discriminator is proportional to the instantaneous frequency, where the instantaneous frequency is defined by:[16]

$$f = \frac{1}{2\pi}\frac{d}{dt} \text{ (argument of sine function)}. \tag{6}$$

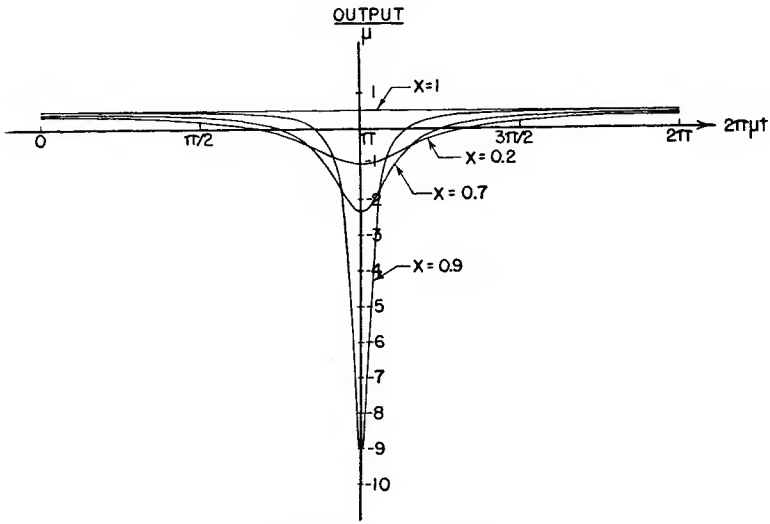


Fig. 7—Audio output, $x < 1$.

For a balanced linear discriminator, tuned to frequency ω, the output is proportional to the deviation in frequency from the center frequency ω. As shown in Appendix I, the output is given by

$$\text{Output} \propto \frac{\mu}{\dfrac{\cos 2\pi\mu t + 1/x}{\cos 2\pi\mu t + x} + 1} \tag{7}$$

Obviously this output is proportional to the slope of the curves of Figure 4, since it represents the first derivative with respect to time.

The curves of Figure 7 show the wave form in the audio output from

[16] J. R. Carson, "Notes on the Theory of Modulation," *Proc. I.R.E.*, Vol. 10, No. 2, p. 57; February, 1922.

a frequency modulation receiver, with perfect limiting and linear phase shift in the tuned circuits. As x approaches one, the output becomes more and more like an impulse, until at $x = 1$, the output has the constant value one-half except when $2\pi\mu t = \pi$; here the output becomes infinite. The area between the line one-half unit above the time axis and the curve for the instantaneous frequency over one cycle is constant for all values of x and equals $-\pi\mu$. This means that as $x \rightarrow 1$ the output is constant except at $2\pi\mu t = \pi$ and at that point is an impulse equal to $\pi\mu$ times a unit-impulse function.

When x becomes greater than one, the polarity of the impulse change but the shape is the same, as shown by Figure 8.

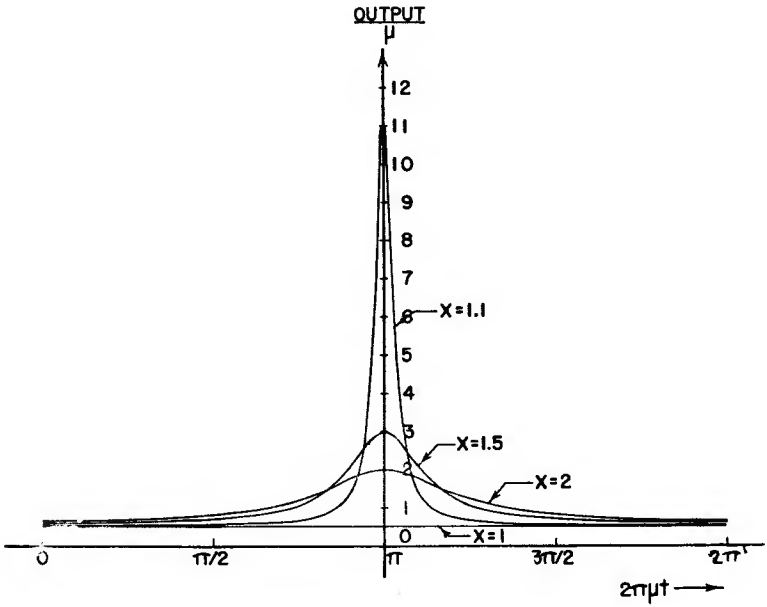


Fig. 8—Audio output, $x > 1$.

Average Value of Instantaneous Frequency. If the discriminator is tuned to the frequency ω, the average audio output is zero when $x < 1$. As shown in Appendix I, the average output is proportional to μ when $x > 1$. The curves of Figure 8 show this shift in average value when e_2 becomes stronger than e_1 and takes control.

Root-Mean-Square Value of Instantaneous Frequency. If the audio output from the discriminator is measured with an root-mean-square meter, the readings will vary as shown by Figure 9. The output increases uniformly from zero when $x = 0$ until it rapidly approaches infinity when $x = 1$. When $x > 1$ the output decreases uniformly to one as x becomes large:

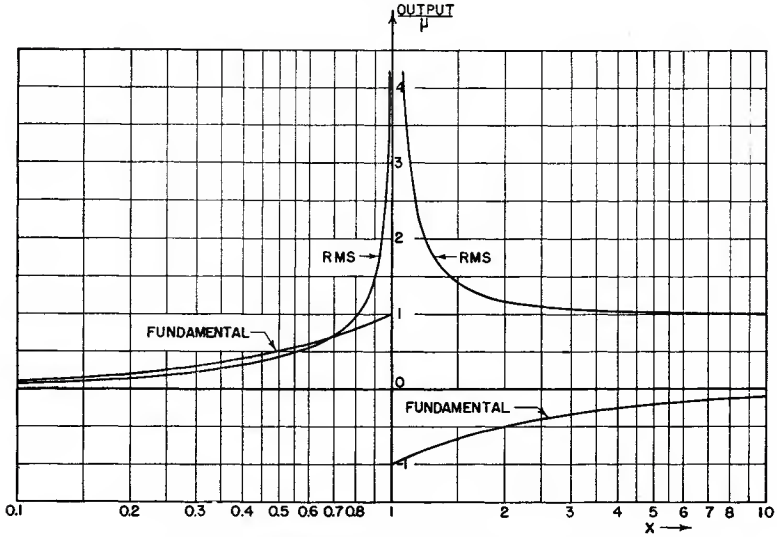


Fig. 9—Root-mean-square voltage output.

As shown in Appendix I:

$$\text{Root-mean-square output} \propto \frac{x\mu}{\sqrt{2(1-x^2)}} \qquad \text{when } x<1 \qquad (8)$$

$$= \mu \sqrt{\frac{2x^2-1}{2(x^2-1)}} \qquad \text{when } x>1. \qquad (9)$$

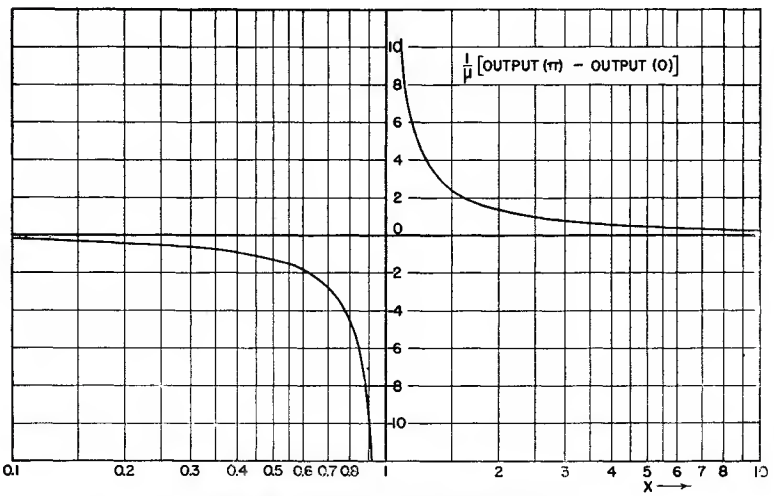


Fig. 10—Peak-to-peak audio output.

Peak-to-Peak Value of Instantaneous Frequency. The output when $2\pi\mu t = \pi$ minus the output at $2\pi\mu t = 0$ gives the peak-to-peak value of the instantaneous frequency. This is given by:

$$\text{Output } (\pi) - \text{Output } (0) = \frac{2x\mu}{x^2 - 1} \tag{10}$$

The curves of Figure 10 show how the peak-to-peak output varies as x increases. When $x = 1$, the peak-to-peak output becomes infinite, and it decreases uniformly beyond this point.

Harmonic Analysis of Instantaneous Frequency. If the harmonic content of the audio output is calculated by means of a Fourier-series analysis, the result can be expressed as:

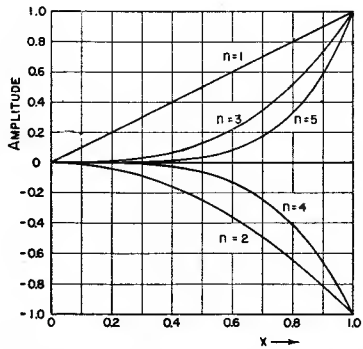


Fig. 11—Harmonic content of audio output.

$$\text{Output} \propto -\mu \sum_{n=1}^{\infty} (-x)^n \cos n(2\pi\mu t). \tag{11}$$

This means that the nth harmonic amplitude is proportional to μx^n. Figure 11 shows the increase of the harmonic amplitudes with increasing x for the first five harmonics. For small values of x, the higher harmonics are much smaller than the fundamental; but as x approaches one, the higher harmonics increase rapidly, until at $x = 1$ all harmonics are equal.

Effect of Limited Band Width. If the audio output from the discriminator is sent through a low-pass filter, having approximately linear phase-shift, the resultant wave form will depend upon how many harmonics are passed by the filter. In Figure 12, the case of $x = 0.9$ is shown for various low-pass filters. The case $n = 1$ means that only one harmonic, the fundamental, is passed by the filter. If two harmonics are passed, $n = 2$, the center begins to dip more because both harmonics are in phase at that point. The cases for $n = 3$ and $n = 5$ are also shown. The effect, there-

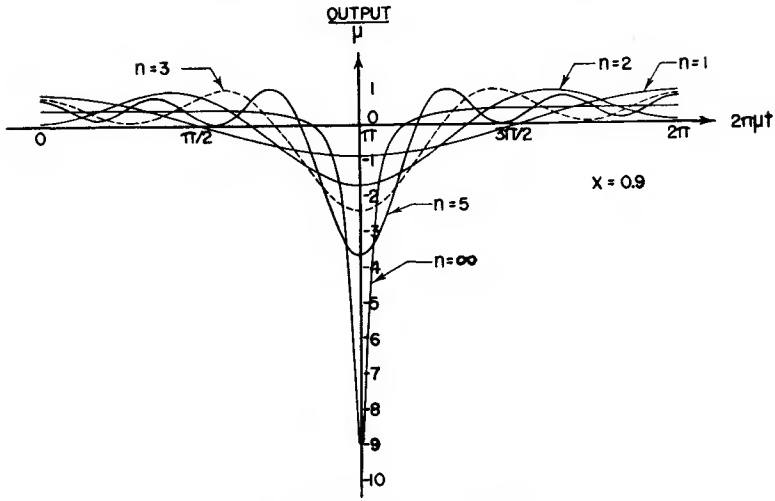


Fig. 12—Effect of low-pass filter.

fore, of limited band width is to reduce the output at $2\pi\mu t = \pi$ and to cause the resulting wave to oscillate about the curve that would be obtained with unlimited band width. For the case when $n = 5$, the peak output is reduced from 9.0 to 3.69, or the output becomes 41 per cent of that for unlimited band width. The curves of Figure 13 show the effect of limited band width. The variable on the axis of abscissas shows the number of harmonics passed by the low-pass filter, and the other axis shows the percent of peak amplitude compared to that for unlimited band width. Thus, if $x = 0.9$ and 10 harmonics are passed by the filter, the peak output

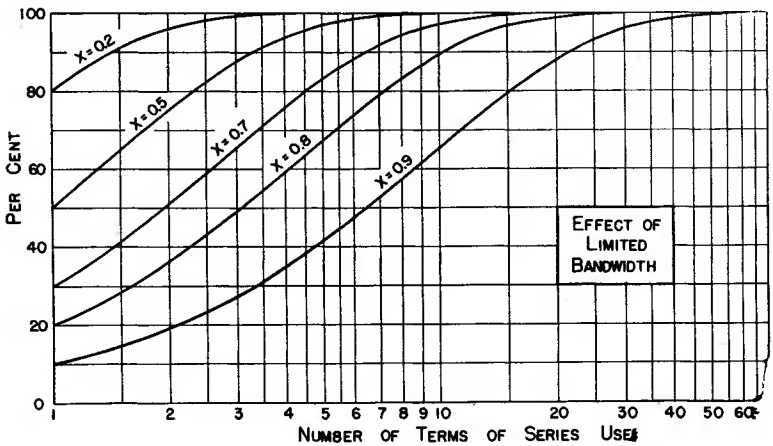


Fig. 13—Effect of limited band width.

will be approximately 65 per cent of what it would be if all harmonics were passed. If $x = 0.5$, it is evident that five or six harmonics will give nearly undistorted output.

As shown by Appendix I, this ratio of the peak output to the corresponding peak for unlimited band width is equal to $1 - x^n$ where n is the number of harmonics passed.

Common- and Adjacent-Channel Interference

The simplest case of frequency modulation interference (that of two

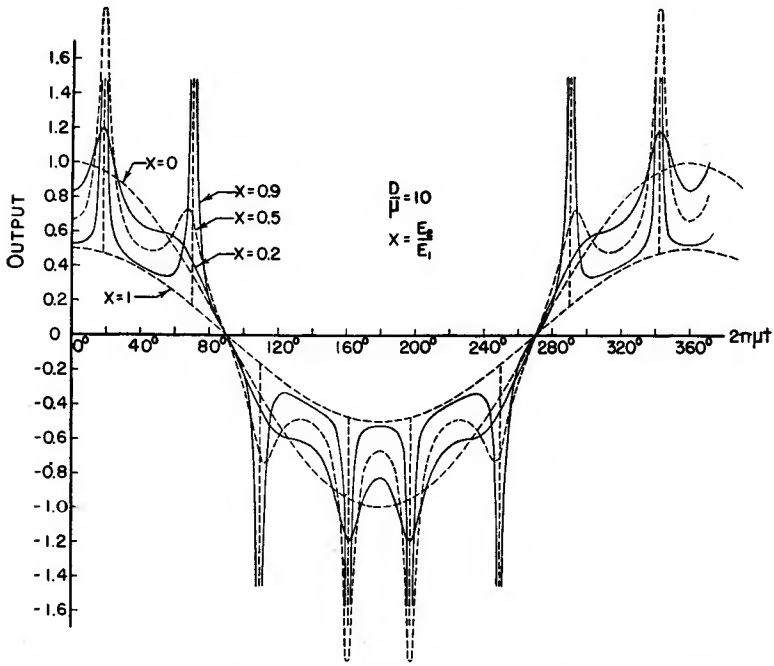


Fig. 14—Variation of distortion as interfering signal becomes stronger.

unmodulated carriers of slightly different frequency) has already been discussed. If now the amplitudes of the two waves are kept constant, but the frequency of one carrier is changed, the problem becomes one of common- or adjacent-channel interference depending upon what range of frequencies the swings of the modulated carrier cover. If the deviations of the one wave are about a mean frequency which coincides with the frequency of the second carrier, the result is common-channel interference. If the mean frequencies are separated by the width of one channel, the result will be adjacent-channel interference.

Common-Channel Interference, Interfering Signal Unmodulated. If a frequency-modulated signal and an unmodulated carrier produce the beat-note interference, the output from a frequency-modulation receiver with limiter will be as shown by Figure 14. This shows the wave form for the various ratios of the interfering signal voltage x. When $x = 0$ (i.e., no interference) the output is an undistorted cosine wave, as shown by the dotted line. As the interference increases, the peaks and dips increase in size, until finally, in the limit, they become very narrow pulses superimposed on a cosine wave of one-half the amplitude obtained with no interference.

As x becomes greater than one, the interfering signal takes control and

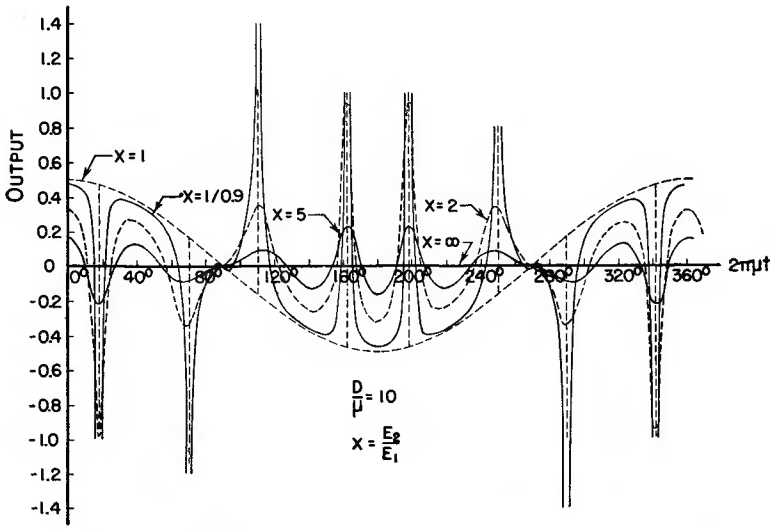


Fig. 15—Variation of distortion as interfering signal becomes stronger.

the modulation of the desired signal is suppressed. Figure 15 shows how the peaks and dips in output decrease when x increases from one to infinity. The envelope of the carrier amplitude corresponding to Figures 14 and 15 is shown by Figure 16. There is one cancellation or hole in the carrier amplitude corresponding to each peak or dip in the output, since the rapid phase change which occurs at cancellation produces the large frequency deviation. If the limiter is not able to maintain a constant voltage input to the discriminator, the amplitude variations of the carrier will cause a reduction in the peaks in the output.

Envelope of Beat-note Pattern. As shown in Appendix II, the beat-note produced in the output of a receiver with a perfect limiter during common-channel interference is a series of peaks and dips which are limited by the

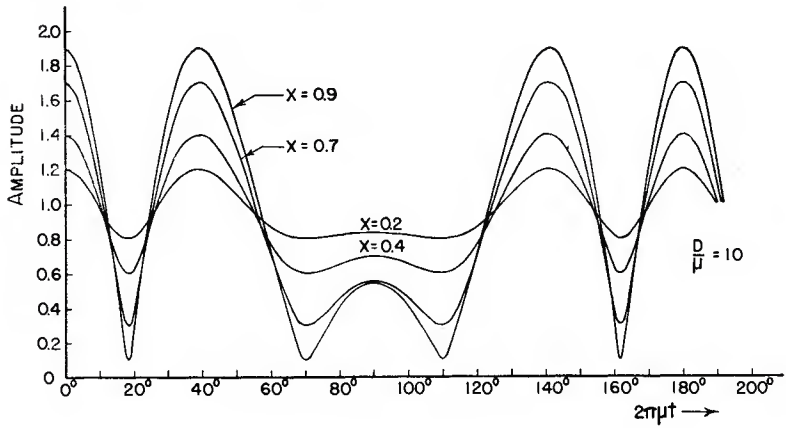


Fig. 16—Heterodyne envelope.

two curves $\frac{D}{1+x}\cos 2\pi\mu t$ and $\frac{D}{1-x}\cos 2\pi\mu t$.

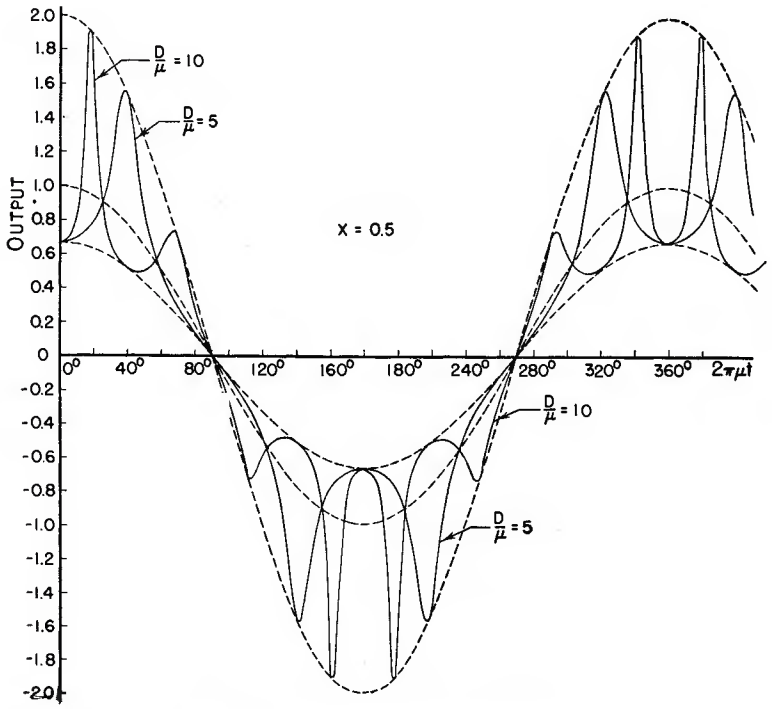


Fig. 17—Common-channel interference.

This will be true for all modulation indexes D/μ. The effect of increasing the modulation index is to produce more peaks and dips in output with no change in the limits. Figure 17 shows these two limits as dotted lines, for $x = 0.5$, and modulation indexes of 5 and 10. The output that would be obtained with no interference is also shown as a cosine wave of unit amplitude. Figure 18 shows how the number of peaks increases when the modulation index increases to 30. The limiting curves are the same as before. The two signals have the common center frequency at $2\pi\mu t = 90$

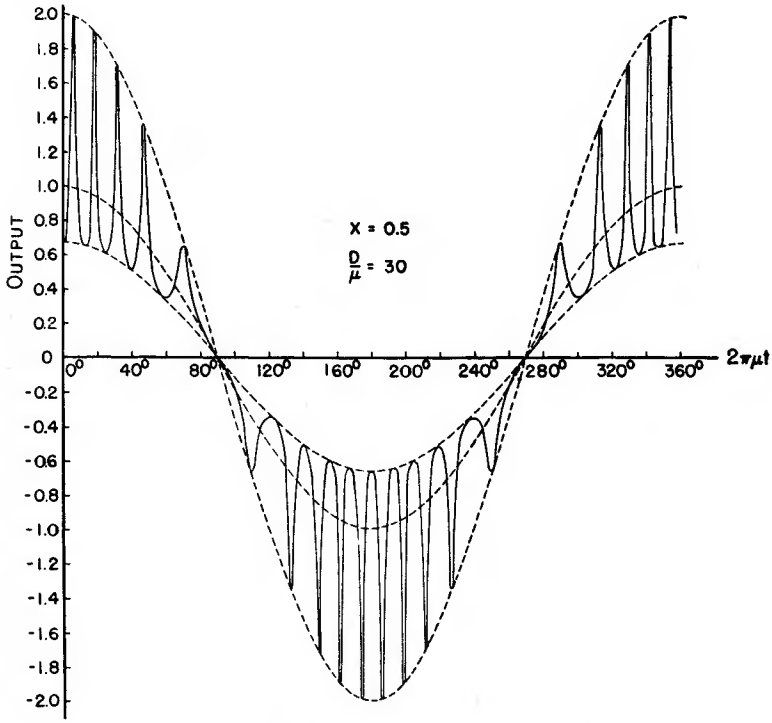


Fig. 18—Common-channel interference.

degrees and 270 degrees so they do not beat together there. As the modulated signal deviates toward the end of the swing, the frequency difference is large, and the peaks come more and more rapidly.

Effect of Detuning Interfering Signal. If the interfering signal is detuned by an amount equal to one-half the deviation of the desired signal, the effect is to move the frequency at which zero beat occurs to that point.

Figure 19 shows how the beat-note then becomes unsymmetrical. At one end of the swing the two signals have nearly the same frequency and the beats come slowly. At the other end of the swing there is a considerable frequency difference and the beats are very much more rapid. The peaks and dips are limited by the two curves:

$$\text{Envelope} = \frac{D}{1+x}\cos 2\pi\mu t + \frac{\alpha}{2\pi}\frac{x}{x+1} \tag{12}$$

and

$$\frac{D}{1-x}\cos 2\pi\mu t + \frac{\alpha}{2\pi}\frac{x}{x-1} \tag{13}$$

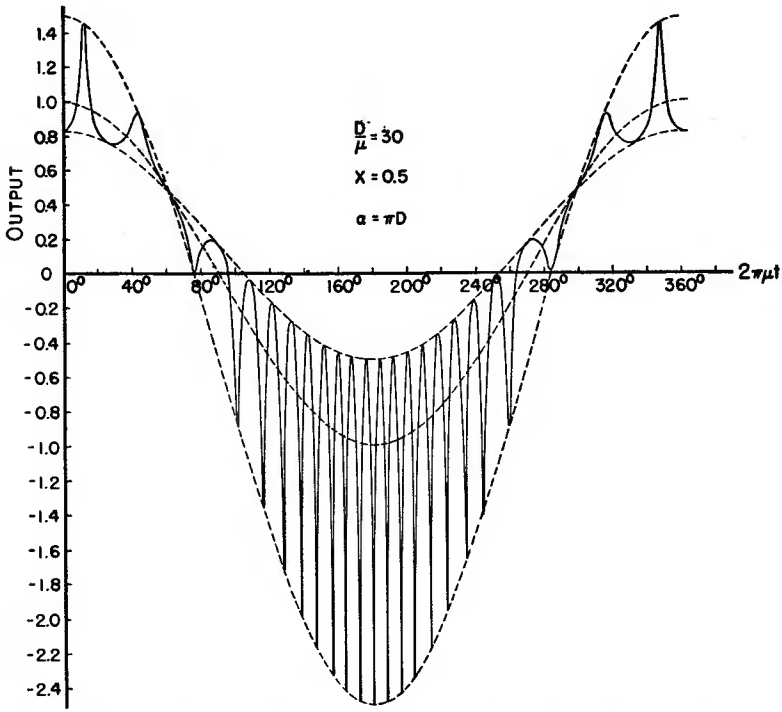


Fig. 19—Common-channel interference, interfering signal detuned.

Fourier-Series Analysis of Distorted Output. If the desired frequency-modulated signal is:

$$e_1 = E_1 \sin\left(\omega t + \frac{D}{\mu}\sin 2\pi\mu t\right) \tag{14}$$

and the interference is an unmodulated *r-f* carrier of angular frequency

$\omega + \alpha$, and phase angle θ, or:

$$e_2 = E_2 \sin \{(\omega + \alpha)t + \theta\} \tag{15}$$

then, as shown in Appendix II, the envelope of the resultant carrier is given by:

$$\text{Envelope} = E_1 \sqrt{1 + x^2 + 2x \cos \beta} \tag{16}$$

where: $$\beta = \frac{D}{\mu} \sin 2\pi\mu t - \alpha t - \theta$$

and the audio output is given by:

$$\text{Output} = D \cos 2\pi\mu t - \frac{D \cos 2\pi\mu t - \alpha/2\pi}{\dfrac{\cos \beta + 1/x}{\cos \beta + x}} + 1 \tag{17}$$

When this is expanded in a Fourier series to determine the harmonic and cross-modulation distortion, the audio output is given by:

$$\text{Output} = D \cos 2\pi\mu t$$

$$+ \sum_{n=1}^{\infty} \sum_{r=-\infty}^{\infty} (-x)^n \left\{ \frac{\mu r}{n} - \frac{\alpha}{2\pi} \right\} J_r(nD/\mu) \cos (r\varepsilon - n\alpha t - n\theta) \tag{18}$$

where $\varepsilon = 2\pi\mu t$, and $x<1$.

This shows that the effect of the interfering signal is to produce cross modulation between the desired signal modulated with audio frequency μ and the interfering unmodulated carrier of angular frequency $\omega + \alpha$. The amplitude of each cross-modulation frequency can be computed with the help of a table of Bessel functions of the first kind.

When $\alpha = 0$, (i.e., common-channel interference) the output becomes:

$$\text{Output} = D \cos 2\pi\mu t$$

$$+ 2\mu \sum_{r=1}^{\infty} (2r - 1)\, C(2r - 1, D/\mu; x, \theta) \cos \{(2r - 1)\, (2\pi\mu t)\}$$

$$+ 2\mu \sum_{r=1}^{\infty} (2r)\, S(2r, D/\mu; x, \theta) \sin \{(2r)\, (2\pi\mu t)\} \tag{19}$$

where the C- and S-functions are defined as follows:

$$C(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m(sn) \cos s\theta \tag{20}$$

$$S(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_m(sn) \sin s\theta \tag{21}$$

$$x^2 \leqq 1$$

To find the amplitudes of the various harmonics produced during common-channel interference, compute the value of the desired $C-$ or $S-$ function from equations 20 and 21, and multiply by the proper factor, which is shown by the above equation 19 for the audio output. A special table of Bessel functions has been prepared for this purpose.[17]

The effect of a de-emphasis network following the discriminator, and of a low-pass audio filter, can be determined by computing the amplitude of each harmonic that falls within the working range, correcting each one for amplitude and phase changes in the audio amplifier and filters, and then recombining them by superposition.

If the signal-noise ratio is defined as the desired audio output with no interfering carrier present, divided by the peak noise (i.e., the maximum departure from the desired audio output when no interference is present), then, as shown by Figures 17 and 18, the signal-noise ratio is independent of the modulation index, but depends only on the ratio of the two voltages, x. This assumes a perfect limiter, adequate band width in the amplifiers and discriminator, and linear-phase-shift circuits.

If a de-emphasis network and a low-pass audio filter are used, many of the harmonics will be attentuated or removed, and the nonlinear phase shift will prevent the remaining harmonics from coming into phase all at the same time. The peaks of noise are therefore reduced considerably. When the modulation index, D/μ, is large, the noise beat-note peaks come very rapidly. This means that the harmonics will be of high order and they will be reduced or removed by the audio selectivity. This accounts for the observed noise reduction with wide-band frequency modulation and shows that it is very important to use a de-emphasis network and low-pass filter.

Common-Channel Interference, Both Signals Modulated. The preceding cases have described the interference produced by an unmodulated carrier on the same channel as the desired signal, and the effect of detuning the interfering carrier. This section is a discussion of the case when both the desired and undesired signals are modulated sinusoidally, and of the resultant distortion, which is even more complicated.

In order to illustrate this form of interference, assume the following conditions:

$D_1/\mu_1 = 10$, $D_2/\mu_2 = 5$, $D_1 = 4D_2$, $\mu_1 = 2\mu_2$, $x = E_2/E_1$
For example, $D_1 = 60$ kc, $\mu_1 = 6$ kc, $D_2 = 15$ kc, $\mu_2 = 3$ kc, $x = 0.5$ and 0.9 could be one set of numerical values.

---

[17] Murlan S. Corrington and William Miehle, "Tables of Bessel Functions $J_n(x)$ for Large Arguments," *Jour. Math. Phys.*, Vol. 24, No. 1, pp. 30–50; Feb., 1945.

The beat-note envelope produced in this case is shown by Figure 20. The characteristic peaks and holes in the resultant carrier amplitude are present, but some of them are modified in shape because the two audio frequencies are present simultaneously.

Near 40 degrees and again near 130 degrees the two voltages start to go out-of-phase, but the two vectors then begin to reverse themselves and only a small decrease in amplitude occurs.

If this signal is sent through a receiver with a perfect limiter and linear discriminator, the resultant audio output will be as shown by Figure 21. Two cycles of the desired signal are shown as a dotted curve. This corresponds to one cycle of the undesired signal. As x increases toward one, the beat-note interference increases in amplitude until in the limit as $x \rightarrow 1$, the pulses become very narrow and long. If x becomes greater than one,

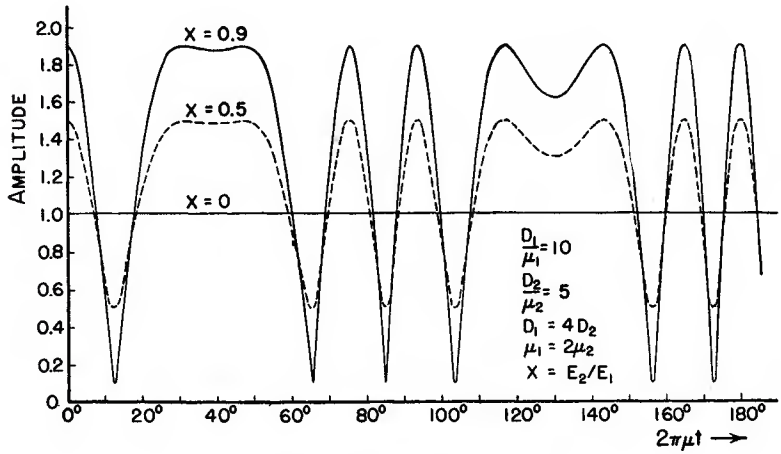


Fig. 20—Heterodyne envelope.

the polarity of the pulses is reversed (as shown by Figure 22) and this undesired signal gains control. When x becomes very large, only the undesired signal is received, as shown by the dotted cosine wave of unit amplitude.

The equations for the envelope and the beat-note interference are derived in Appendix III. If D_1 and D_2 are the two deviations and μ_1 and μ_2 are the corresponding audio frequencies, the envelope of the carrier is:

$$\text{Envelope} = E_1 \sqrt{1 + x^2 + 2x \cos \{D_1/\mu_1 \sin 2\pi\mu_1 t - D_2/\mu_2 \sin 2\pi\mu_2 t\}} \tag{22}$$

and the audio output from a receiver with limiter and balanced discriminator is:

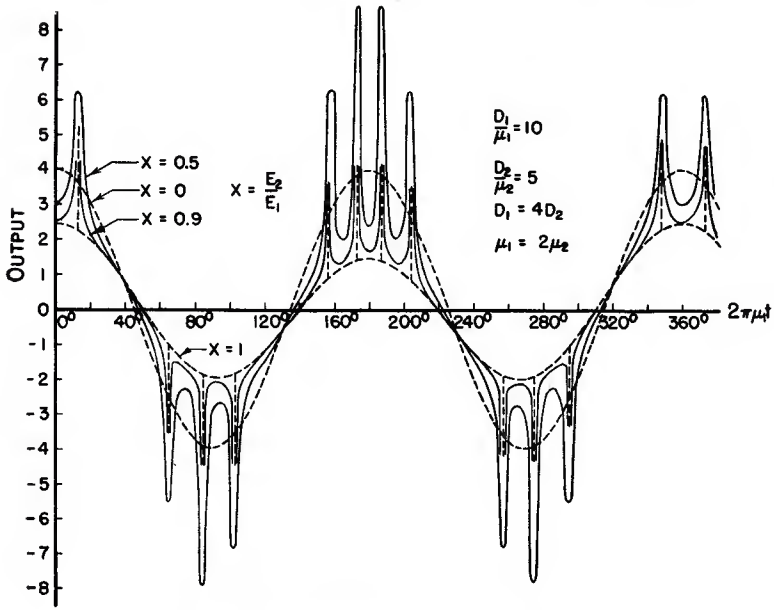


Fig. 21—Beat-note interference, $x < 1$.

$$\text{Output} \propto D_1 \cos 2\pi\mu_1 t - \frac{D_1 \cos 2\pi\mu_1 t - D_2 \cos 2\pi\mu_2 t}{\dfrac{\cos\theta + 1/x}{\cos\theta + x} + 1} \tag{23}$$

where

$$\theta = \frac{D_1}{\mu_1} \sin 2\pi\mu_1 t - \frac{D_2}{\mu_2} \sin 2\pi\mu_2 t. \tag{24}$$

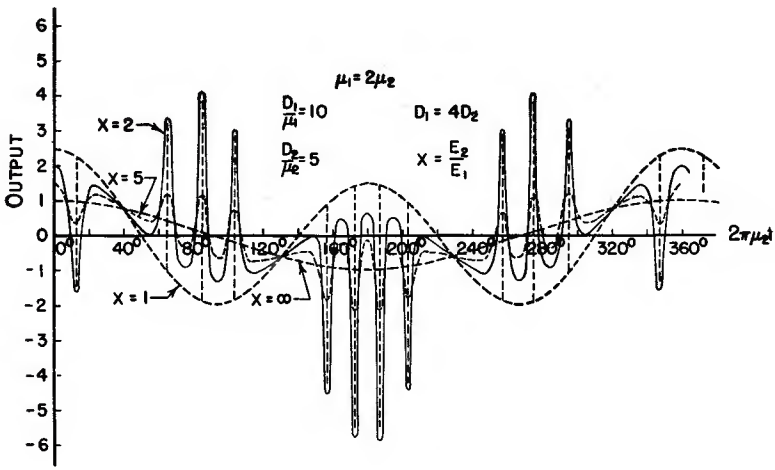


Fig. 22—Beat-note interference, $x > 1$.

The audio output is composed of a beat-note pattern which is limited by the two envelopes:

$$\text{Envelope} = \frac{D_1}{1+x}\cos 2\pi\mu_1 t + \frac{D_2 x}{1+x}\cos 2\pi\mu_2 t \tag{25}$$

and

$$\frac{D_1}{1-x}\cos 2\pi\mu_1 t + \frac{D_2 x}{x-1}\cos 2\pi\mu_2 t \tag{26}$$

This effect is shown by Figure 23 for the set of values given. In case of imperfect limiting, limited band width, or nonlinear phase shift in the amplifiers, these peaks will not be so long and narrow; the two envelopes shown represent the limits of the distortion.

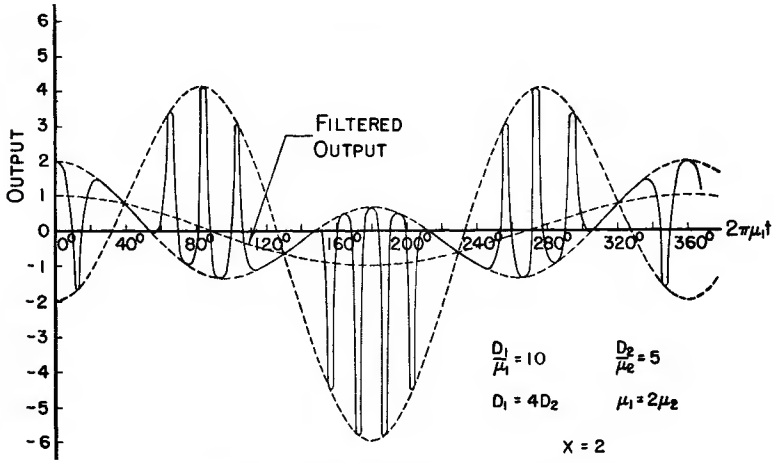


Fig. 23—Envelope of beat-note pattern.

The effect of low-pass filters or other audio selectivity can be determined from a study of the harmonic content of the distortion. As shown in Appendix III, the audio output can be expressed as a Fourier series which gives the cross modulation terms produced and their amplitudes.

Thus: Output $\propto D_1 \cos 2\pi\mu_1 t$

$$+ \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} (r\mu_1 - s\mu_2)\, C(r, D_1/\mu_1; s, D_2/\mu_2; x, 0)\cos(r\alpha - s\beta) \tag{27}$$

where $\alpha = 2\pi\mu_1 t$, $\beta = 2\pi\mu_2 t$ and the generalized C-function is defined as:

$$C(k, l; m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_k(sl)\, J_m(sn)\cos s\theta. \tag{28}$$

The amplitude of any desired combination tone can be determined by choosing the appropriate values of r and s and by computing the desired C-function. Since the C-function cannot exceed unity for a given combination tone, it is evident that if $D_1 \gg \mu_1$ or μ_2 the distortion will be reduced with increasing modulation index.

Conclusions

Frequency-modulated radio broadcasting offers the advantage of improved noise reduction when compared with the usual amplitude-modulation systems. There is less interference between stations operating on the same frequency than for the corresponding case of amplitude modulation, and less power is required to cover a given area.

A difficulty arose occasionally in the 42–50 megacycle frequency modulation band because long-distance transmission could be observed from frequency modulation broadcast stations during hot weather or before a storm. It sometimes happened that such an interfering station became stronger than a desired station in the same channel for short intervals. When this happened there was a small amount of noise and the programs suddenly were interchanged. This change often lasted for several seconds but sometimes was limited to a word or two or a few notes of music. If the proposed new frequency modulation stations are all completed, this interference may occur again. When the interfering station has nearly the same carrier frequency as the desired station this effect is called common-channel interference. If the two carrier frequencies are separated by the width of one channel the result is called adjacent-channel interference.

The simplest case of frequency modulation interference occurs when two modulated carriers, having nearly the same frequency, beat together to produce a resultant signal. As the two voltages alternately reinforce and cancel each other, the result is a heterodyne envelope consisting of a series of broad peaks and sharp dips. Each time the two interfering voltages cancel each other to produce a hole in the envelope, there is a rapid phase shift of the resultant voltage. Since the audio output from a frequency-modulation receiver is proportional to the rate of change of the phase of this resultant, the rapid phase shift produces a distorted audio output, which becomes more and more like an impulse as the interfering carrier voltage becomes nearly equal to the desired carrier voltage.

When the two amplitudes of the interfering voltages are kept constant but the frequency of one is changed, the result is common- or adjacent-channel interference depending upon what range of frequencies the swings of the modulated carrier cover. The beat-note produced by this interference consists of a series of sharp peaks and dips of noise and is super-

imposed on the desired audio output. When the modulation index is increased, these peaks occur more and more rapidly, and the harmonics produced are redistributed to higher and higher orders. If the receiver has sufficient band width, a perfect limiter, and a wide-band audio system, the signal-noise ratio does not depend on the modulation index, but is determined solely by the ratio of the desired signal voltage to the undesired signal voltage.

Formulas are given for computing the amplitudes of the harmonics and cross-modulation frequencies produced by the interference. The effect of a de-emphasis network following the discriminator, of a low-pass audio filter, and of nonlinear phase-shift can be determined by computing the amplitude of each harmonic that falls within the working range. Each such harmonic is then corrected for amplitude and phase changes in the audio amplifier and filters, and they are then recombined by superposition to obtain the filtered audio output.

When the modulation index is large, the beat-notes of the noise come very rapidly, and since this means that the harmonics are then of high order, most of the distortion will be removed by the audio selectivity. This accounts for the observed noise reduction with wide-band frequency modulation and shows that it is very important that a de-emphasis network and low-pass filter be used to obtain maximum performance. In order to obtain the maximum signal-noise ratio, it is necessary to use some means for removing the variations in the amplitude of the resultant signal so that the discriminator responds to the variations in the instantaneous frequency, but is not affected by amplitude variations.

* * *

APPENDIX I.

Analysis of Fundamental Case

Let the two interfering signals be $e_1 \sin \omega t$ and $e_2 \sin (\omega + 2\pi\mu)t$ The resultant voltage is then:

$$e_1 \sin \omega t + e_2 \sin (\omega + 2\pi\mu)t$$

$$= e_1\sqrt{1 + x^2 + 2x \cos 2\pi\mu t} \sin (\omega t + \varphi) \qquad (29)$$

where $e_2/e_1 = x$ and $\tan \varphi = \dfrac{x \sin 2\pi\mu t}{1 + x \cos 2\pi\mu t}$.

The instantaneous frequency becomes:

$$f = \frac{1}{2\pi}\frac{d}{dt}(\omega t + \varphi) = \frac{\omega}{2\pi} + \frac{1}{2\pi}\frac{d}{dt}\tan^{-1}\frac{x \sin 2\pi\mu t}{1 + x\cos 2\pi\mu t}$$

$$= \frac{\omega}{2\pi} + \mu\frac{x\cos 2\pi\mu t + x^2}{1 + x^2 + 2x\cos 2\pi\mu t}$$

$$= \frac{\omega}{2\pi} + \frac{\mu}{\dfrac{\cos 2\pi\mu t + 1/x}{\cos 2\pi\mu t + x} + 1} \tag{30}$$

This is valid for all values of x.

For a balanced linear discriminator, the audio output is proportional to:

$$\text{Output} \propto \frac{\mu}{\dfrac{\cos 2\pi\mu t + 1/x}{\cos 2\pi\mu t + x} + 1} \tag{31}$$

When $x \ll 1$, this is, approximately,

$$\text{Output} \propto \mu x \cos 2\pi\mu t \tag{32}$$

The instantaneous frequency can be written:

$$f = \frac{\omega}{2\pi} + \mu - \mu\frac{\dfrac{1}{x}\cos 2\pi\mu t + \dfrac{1}{x^2}}{1 + \dfrac{1}{x^2} + \dfrac{2}{x}\cos 2\pi\mu t} \tag{33}$$

This means that as x goes from less than one to greater than one (i.e., if x is changed to $1/x$) there is an apparent change in frequency equal to μ and a reversal in polarity of the modulation. This means that e_2 becomes stronger than e_1 and takes control.

Average Voltage of Rectified Envelope

The average voltage of the carrier envelope is:

$$\text{Average voltage} = \frac{1}{\pi}\int_0^{\pi} e_1\sqrt{1 + x^2 + 2x\cos\theta}\, d\theta$$

$$= \frac{2(1+x)\ e_1}{\pi}\int_0^{\pi/2}\sqrt{1 - \frac{4x}{(1+x)^2}\sin^2\alpha}\, d\alpha$$

$$= \frac{2(1+x)e_1}{\pi} E\left\{\frac{2\sqrt{x}}{1+x}\right\} \tag{34}$$

where $E\left\{\frac{2\sqrt{x}}{1+x}\right\}$ is a complete elliptic integral of the second kind with modulus $\frac{2\sqrt{x}}{1+x}$.

Root-Mean-Square Voltage of Rectified Envelope

The Root-Mean-Square voltage of the rectified carrier envelope is:

$$\text{Root-Mean-Square voltage} = e_1\sqrt{\frac{1}{\pi}\int_0^{\pi}(1+x^2+2x\cos\theta)\,d\theta}$$

$$= e_1\sqrt{\frac{1}{\pi}\Big[(1+x^2)\theta+2x\sin\theta\Big]_0^{\pi}}$$

$$= e_1\sqrt{1+x^2} \qquad (35)$$

Fourier-Series Analysis of Envelope

The envelope of the carrier is given by:

$$\text{Envelope} = e_1\sqrt{1+x^2+2x\cos 2\pi\mu t} \qquad \text{where } x \leqq 1. \qquad (36)$$

Consider the expression:

$$\sqrt{1+x^2+2x\cos\beta} = (1+xe^{i\beta})^{\frac{1}{2}}(1+xe^{-i\beta})^{\frac{1}{2}}$$

$$= \left\{1+\frac{1}{2}xe^{i\beta}-\frac{1(1)}{2(4)}x^2e^{2i\beta}+\frac{1(1)\ (3)}{2(4)\ (6)}x^3e^{3i\beta}-\frac{1(1)\ (3)\ (5)}{2(4)\ (6)\ (8)}x^4e^{4i\beta}+\ldots\right\}$$

$$\times\left\{1+\frac{1}{2}xe^{-i\beta}-\frac{1(1)}{2(4)}x^2e^{-2i\beta}+\frac{1(1)\ (3)}{2(4)\ (6)}x^3e^{-3i\beta}-\frac{1(1)\ (3)\ (5)}{2(4)\ (6)\ (8)}x^4e^{-4i\beta}+\ldots\right\} \qquad (37)$$

by the usual binomial series expansion.[18]

Multiply these factors together, term by term, then:

$$\sqrt{1+x^2+2x\cos\beta} = a_0+a_1\cos\beta+a_2\cos 2\beta+\ldots \qquad (38)$$

where:

$$a_0 = 1+\frac{x^2}{4}+\frac{x^4}{64}+\frac{x^6}{256}+\frac{25x^8}{16384}+\ldots \qquad (39)$$

$$a_1 = x\left\{1-\frac{x^2}{8}-\frac{x^4}{64}-\frac{5x^6}{1024}-\frac{35x^8}{16384}-\ldots\right\} \qquad (40)$$

[18] Edwin P. Adams, SMITHSONIAN MATHEMATICAL FORMULAE AND TABLES OF ELLIPTIC FUNCTIONS, Smithsonian Institution, Washington, D. C., 1939, p. 117.

$$a_2 = -\frac{x^2}{4}\left\{1 - \frac{x^2}{4} - \frac{5x^4}{128} - \frac{7x^6}{512} - \frac{105x^8}{16384} - \ldots\right\} \tag{41}$$

$$a_3 = \frac{x^3}{8}\left\{1 - \frac{5x^2}{16} - \frac{7x^4}{128} - \frac{21x^6}{1024} - \frac{165x^8}{16384} - \ldots\right\} \tag{42}$$

$$a_4 = -\frac{5x^4}{64}\left\{1 - \frac{7x^2}{20} - \frac{21x^4}{320} - \frac{33x^6}{1280} - \frac{429x^8}{32768} - \ldots\right\} \tag{43}$$

$$a_n = 2(-1)^n \left\{\frac{1(3)\ \ldots\ (2n-1)}{n!}\right\} \frac{x^n}{2^n}\left\{\frac{-1}{2n-1} + \frac{1}{1(n+1)}\frac{x^2}{2^2}\right.$$

$$\left. + \sum_{k=2}^{\infty} \frac{1(3)\ \ldots\ (2k-3)}{k!\ 2^{2k}} \frac{(2n+1)\ (2n+3)\ \ldots\ (2n+2k-3)}{(n+1)\ \ (n+2)\ \ldots\ (n+k)} x^{2k}\right\} \tag{44}$$

This expression for a_n was previously obtained by Vigoureux[19] and Moullin[20]

In the limit as $x \to 1$:

$$\text{Envelope} = \sqrt{2}\, e_1 \sqrt{1 + \cos 2\pi\mu t} = 2e_1 \left|\cos \pi\mu t\right|$$

$$= \frac{4e_1}{\pi}\left\{1 + \frac{2}{3}\cos\theta - \frac{2}{15}\cos 2\theta\right.$$

$$\left. + \frac{2}{35}\cos 3\theta - \frac{2}{63}\cos 4\theta - \ldots\right\}$$

$$= \frac{4e_1}{\pi}\left\{1 - 2\sum_{n=1}^{\infty} \frac{(-1)^n \cos n\theta}{(2n)^2 - 1}\right\}$$

where $\theta = 2\pi\mu t$. (45)

Calculation of Average Value of Instantaneous Frequency

Consider the integral:

$$I = \frac{\mu}{\pi}\int_0^{\pi} \frac{x^2 + x\cos\varepsilon}{1 + x^2 + 2x\cos\varepsilon}\, d\varepsilon \tag{46}$$

Make the transformation: $\cos\varepsilon = \dfrac{1 - t^2}{1 + t^2}$, $d\varepsilon = \dfrac{2dt}{1 + t^2}$

[19] F. M. Colebrook, "A Note on the Frequency Analysis of the Heterodyne Envelope. Its Relation to Problems of Interference." *Wireless Engineer & Experimental Wireless*, Vol. 9, p. 200, April, 1932.

[20] E. B. Moullin, "The Detection by a Straight Line Rectifier of Modulated and Heterodyne Signals," *Wireless Engineer & Experimental Wireless*, Vol. 9, pp. 378–383; July, 1932.

Then.

$$I = \frac{2x\,\mu}{\pi}\int_0^\infty \frac{(1+x)-(1-x)t^2}{(1+x)^2+(1-x)^2t^2}\,\frac{dt}{1+t^2}$$

$$= \frac{(x^2-1)\mu}{\pi}\int_0^\infty \frac{dt}{(1+x)^2+(1-x)^2t^2} + \frac{\mu}{\pi}\int_0^\infty \frac{dt}{1+t^2}$$

$$= \frac{(x^2-1)\mu}{\pi}\left\{\frac{\pi}{2(1-x^2)}\right\} + \frac{\mu}{\pi}\left(\frac{\pi}{2}\right) = 0 \text{ when } x<1$$

$$= \mu \text{ when } x > 1. \qquad (47)$$

The average value of the instantaneous frequency therefore equals zero when $x<1$ and is proportional to μ when $x>1$.

Calculation of the Root-Mean-Square Value of Instantaneous Frequency

Consider the integral:

$$I = \frac{1}{\pi}\int_0^\pi \left\{\frac{x^2+x\cos\varepsilon}{1+x^2+2x\cos\varepsilon}\right\}^2 d\varepsilon \qquad (48)$$

Make the transformation: $\cos\varepsilon = \dfrac{1-t^2}{1+t^2}$, $d\varepsilon = \dfrac{2\,dt}{1+t^2}$.

Then:

$$I = \frac{2x^2}{\pi}\int_0^\infty \left\{\frac{(1+x)-(1-x)t^2}{(1+x)^2+(1-x)^2t^2}\right\}^2 \frac{dt}{1+t^2}$$

$$= \frac{-2x(1+x)^2}{\pi}\int_0^\infty \frac{dt}{\{(1+x)^2+(1-x)^2t^2\}^2}$$

$$+ \frac{(3x-1)(1+x)}{2\pi}\int_0^\infty \frac{dt}{\{(1+x)^2+(1-x)^2t^2\}} + \frac{1}{2\pi}\int_0^\infty \frac{dt}{1+t^2} \qquad (49)$$

Consider the integral:

$$\int_0^\infty \frac{dt}{\{a^2+b^2t^2\}^2} = \frac{1}{2a^2}\left[\frac{t}{a^2+b^2t^2}\right]_0^\infty + \frac{1}{2a^2}\int_0^\infty \frac{dt}{a^2+b^2t^2}$$

$$= \frac{1}{2a^2}\int_0^\infty \frac{dt}{a^2+b^2t^2} = \frac{1}{2a^3b}\left[\tan^{-1}\frac{b}{a}\,t\right]_0^\infty = \frac{\pi}{4a^3b} \text{ when } x<1 \qquad (50)$$

$$= -\frac{\pi}{4a^3b} \text{ when } x > 1. \qquad (51)$$

Therefore I becomes:

$$I = \frac{-2x(1+x)^2}{\pi}\left\{\frac{\pi}{4(1+x)^3(1-x)}\right\} + \frac{(3x-1)(1+x)}{2\pi}\left\{\frac{\pi}{2(1+x)(1-x)}\right\} + \frac{1}{2\pi}\left\{\frac{\pi}{2}\right\}$$

$$= \frac{-x}{2(1-x^2)} + \frac{3x-1}{4(1-x)} + \frac{1}{4} = \frac{x^2}{2(1-x^2)} \text{ when } x < 1, \tag{52}$$

$$= \frac{1-2x^2}{2(1-x^2)} \text{ when } x > 1. \tag{53}$$

The root-mean-square voltage is proportional to:

$$\mu\sqrt{I} = \frac{x\mu}{\sqrt{2(1-x^2)}} \text{ when } x < 1 \tag{54}$$

$$= \sqrt{\frac{2x^2-1}{2(x^2-1)}}\,\mu \text{ when } x > 1. \tag{55}$$

Calculation of the Area of One Cycle of the Instantaneous Frequency

The area bounded by one cycle of the variation of the instantaneous frequency and a line one-half unit above the time axis, as shown by Figure 7, will now be computed. From equation 2 the instantaneous frequency is given by:

$$f = \frac{\omega}{2\pi} + \frac{1}{2\pi}\frac{d}{dt}\tan^{-1}\frac{x\sin 2\pi\mu t}{1+x\cos 2\pi\mu t} \tag{56}$$

$$\text{Area} = 2\int_0^{\pi}\left\{\frac{1}{2\pi}\frac{d}{dt}\tan^{-1}\frac{x\sin 2\pi\mu t}{1+x\cos 2\pi\mu t} - \frac{1}{2}\right\} d(2\pi\mu t)$$

$$= 2\mu\int_0^{\pi}\left\{\frac{d}{d\theta}\tan^{-1}\frac{x\sin\theta}{1+x\cos\theta} - \frac{1}{2}\right\} d\theta$$

$$= 2\left[\tan^{-1}\frac{x\sin\theta}{1+x\cos\theta} - \frac{\theta}{2}\right]_0^{\pi} = -\pi\mu \tag{57}$$

for all values of x.

Fourier-Series Analysis of Instantaneous Frequency

The audio output is proportional to:

$$\mu\frac{x\cos 2\pi\mu t + x^2}{1+x^2+2x\cos 2\pi\mu t} = \frac{1}{2\pi}\frac{d}{dt}\tan^{-1}\frac{x\sin 2\pi\mu t}{1+x\cos 2\pi\mu t} \tag{58}$$

Let $2\pi\mu t = \beta$ and $\tan \alpha = \dfrac{x \sin \beta}{1 + x \cos \beta}$

Then $k \sin \alpha = x \sin \beta$ (59)

$k \cos \alpha = 1 + x \cos \beta$ (60)

where: $k = \sqrt{1 + x^2 + 2x \cos \beta}$ (61)

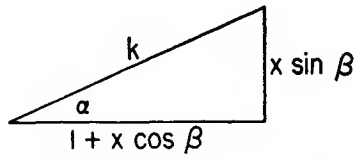


Fig. 24—Determination of k.

Multiply equation 59 by i and add equation 60.

Then: $1 + x \cos \beta + ix \sin \beta = k (\cos \alpha + i \sin \alpha)$ (62)

or: $1 + x e^{i\beta} = k e^{i\alpha}$. Take logarithms of both sides. Then:

$\log (1 + x e^{i\beta}) = \log k + i\alpha.$

Since:

$$\log (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \qquad -1 < x \leqq 1$$

$$\log (1 + x e^{i\beta}) = x e^{i\beta} - \frac{x^2}{2} e^{2i\beta} + \frac{x^3}{3} e^{3i\beta} - \ldots$$

so:

$$\log k + i\alpha = x(\cos \beta + i \sin \beta) - \frac{x^2}{2} (\cos 2\beta + i \sin 2\beta)$$

$$+ \frac{x^3}{3} (\cos 3\beta + i \sin 3\beta) - \ldots \tag{63}$$

Equate imaginary terms:

$$\alpha = x \sin \beta - \frac{x^2}{2} \sin 2\beta + \frac{x^3}{3} \sin 3\beta - \ldots \tag{64}$$

Differentiate:

$$\frac{1}{2\pi} \frac{d\alpha}{dt} = \mu (x \cos \beta - x^2 \cos 2\beta + x^3 \cos 3\beta - \ldots) \tag{65}$$

The audio output is therefore proportional to:

$$\text{Output} \propto -\mu \sum_{n=1}^{\infty} (-x)^n \cos n\beta \qquad -1 < x \leqq 1 \tag{66}$$

When $\beta = 0$,

$$\text{Output} \propto -\mu \sum_{n=1}^{\infty} (-x)^n = \frac{\mu x}{x+1} \tag{67}$$

When $\beta = \pi$, $$\text{Output} \propto -\mu \sum_{n=1}^{\infty} x^n = \frac{\mu x}{x-1} \tag{68}$$

Effect of Limited Band Width

To show the effect of a limited band width, consider the geometrical progression:

$$S = -\mu \sum_{n=1}^{n} x^n \tag{69}$$

By ordinary long division:

$$\frac{p^n - 1}{p - 1} = 1 + p + p^2 + \ldots + p^{n-1} \tag{70}$$

so,

$$S = -\mu x \frac{x^n - 1}{x - 1} \tag{71}$$

The ratio of the partial sum to the output at $\beta = \pi$ equals $1 - x^n$.

APPENDIX II.

Common- and Adjacent-Channel Interference

In order to show the effect of common- and adjacent-channel interference, let the desired frequency-modulated signal be:

$$e_1 = E_1 \sin \left(\omega t + \frac{D}{\mu} \sin 2\pi\mu t\right) \tag{72}$$

and let the interference be an unmodulated radio-frequency carrier at angular frequency $\omega + \alpha$, and phase angle θ, or

$$e_2 = E_2 \sin \{(\omega + \alpha)t + \theta\} \tag{73}$$

Then:

$$e_1 + e_2 = E_1\sqrt{1 + x^2 + 2x\cos\beta}\,\sin\left\{\omega t + \frac{D}{\mu}\sin 2\pi\mu t - \varphi\right\} \quad (74)$$

where $x = E_2/E_1$, $\beta = \dfrac{D}{\mu}\sin 2\pi\mu t - \alpha t - \theta$

and $\tan\varphi = \dfrac{x\sin\beta}{1 + x\cos\beta}$

The instantaneous frequency becomes:

$$\frac{\omega}{2\pi} + D\cos 2\pi\mu t - \frac{1}{2\pi}\frac{d}{dt}\tan^{-1}\frac{x\sin\beta}{1 + x\cos\beta}$$

$$= \frac{\omega}{2\pi} + D\cos 2\pi\mu t - \frac{x\cos\beta + x^2}{1 + x^2 + 2x\cos\beta}\left\{D\cos 2\pi\mu t - \frac{\alpha}{2\pi}\right\}$$

$$= \frac{\omega}{2\pi} + D\cos 2\pi\mu t - \frac{D\cos 2\pi\mu t - \alpha/2\pi}{\dfrac{\cos\beta + 1/x}{\cos\beta + x} + 1} \quad (75)$$

Envelope of Beatnote Pattern

The beatnote produced in the output of a receiver with a perfect limiter is given by:

$$\text{Output} = D\cos 2\pi\mu t - \frac{D\cos 2\pi\mu t - \alpha/2\pi}{\dfrac{\cos\beta + 1/x}{\cos\beta + x} + 1} \quad (76)$$

where: $\beta = \dfrac{D}{\mu}\sin 2\pi\mu t - \alpha t - \theta$. The two envelopes of the maxima and minima of the beat-note pattern are obtained by setting $\beta = 2n\pi$ or $(2n + 1)\pi$ where n is an integer. This gives the two envelopes:

$$\text{Envelope} = \frac{D}{1 + x}\cos 2\pi\mu t + \frac{\alpha}{2\pi}\frac{x}{x + 1} \quad (77)$$

and:

$$\frac{D}{1 - x}\cos 2\pi\mu t + \frac{\alpha}{2\pi}\frac{x}{x - 1} \quad (78)$$

Fourier-Series Analysis of Instantaneous Frequency

In accordance with the analysis of Appendix I, equation 65:

$$\frac{d}{dt}\tan^{-1}\frac{x\sin\beta}{1+x\cos\beta} = -\sum_{n=1}^{\infty}(-x)^n\cos n\beta\,\frac{d\beta}{dt} \tag{79}$$

When: $\beta = \frac{D}{\mu}\sin 2\pi\mu t - \alpha t - \theta$ and $\varepsilon = 2\pi\mu t$

the instantaneous frequency is:

$$f=\frac{\omega}{2\pi} + D\cos\varepsilon + \sum_{n=1}^{\infty}(-x)^n\cos n\left\{\frac{D}{\mu}\sin\varepsilon - \alpha t - \theta\right\}\left\{D\cos\varepsilon - \frac{\alpha}{2\pi}\right\} \tag{80}$$

Let: $\gamma = D/\mu$, then

$$f = \frac{\omega}{2\pi} + D\cos\varepsilon$$

$$+\frac{1}{4}\sum_{n=1}^{\infty}(-x)^n\left\{D(e^{i\varepsilon}+e^{-i\varepsilon}) - \frac{\alpha}{\pi}\right\}\left\{e^{in\gamma\sin\varepsilon - in\alpha t - in\theta} + e^{-in\gamma\sin\varepsilon + in\alpha t + in\theta}\right\}$$

$$= \frac{\omega}{2\pi} + D\cos\varepsilon$$

$$+\frac{1}{4}\sum_{n=1}^{\infty}(-x)^n\left\{D\,e^{i(\varepsilon - n\alpha t - n\theta)}\,e^{in\gamma\sin\varepsilon} + D\,e^{i(\varepsilon + n\alpha t + n\theta)}\,e^{-in\gamma\sin\varepsilon}\right.$$

$$+ D\,e^{-i(\varepsilon + n\alpha t + n\theta)}\,e^{in\gamma\sin\varepsilon} + D\,e^{-i(\varepsilon - n\alpha t - n\theta)}\,e^{-in\gamma\sin\varepsilon}$$

$$\left. - \frac{\alpha}{\pi}e^{-i(n\alpha t + n\theta)}\,e^{in\gamma\sin\varepsilon} - \frac{\alpha}{\pi}e^{i(n\alpha t + n\theta)}\,e^{-in\gamma\sin\varepsilon}\right\} \tag{81}$$

Using the identities:

$$e^{ix\sin\varepsilon} = \sum_{k=-\infty}^{\infty} J_k(x)\,e^{ik\varepsilon} \tag{82}$$

and $J_k(-x) = J_{-k}(x)$ where $J_k(x)$ is a Bessel function of the first kind of order k and argument x, the instantaneous frequency becomes:

$$f = \frac{\omega}{2\pi} + D\cos\varepsilon$$

$$+\frac{1}{4}\sum_{n=1}^{\infty}(-x)^n\left\{D\,e^{i(\varepsilon - n\alpha t - n\theta)}\sum_{k=-\infty}^{\infty}J_k(n\gamma)\,e^{ik\varepsilon}\right.$$

$$+ D\, e^{i(\varepsilon + n\alpha t + n\theta)} \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{ik\varepsilon}$$

$$+ D\, e^{-i(\varepsilon + n\alpha t + n\theta)} \sum_{k=-\infty}^{\infty} J_{k}(n\gamma)\, e^{ik\varepsilon}$$

$$+ D\, e^{-i(\varepsilon - n\alpha t - n\theta)} \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{ik\varepsilon}$$

$$- \frac{\alpha}{\pi} e^{-i(n\alpha t + n\theta)} \sum_{k=-\infty}^{\infty} J_{k}(n\gamma)\, e^{ik\varepsilon}$$

$$- \frac{\alpha}{\pi} e^{i(n\alpha t + n\theta)} \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{ik\varepsilon} \Bigg\}$$

$$= \frac{\omega}{2\pi} + D \cos \varepsilon$$

$$+ \frac{1}{4} \sum_{n=1}^{\infty} (-x)^n \Bigg\{ D \sum_{k=-\infty}^{\infty} J_{k}(n\gamma)\, e^{i\{(k+1)\varepsilon - n\alpha t - n\theta\}}$$

$$+ D \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{i\{(k+1)\varepsilon + n\alpha t + n\theta\}}$$

$$+ D \sum_{k=-\infty}^{\infty} J_{k}(n\gamma)\, e^{i\{(k-1)\varepsilon - n\alpha t - n\theta\}}$$

$$+ D \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{i\{(k-1)\varepsilon + n\alpha t + n\theta\}}$$

$$- \frac{\alpha}{\pi} \sum_{k=-\infty}^{\infty} J_{k}(n\gamma)\, e^{i\{k\varepsilon - n\alpha t - n\theta\}}$$

$$- \frac{\alpha}{\pi} \sum_{k=-\infty}^{\infty} J_{-k}(n\gamma)\, e^{i\{k\varepsilon + n\alpha t + n\theta\}} \Bigg\} \qquad (83)$$

Make the substitutions:

$k + 1 = r$ in term 1	$k - 1 = -r$ in term 4
$k + 1 = -r$ in term 2	$k = r$ in term 5
$k - 1 = r$ in term 3	$k = -r$ in term 6

Then:

$$f = \frac{\omega}{2\pi} + D \cos \varepsilon$$

$$+\frac{1}{4}\sum_{n=1}^{\infty}(-x)^n\left\{D\sum_{r=-\infty}^{\infty}\left[J_{r-1}(n\gamma)+J_{r+1}(n\gamma)\right]e^{i(r\varepsilon-n\alpha t-n\theta)}\right.$$

$$+D\sum_{r=-\infty}^{\infty}\left[J_{r-1}(n\gamma)+J_{r+1}(n\gamma)\right]e^{-i(r\varepsilon-n\alpha t-n\theta)}$$

$$\left.-\frac{\alpha}{\pi}\sum_{r=-\infty}^{\infty}J_r(n\gamma)\left[e^{i(r\varepsilon-n\alpha t-n\theta)}+e^{-i(r\varepsilon-n\alpha t-n\theta)}\right]\right\} \tag{84}$$

Apply the relation:

$$J_{r-1}(n\gamma)+J_{r+1}(n\gamma)=\frac{2r}{n\gamma}J_r(n\gamma) \tag{85}$$

Then:

$$f=\frac{\omega}{2\pi}+D\cos\varepsilon$$

$$+\sum_{n=1}^{\infty}(-x)^n\left\{D\sum_{r=-\infty}^{\infty}\frac{r}{n\gamma}J_r(n\gamma)\cos(r\varepsilon-n\alpha t-n\theta)\right.$$

$$\left.-\frac{\alpha}{2\pi}\sum_{r=-\infty}^{\infty}J_r(n\gamma)\cos(r\varepsilon-n\alpha t-n\theta)\right\} \tag{86}$$

$$=\frac{\omega}{2\pi}+D\cos 2\pi\mu t$$

$$+\sum_{n=1}^{\infty}\sum_{r=-\infty}^{\infty}(-x)^n\left\{\frac{\mu r}{n}-\frac{\alpha}{2\pi}\right\}J_r\left(\frac{nD}{\mu}\right)\cos(2\pi r\mu t-n\alpha t-n\theta) \tag{87}$$

where: $x<1$.

This shows that the effect of the interfering signal is to produce cross modulation between the desired audio signal of frequency μ and the difference angular frequency α. The amplitude of each frequency can be computed from this equation. When $\alpha = 0$, i.e., common-channel interference, this reduces to:

$$f=\frac{\omega}{2\pi}+D\cos 2\pi\mu t$$

$$+\sum_{n=1}^{\infty}\sum_{r=-\infty}^{\infty}(-x)^n\frac{\mu r}{n}J_r\left(\frac{nD}{\mu}\right)\left\{\cos 2\pi r\mu t\cos n\theta+\sin 2\pi r\mu t\sin n\theta\right\}$$

$$=\frac{\omega}{2\pi}+D\cos 2\pi\mu t$$

$$+ 2\mu \sum_{r=1}^{\infty} (2r-1)\, C\left(2r-1, \frac{D}{\mu}; x, \theta\right) \cos\{(2r-1)(2\pi\mu t)\}$$

$$+ 2\mu \sum_{r=1}^{\infty} (2r)\, S\left(2r, \frac{D}{\mu}; x, \theta\right) \sin\{(2r)(2\pi\mu t)\} \tag{88}$$

where the C- and S- functions are defined as follows:

$$C(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_n(ms) \cos s\theta \tag{89}$$

$$S(m, n; x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_n(ms) \sin s\theta \tag{90}$$

$$x^2 \leqq 1.$$

APPENDIX III.

Common-Channel Interference, Both Signals Modulated

When two frequency-modulated signals with a common carrier frequency produce interference, the effect is similar to that which occurs when one wave is not modulated. The exact relations can be obtained in the following way:

Let: $$e_1 = E_1 \sin\left(\omega t + \frac{D_1}{\mu_1} \sin 2\pi\mu_1 t\right) \tag{91}$$

and: $$e_2 = E_2 \sin\left(\omega t + \frac{D_2}{\mu_2} \sin 2\pi\mu_2 t\right) \tag{92}$$

be the two interfering waves. Then:

$$e_1 + e_2 = \sqrt{E_1^2 + E_2^2 + 2E_1E_2 \cos\psi}\ \sin\left(\omega t + \frac{D_1}{\mu_1} \sin 2\pi\mu_1 t - \varphi\right) \tag{93}$$

where: $$\tan\varphi = \frac{x \sin\psi}{1 + x\cos\psi}$$

and: $$\psi = \frac{D_1}{\mu_1} \sin 2\pi\mu_1 t - \frac{D_2}{\mu_2} \sin 2\pi\mu_2 t$$

The instantaneous frequency becomes:

$$f = \frac{\omega}{2\pi} + D_1 \cos 2\pi\mu_1 t - \frac{D_1 \cos 2\pi\mu_1 t - D_2 \cos 2\pi\mu_2 t}{\dfrac{\cos\psi + 1/x}{\cos\psi + x} + 1} \tag{94}$$

Envelope of Beat-note Pattern

The beat-note produced in the output of a receiver with a perfect limiter and balanced discriminator is given by:

$$\text{Output} = D_1 \cos 2\pi\mu_1 t - \frac{D_1 \cos 2\pi\mu_1 t - D_2 \cos 2\pi\mu_2 t}{\dfrac{\cos\psi + 1/x}{\cos\psi + x} + 1} \tag{95}$$

The two envelopes of the maxima and minima of the beat-note pattern are obtained by setting $\psi = 2n\pi$ or $\psi = (2n + 1)\pi$, where n is an integer. This gives the result:

$$\text{Envelope} = \frac{D_1}{1 + x} \cos 2\pi\mu_1 t + \frac{D_2 x}{x + 1} \cos 2\pi\mu_2 t \tag{96}$$

and:

$$\frac{D_1}{1 - x} \cos 2\pi\mu_1 t + \frac{D_2 x}{x - 1} \cos 2\pi\mu_2 t \tag{97}$$

Fourier-Series Analysis of Instantaneous Frequency

The distortion present in the instantaneous frequency is given by

$$-\frac{d}{dt} \tan^{-1} \frac{x \sin\psi}{1 + x\cos\psi} = \sum_{n=1}^{\infty} (-x)^n \cos n\psi \frac{d\psi}{dt} \tag{98}$$

Consider the expression:

$$\cos n\psi \frac{d\psi}{dt} = 2\pi \cos n\psi \, (D_1 \cos 2\pi\mu_1 t - D_2 \cos 2\pi\mu_2 t) \tag{99}$$

Make the substitutions:

$$\alpha = 2\pi\mu_1 t \qquad \gamma = n\frac{D_1}{\mu_1}$$

$$\beta = 2\pi\mu_2 t \qquad \delta = n\frac{D_2}{\mu_2}$$

Then the first term becomes:

$$D_1 \cos\alpha \cos\{\gamma \sin\alpha - \delta \sin\beta\}$$

$$= \frac{D_1}{4}\left\{e^{i\alpha} + e^{-i\alpha}\right\}\left\{e^{i\gamma \sin\alpha} e^{-i\delta \sin\beta} + e^{-i\gamma \sin\alpha} e^{i\delta \sin\beta}\right\}$$

$$= \frac{D_1}{4}\left\{e^{i\alpha} \sum_{r=-\infty}^{\infty} J_r(\gamma)\, e^{ir\alpha} \sum_{s=-\infty}^{\infty} J_{-s}(\delta)\, e^{is\beta}\right.$$

$$+ e^{-t\alpha} \sum_{r=-\infty}^{\infty} J_r(\gamma)\, e^{tr\alpha} \sum_{s=-\infty}^{\infty} J_{-s}(\delta)\, e^{ts\beta}$$

$$+ e^{t\alpha} \sum_{r=-\infty}^{\infty} J_{-r}(\gamma)\, e^{tr\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{ts\beta}$$

$$+ e^{-t\alpha} \sum_{r=-\infty}^{\infty} J_{-r}(\gamma)\, e^{tr\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{ts\beta} \Big\}$$

$$= \frac{D_1}{4} \Big\{ \sum_{r=-\infty}^{\infty} \Big[J_r(\gamma)\, e^{t(r+1)\alpha} + J_r(\gamma)\, e^{t(r-1)\alpha} \Big] \sum_{s=-\infty}^{\infty} J_{-s}(\delta)\, e^{ts\beta}$$

$$+ \sum_{r=-\infty}^{\infty} \Big[J_{-r}(\gamma)\, e^{t(r+1)\alpha} + J_{-r}(\gamma)\, e^{t(r-1)\alpha} \Big] \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{ts\beta} \Big\} \qquad (100)$$

Make the following substitutions:

$r + 1 = k$ in the first expression in the first bracket
$r - 1 = k$ in the second expression in the first bracket
$r + 1 = -k$ in the first expression in the second bracket
$r - 1 = -k$ in the second expression in the second bracket

and apply the identity:

$$J_{k-1}(\gamma) + J_{k+1}(\gamma) = \frac{2k}{\gamma} J_k(\gamma). \qquad (101)$$

This gives:

$$D_1 \cos\alpha \cos\{\gamma \sin\alpha - \delta \sin\beta\}$$

$$= \frac{D_1}{4} \Big\{ \sum_{k=-\infty}^{\infty} \Big[J_{k-1}(\gamma) + J_{k+1}(\gamma) \Big] e^{ik\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{-ts\beta}$$

$$+ \sum_{k=-\infty}^{\infty} \Big[J_{k-1}(\gamma) + J_{k+1}(\gamma) \Big] e^{-tk\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{ts\beta} \Big\}$$

$$= \frac{D_1}{2} \Big\{ \sum_{k=-\infty}^{\infty} \frac{k}{\gamma} J_k(\gamma)\, e^{tk\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{-ts\beta}$$

$$+ \sum_{k=-\infty}^{\infty} \frac{k}{\gamma} J_k(\gamma)\, e^{-tk\alpha} \sum_{s=-\infty}^{\infty} J_s(\delta)\, e^{ts\beta} \Big\}$$

$$= D_1 \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \frac{r}{\gamma} J_r(\gamma)\, J_s(\delta) \cos(r\alpha - s\beta) \qquad (102)$$

By the same process:

$$D_2 \cos\beta \cos\{\gamma \sin\alpha - \delta \sin\beta\}$$

$$= D_2 \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \frac{s}{\delta} J_r(\gamma)\, J_s(\delta) \cos(r\alpha - s\beta) \qquad (103)$$

These two results give the expression:

$$\cos n\psi \frac{d\psi}{dt} = 2\pi \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \left\{ \frac{rD_1}{\gamma} - \frac{sD_2}{\gamma} \right\} J_r(\gamma)\, J_s(\delta) \cos(r\alpha - s\beta)$$

$$= \frac{2\pi}{n} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} (r\mu_1 - s\mu_2)\, J_r(\gamma)\, J_s(\delta) \cos(r\alpha - s\beta) \qquad (104)$$

The audio output from a balanced discriminator thus becomes:

Output $\propto D_1 \cos 2\pi\mu_1 t$

$$\sum_{n=1}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \frac{(-x)^n}{n} (r\mu_1 - s\mu_2)\, J_r\left(\frac{nD_1}{\mu_1}\right) J_s\left(\frac{nD_2}{\mu_2}\right) \cos(r\alpha - s\beta)$$

$$= D_1 \cos 2\pi\mu_1 t + \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} (r\mu_1 - s\mu_2)\, C\left(r, \frac{D_1}{\mu_1}; s, \frac{D_2}{\mu_2}: x, 0\right)$$

$$\cos(r\alpha - s\beta) \qquad (105)$$

where the generalized C-function is defined as follows:

$$C(k, l; m, n: x, \theta) = \sum_{s=1}^{\infty} \frac{(-x)^s}{s} J_k(s\,l)\, J_m(sn) \cos s\theta, \qquad (106)$$

$\alpha = 2\pi\mu_1 t$ and $\beta = 2\pi\mu_2 t$.

THE RATIO DETECTOR*†

By

STUART WM. SEELEY AND JACK AVINS

Industry Service Laboratory, RCA Laboratories Division,
New York, N. Y.

Summary—A new circuit for f-m‡ detection known as the ratio detector is coming into wide use. In this circuit two frequency-sensitive voltages are applied to diodes and the sum of the rectified voltages held constant. The difference voltage then constitutes the desired a-f‡ signal. This means of operation makes the output insensitive to amplitude variations.

The basic principles governing the operation and design of the ratio detector are described. The ratio between the primary and secondary components of the frequency-sensitive voltages in a phase-shift type of ratio detector is a function of the instantaneous signal amplitude. The a-m† rejection properties, however, are shown to depend upon the mean ratio between these voltages. An expression which is developed for this ratio in terms of the circuit parameters provides the basis for arriving at an optimum design. The measurements necessary in the design of a ratio detector and in checking its performance are described.

GENERAL DESCRIPTION

THE principle underlying most circuits for f-m detection has been the peak rectification of two i-f‡ voltages whose relative amplitudes are a function of frequency, together with means for combining the rectified voltages in reversed polarity. This introduces a difference voltage, proportional to the difference of the two applied i-f voltages, which is a function of the instantaneous applied frequency. A representative circuit of this type is shown in Figure 1(a).

While this circuit represents a considerable improvement over unbalanced f-m detectors in that it is amplitude-insensitive at the center frequency, it has the disadvantage that changes in amplitude at other than the center frequency will cause the audio output to fluctuate proportionally as shown in Figure 1(b). This has been recognized and has resulted in the use of limiter circuits to remove extraneous amplitude variations from the applied signal.

In the ratio type of f-m detector, the problem of making the de-

* Decimal Classification: R362 × R361.111.
† Reprinted from *RCA REVIEW*, June, 1947.
‡ Throughout this paper the following abbreviations are used: f-m: frequency modulation or frequency-modulated; a-m: amplitude modulation or amplitude-modulated; a-f: audio frequency; i-f: intermediate frequency.

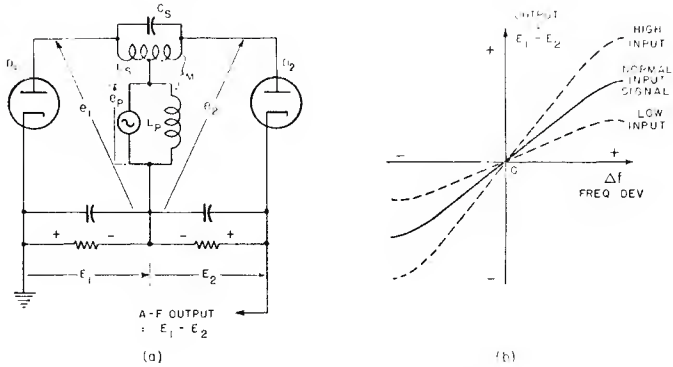


Fig. 1—Basic circuit of a balanced discriminator. The output characteristic in (b) shows the dependence of the output on the input signal amplitude.

tector insensitive to amplitude variations has been met by splitting the rectified i-f voltages into two parts in such a way that the ratio of the rectified voltages is proportional to the ratio of the applied frequency-sensitive i-f voltages. It follows that if the sum of these rectified voltages is maintained constant by a suitable means, *and if their ratio remains constant,* the individual rectified voltages will also remain constant. The output will therefore tend to be independent of amplitude variations in the input signal. A representative simplified ratio-detector circuit is shown in Figure 2. The circuit connections are such that both diodes carry the same direct current. The rectified voltages add to produce the sum voltage, which is held constant.

The sum voltage may be stabilized by using a battery or by shunt-

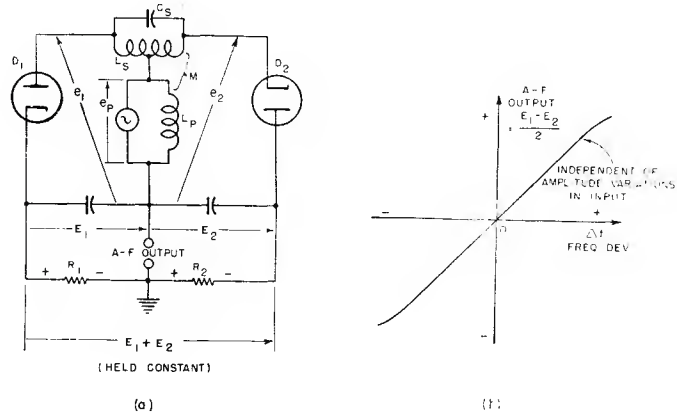


Fig. 2—Basic circuit of a ratio detector using the same phase shift input circuit as in Figure 1. The rectified voltage is stabilized so that the output can be independent of the input amplitude.

ing a large condenser across the load resistors. If a battery is used, operation is limited to a signal at least strong enough to overcome the fixed bias created by the battery. On the other hand, if a large condenser is used, the voltage across the condenser will vary in proportion to the average signal amplitude and thus automatically adjust itself to the optimum operating level. In this way amplitude rejection can be secured over a wide range of input signals, the lowest signal being determined by deterioration of the rectifier characteristics at low levels.

When a condenser is used to stabilize the rectified output voltage its capacitance must be large enough so that the rectified output

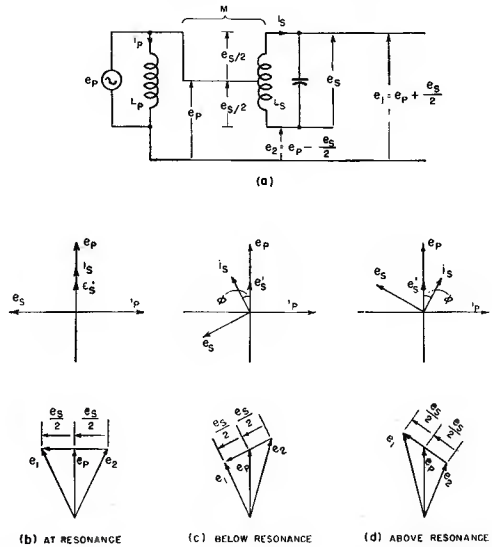


Fig. 3—Basic phase shift circuit used in frequency discriminators and ratio detectors. The phase relations between the primary and secondary voltages are shown, as well as the resultant voltages which are applied to each diode.

voltage does not vary at audio frequency. This calls for a time constant on the order of 0.2 second and for capacitance values on the order of several microfarads.

Elementary Operation

Although the exact analysis of a practical ratio detector is complicated by many factors, it is not difficult to obtain a qualitative understanding of the method of operation. A more detailed investigation of the circuit is made later.

Since the ratio detectors described here use the balanced phase shift type of input circuit, the characteristics of this circuit are briefiy reviewed. Figure 3 shows the basic phase shift circuit and the manner in which the two frequency-sensitive voltages are obtained. These voltages are combined in the ratio detector as shown in Figure 4(a). The variations in the open-circuit voltages e_1 and e_2 are conveniently described by Figure 4(b). The rectified voltages E_1 and E_2 are approximately proportional to OA and OB. The phase angle ϕ is zero at the center frequency and increases as the deviation increases.

Using this diagram it is possible to draw the detector output characteristic as shown by the $(E_1\text{-}E_2)$ curve in Figure 5. This is

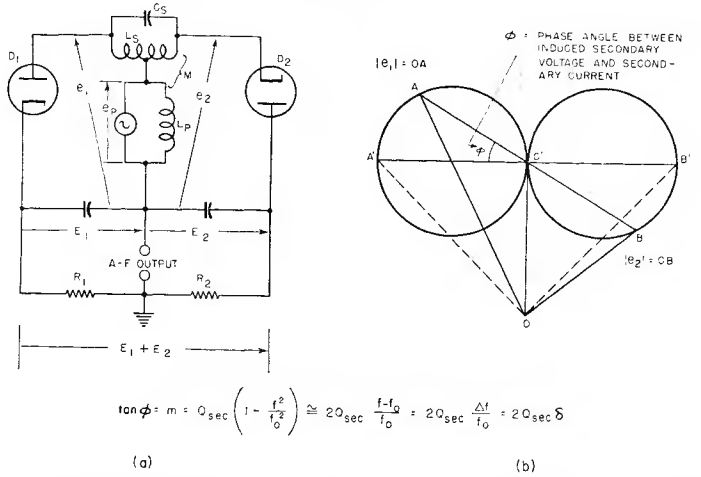


Fig. 4—Basic ratio-detector circuit showing the variation with frequency of the voltages applied to each of the diodes.

similar to the characteristic obtained with a conventional discriminator. A factor ½, however, should be applied to indicate the six-decibel loss which is inherent in the reversed polarity connection of the diodes and the change in the a-f take-off point. Thus, when, $E_1\text{-}E_2$ changes by a given amount, the change in potential at the a-f take-off point in Figure 4 is $(E_1\text{-}E_2)/2$.

If now the rectified voltage is stabilized as in Figure 2, the output characteristic changes. Over the region where $E_1 + E_2$ is constant, i.e. for small deviation from the center frequency, Figure 5 shows that a ratio detector will have essentially the same output (in the absence of amplitude modulation) as a conventional discriminator using the same transformer and load resistors but with one of the diodes reversed. For large deviations, $E_1 + E_2$ falls and the application of

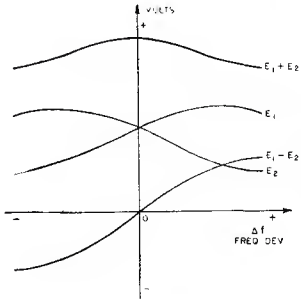
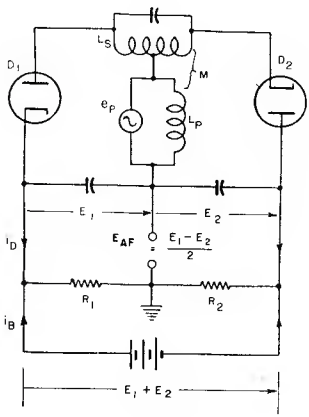


Fig. 5—The voltages E_1, E_2, E_1+E_2, E_1-E_2 which are present in Figure 4(a) can be plotted as shown, with the aid of Figure 4(b).

the stabilizing voltage to maintain $E_1 + E_2$ constant has the effect of improving the linearity of the detector characteristic and extending the maximum deviation that can be handled for a given amount of distortion.

The manner in which the ratio detector maintains a constant output independent of amplitude modulation may be seen by reference to Figure 6. If the amplitude of the input signal is constant as in (a), the stabilizing current is zero and the circuit has essentially the same output characteristic as a balanced discriminator. On the other hand, if the input signal increases as in (b), the diodes are driven harder, the average diode current increases, and this increased direct current flows into the stabilizing voltage source. Similarly, if the input signal decreases, as in (c), the diode current decreases and the stabilizing



(a) Constant normal input amplitude $e_P = e_0$
Stabilizing current $i_B = 0$
Diode current $i_D = \frac{E_1+E_2}{R_1+R_2}$
Effective diode load resistance $= R_1+R_2$

(b) Increase in input amplitude $e_P > e_0$
Stabilizing current $i_B < 0$
Diode current $i_D > \frac{E_1+E_2}{R_1+R_2}$
Effective diode load resistance $< R_1+R_2$

(c) Decrease in input amplitude $e_P < e_0$
Stabilizing current $i_B > 0$
Diode current $i_D < \frac{E_1+E_2}{R_1+R_2}$
Effective diode load resistance $> R_1+R_2$

Fig. 6—The effect of changes in input signal level on the diode current, the stabilizing current, and the effective load resistance presented to the diodes. The stabilizing voltage $E_1 + E_2$ is held constant at the voltage equal to the rectified output when $e_p = e_o$.

voltage makes up the decrease in diode current in order to maintain the sum voltage across the two diode load resistors constant. Stabilizing the rectified voltage, therefore, results in the equivalent load resistance varying in such a way that it decreases when the input signal rises and increases when the input signal falls.

This action of the stabilizing voltage in varying the effective diode load resistance provides a convenient method for analyzing the mechanism of amplitude rejection. A first approximation to the behavior of the ratio detector, then, is to consider how the output is affected when the load resistance is varied above and below its mean operating value. Figure 7 shows the result of varying R, while all other parameters, including the input signal, are held constant.

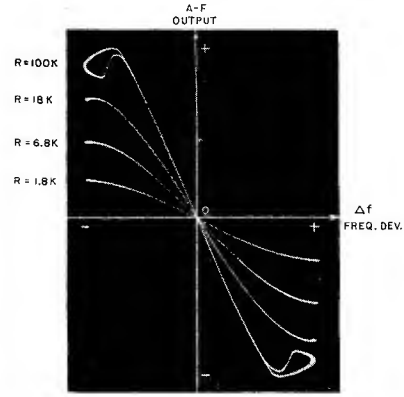


Fig. 7—Effect on the output of the ratio detector when the load resistance is varied. The curves were taken with the circuit shown in Figure 8.

If the rectified outputs $E_1 + E_2$ and $E_1 - E_2$ are measured, it is readily possible to determine whether the detector constants, the primary-secondary inductance ratio, the primary and secondary Q's, the coupling, the diode perveance, and the operating load resistance are such as to yield good amplitude rejection. If the circuit is designed properly, it will be found that, as R is varied, $E_1 - E_2$ will increase (or decrease) at the same rate as $E_1 + E_2$. In other words, $(E_1 - E_2)/(E_1 + E_2)$ must be independent of R and hence independent of the rectified current. If $(E_1 - E_2)/(E_1 + E_2)$ is independent of the rectified current, then it follows that the ratio E_1/E_2 must also be independent of the rectified current.

To take a specific example, suppose R is increased so that, for a fixed deviation, $(E_1 - E_2)/2$ increases to twice its original value,

Then $E_1 + E_2$ will also double. The input signal may therefore be reduced to one half its original value and the a-f output will then be the same as it was before the change in R and the reduction in input signal. In practice, when a battery or a large condenser is used across the load resistors, any variation in the input signal automatically causes the equivalent load resistance to vary in such a way as to keep the a-f output constant, provided of course that the detector circuit parameters are properly related.

The amount of downward amplitude modulation which can be rejected, that is, the extent to which the signal can fall below its initial value without causing the output to change, is an important characteristic. The limiting factor here is that, if the input signal drops to too low a value, the voltages across the primary and secondary are not sufficient to cause the diodes to conduct. Effectively, the diodes are biased off by the stabilizing voltage and the detector becomes inoperative until the signal rises to a level sufficiently great to cause the diodes to conduct.

It is desirable to design the ratio detector so that the signal can fall to as low a value as possible without the diodes being cut off. This is accomplished, in general, by using a high value of secondary Q and by using relatively low values of diode load resistance so that the operating Q of the secondary is on the order of 25 per cent of its unloaded Q. Other considerations are also involved, but the extent to which the operating Q is smaller than the unloaded Q is the principal factor in determining the maximum percentage of downward amplitude modulation which can be rejected without the diodes being cut off.

The limitation of the diodes being biased off by the stabilized rectified voltage does not exist as far as upward modulation is concerned. In general, the ratio detector can reject higher percentages of upward modulation than it can downward modulation.

Ratio Detector Receivers

Receivers using the ratio detector have characteristics which differ from limiter and locked-oscillator receivers in a number of respects. Since the rectified voltage across the long-time-constant load circuit of the ratio detector automatically adjusts itself to the input signal level, there is no fixed threshold, and the a-f output and the stabilizing voltage (which may be used for automatic-volume-control (a-v-c)) are proportional to the input signal. Less i-f gain is required since the ratio detector in typical designs will provide appreciable amplitude rejection with as little as ten to fifty millivolts of input signal to the grid of the ratio-detector driver tube. As a result, receivers using

this type of detector tend to be quiet between stations. When a-v-c is used, the selectivity of the receiver is maintained for strong as well as weak signals.

The tuning characteristic of a receiver using the ratio detector is characterized by comparatively low side responses. This results, in part, because the high degree of amplitude modulation which an off-resonant f-m signal acquires as it is modulated up and down the steep i-f selectivity curve produces relatively little output from the ratio detector. The extent to which the side responses will fall below the main response depends upon the amount of a-v-c, the i-f selectivity, and the ratio-detector peak separation.

Typical Ratio Detector

The circuit and performance data for a ratio detector circuit oper-

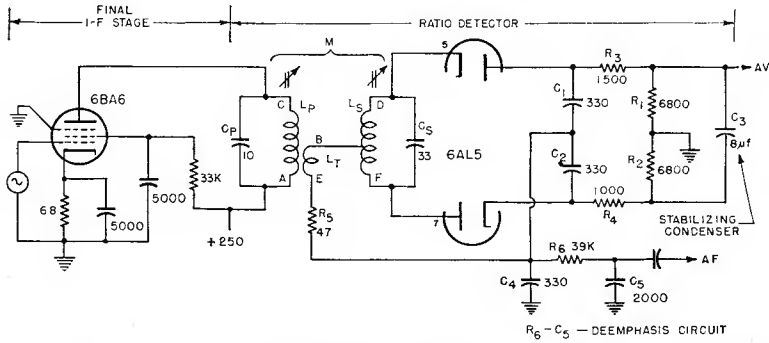
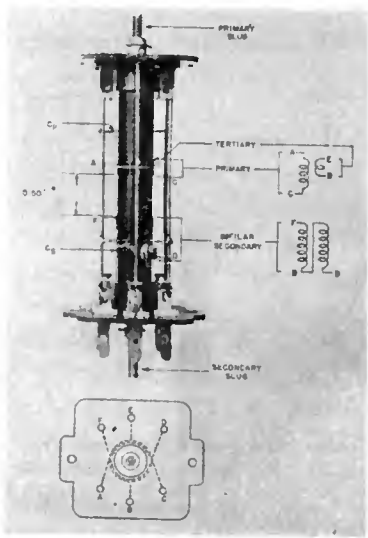


Fig. 8—A ratio-detector circuit using a balanced phase shift transformer feeding a 6AL5 double diode. The closely coupled tertiary winding L_T is wound over the $B+$ end of the primary to provide impedance matching and improved sensitivity.

ating at a center frequency of 10.7 megacycles is shown in Figure 8. In order to improve the sensitivity, a form of impedance matching is used between the high impedance of the driving pentode and the relatively low impedance which the diodes reflect across the winding that injects the primary voltage into the diode circuit. This is, of course, the winding L_t which is wound over the cold end of the primary winding L_p and is tightly coupled to the primary. The phase of the voltages across L_p and L_t is therefore substantially the same. The voltages across L_t and L_s are in quadrature (at the center frequency) and in accordance with the action which has been previously described, the demodulated f-m output is taken off through a connection to the secondary winding.

The construction of the transformer is evident from the photograph in Figure 9. Slug tuning is used for both primary and secondary windings. Since a bifilar construction is used for the secondary winding, the slug penetrates each half of the winding to essentially the same degree, and hence variation in the position of the slug does not unbalance the two halves of the secondary winding. Even where the secondary is capacitively tuned, this bifilar construction is desirable because it provides close coupling between the two halves of the secondary and uniform coupling of each half of the secondary to the primary winding.

As shown in Figure 8, the major part of the rectified output voltage is stabilized by means of the 8-microfarad condenser C_3. This, in con-



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*

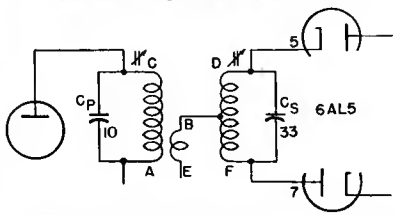


Fig. 9—Details of construction of the ratio-detector transformer used in the circuit of Figure 8. The bifilar secondary connections are brought out symmetrically to the vertical rods to prevent unbalance.

junction with the two diode load resistors R_1 and R_2, gives a discharge time constant of approximately 0.1 second. The rectified voltage drop across R_3 and R_4 is not stabilized against changes in diode current. This permits minimizing the residual balanced component of amplitude modulation (See Figure 21) for this particular circuit design. The fact that the two resistors R_3 and R_4 are not equal in value makes it possible to produce a compensating unbalanced component which cancels the unbalanced component which could otherwise appear in the output (see Figure 22). This unbalance is principally due to the variation in the dynamic input reactance of the diodes and is not

necessarily an indication that the circuit itself is unbalanced in any way. The 47-ohm resistor R_5 modifies the peak diode currents and further reduces the unbalanced a-m component, particularly at high signal levels.

The condenser C_4 bypasses the secondary system to ground at the intermediate frequency. The conventional deemphasis circuit which is conveniently placed at this point provides further filtering against i-f getting into other circuits and causing feedback. A 39,000-ohm resistor and a 0.002-microfarad condenser are suitable values for the deemphasis circuit.

The alignment procedure for this detector consists of applying an unmodulated i-f signal to the ratio-detector driver tube and peaking

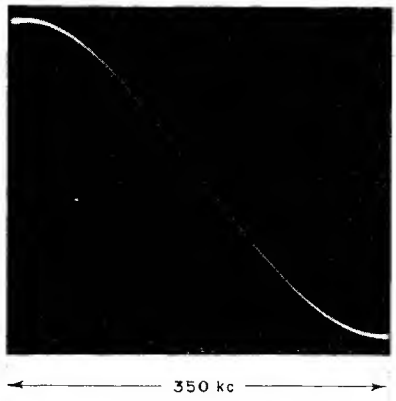


Fig. 10—Output characteristic and performance data for the ratio detector shown in Figure 8. Input signal = 100 millivolts; Rectified voltage = 6 volts; A-V-C voltage = 2.5 volts; A-F output (75-kilocycle deviation) = 0.7 volt root-mean-square; Distortion (100 per cent modulation) = 2.5 per cent root-mean-square; Distortion (30 per cent modulation) = 0.7 per cent root-mean-square; Maximum downward amplitude modulation = 70 per cent.

the primary so that maximum rectified voltage is obtained, as indicated by a d-c voltmeter connected to the a-v-c takeoff point (Figure 8). The secondary tuning is then adjusted for zero voltage as indicated by a d-c voltmeter connected to measure the voltage at the audio take-off point. It will be noted that the alignment procedure is essentially the same as for a balanced discriminator. Because the primary and secondary windings are below critical coupling, there is negligible interaction between the adjustments.

The output characteristic of this detector is shown in Figure 10, along with the data giving the audio recovery and rectified voltage for 100 millivolts input to the last i-f tube, i.e., to the ratio-detector

driver tube. Both the rectified output voltage and the a-f voltage are closely proportional to the average signal input.

One way of showing the amplitude-rejecting properties of this detector is by means of the output characteristic curves shown in Figure 11. These are taken with a battery across the stabilizing condenser to hold its voltage constant at the normal value which was present just before the battery was applied. The input signal is then reduced in strength until the diodes are cut off. Note that the output remains essentially constant over a band of more than 150 kilocycles while the signal is reduced to one-third of its initial value. The peak separation, however, is reduced as the input signal is reduced.

Another more convenient way of showing the amplitude rejection is to use an f-m signal which is simultaneously amplitude modulated.

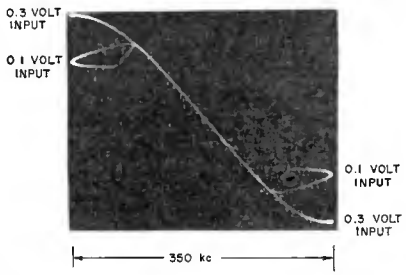


Fig. 11—Effect of reduction in the input signal. The stabilizing voltage is held constant at a level corresponding to the 0.3-volt input signal. Note the reduction in peak separation for decreasing input signal.

When this type of signal is applied, no stabilizing battery is required since the stabilizing condenser will maintain the rectified voltage constant for a-m frequencies above about 100 cycles. As explained in detail later, the extent to which the characteristics in Figure 12 differ from a line, indicates the relative amount of residual amplitude modulation in the f-m output.

In arriving at the design of this detector (Figure 8), a number of circuit parameters, in addition to those previously mentioned, are involved. These include the unloaded and loaded values of primary and secondary Q, the per cent of critical coupling between primary and secondary, and the grid to plate gain under various conditions. They are tabulated below:

Primary: Unloaded $Q = 70$
Operating Q (with diode loading) $= 40$

Secondary: Unloaded $Q = 89$.
Operating Q (with diode loading) $= 21$

Half-secondary/tertiary voltage ratio: 0.65

Coupling: 0.50 of critical. (This includes the effect of the capacitance unbalance due to the difference between the input capacitances of the two diodes.)

Grid-to-plate gain: The normal gain is 100. (Under the conditions yielding maximum gain, i.e., with downward amplitude modulation off the center frequency, the grid-plate gain rises to 130.)

Circuit Analysis

The analysis of the ratio detector is more difficult than that of the

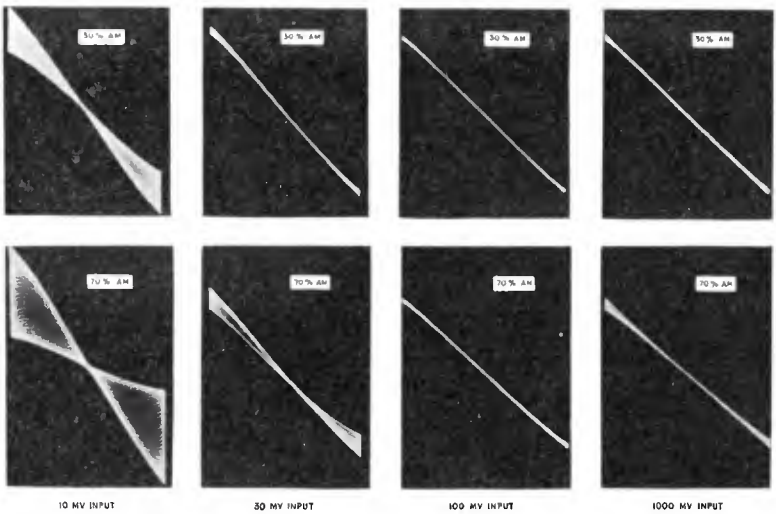


Fig. 12—The demodulated output when an f-m signal, 75 kilocycles deviation, having simultaneous a-m modulation, is applied to the ratio detector in Figure 8. The f-m modulation is at 400 cycles; the a-m modulation is at 250 cycles. The width of the detector characteristic indicates the residual a-m output.

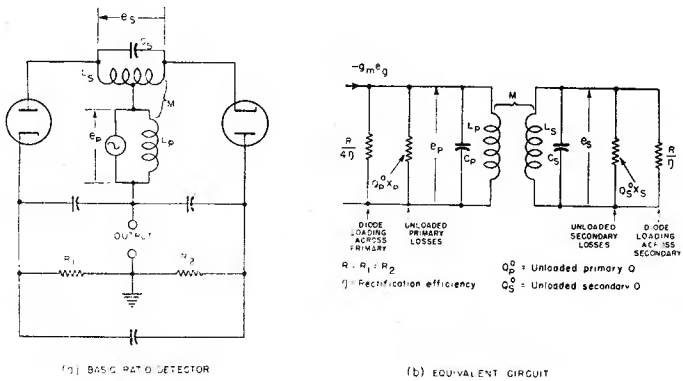
balanced discriminator primarily because of the variation in equivalent load resistance during the a-m cycle as well as during the f-m cycle. In addition, the problem is complicated by the complex impedance characteristics of the frequency-sensitive driving circuit, the non-linearity of the diode characteristics, and the input reactance variations of the diodes which tend to modify the frequency characteristic of the driving circuit. A productive method of attack is to consider first the equivalent circuit and to develop an expression for the ratio of primary and secondary voltage. This expression is impor-

tant because the degree of a-m rejection and the peak separation are closely related to this ratio.

Equivalent Circuit

The effect of the diodes on the ratio detector circuit can be simulated, as shown in Figure 13, by placing a resistance $R/4\eta$ across the primary and a resistance R/η across the secondary. This is the same relationship which applies to the balanced discriminator circuit.

From Figure 13(b) it is clear why some form of impedance matching is of great importance if good sensitivity is to be obtained. It is not uncommon for R to have values of the order of 10,000 ohms or less, since R must be low enough to reduce the secondary Q to about one-



(a) BASIC RATIO DETECTOR (b) EQUIVALENT CIRCUIT

Fig. 13—The equivalent circuit for a basic phase shift ratio detector shows that the diodes reflect an equivalent shunt resistance $R/4\eta$ across the primary and a resistance R/η across the secondary.

fourth of its unloaded value. This would result in a load impedance of 2500 ohms which is too low to provide good energy transfer from a pentode.

A typical circuit using a tapped primary to obtain impedance matching has already been described. The equivalent circuit for this arrangement is shown in Figures 14 and 15. This circuit differs from Figure 13 in that the diode loading is across the tertiary and appears across the primary as a resistance n^2 times as great.

Referring to Figures 14 and 15, it is shown in the Appendix that

$$\frac{S}{P} = \alpha \sqrt{1 + \frac{n^2 Q_s^1}{4\alpha Q_p^0}} \qquad (1)$$

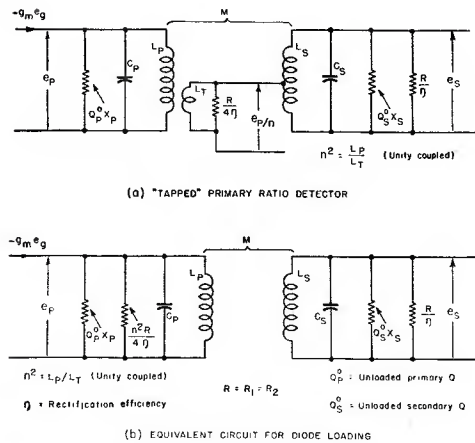


Fig. 14—The equivalent circuit for a ratio detector using a closely coupled tertiary winding (Figure 8) or a tapped primary to obtain impedance matching.

where S = half secondary voltage at the center frequency

P = "primary" (i.e. tertiary) voltage effective in the diode circuit, at the center frequency

$100\,\alpha$ = per cent of critical coupling existing between L_p and L_s

a = ratio of primary and secondary inductance = L_p/L_s

n = ratio between the primary and tertiary voltages, which corresponds approximately to the primary being tapped at $1/n$ turns

Q_s^1 = operating Q of the secondary, taking into account the diode loading

Q_p^0 = unloaded Q of the primary

The assumption is made in deriving this expression that the oper-

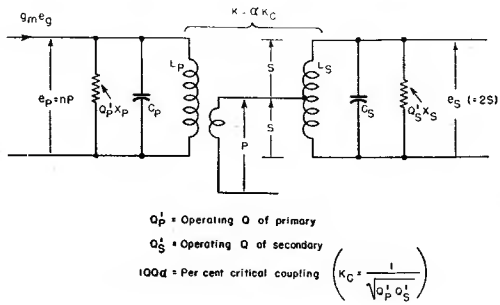


Fig. 15—A further simplification of the equivalent circuit of Figure 14, showing the voltages P and S which are applied to the rectifier circuit.

ating Q of the secondary (Q_s^1) is small in comparison with the unloaded Q (Q_s^0). This assumption is usually justified and accounts for the factor R/η not appearing in the equation. Q_p^0 appears in the equation because it cannot usually be assumed that the loaded primary Q will be small in comparison with the unloaded primary Q.

It is shown in the Appendix that the half-secondary voltage S is given by

$$S = g_m \omega L_s Q_s^1 n \frac{1}{4\sqrt{1 + \frac{n^2 Q_s^1}{4a Q_p^0}}} \frac{\alpha}{1+\alpha^2} \tag{2}$$

or

$$S = \frac{g_m \omega L_s Q_s^1 n}{4} \left(\frac{P}{S}\right) \frac{\alpha^2}{1+\alpha^2} \tag{3}$$

The expression for S is put in this form because in general S/P will not be an independent variable but will be fixed by considerations which determine the amplitude rejection and the peak separation.

Effect of Varying S/P

Before discussing the significance of these equations, it is desirable to show the dependence of peak separation and a-m rejection on the value of S/P. For convenience, constant primary voltage is assumed; this simplifies the analysis and provides a basis for obtaining a solution when the primary voltage is not constant. In this connection, whether the primary voltage variation is due to external amplitude modulation, or whether it is due to the primary selectivity makes no appreciable difference in the analysis.

The curves in Figure 16 were taken with constant primary voltage. As indicated, the a-f output characteristic is plotted for several values of S/P, P being held constant at the same value for all curves. The output characteristics are shown with and without the rectified voltage being stabilized. In addition, the variation in the rectified voltage is shown for each value of S/P. The same voltage scale is used throughout for the a-f output and the rectified voltage variation.

From Figure 16 it is clear that as S/P is increased, the peak separation increases. This is true whether the rectified voltage is stabilized or not; that is, it is equally true for the balanced discriminator or for the ratio detector. When the rectified voltage is stabilized, how-

ever, the effect of the stabilization is much more pronounced when S/P is large than when S/P is small. The reason for this effect is apparent from a consideration of the rectified voltage variation. As shown in Figure 16 this variation is much greater when S/P is large than when S/P is small. It follows that the effect of stabilizing the rectified voltage and removing the variations referred to above will be greater when S/P is large. Furthermore, since $E_1 - E_2$ and $E_1 + E_2$ (Figure 16) tend to fall as the deviation increases, the effect of stabilizing the rectified voltage is to produce an increase in the peak separation.

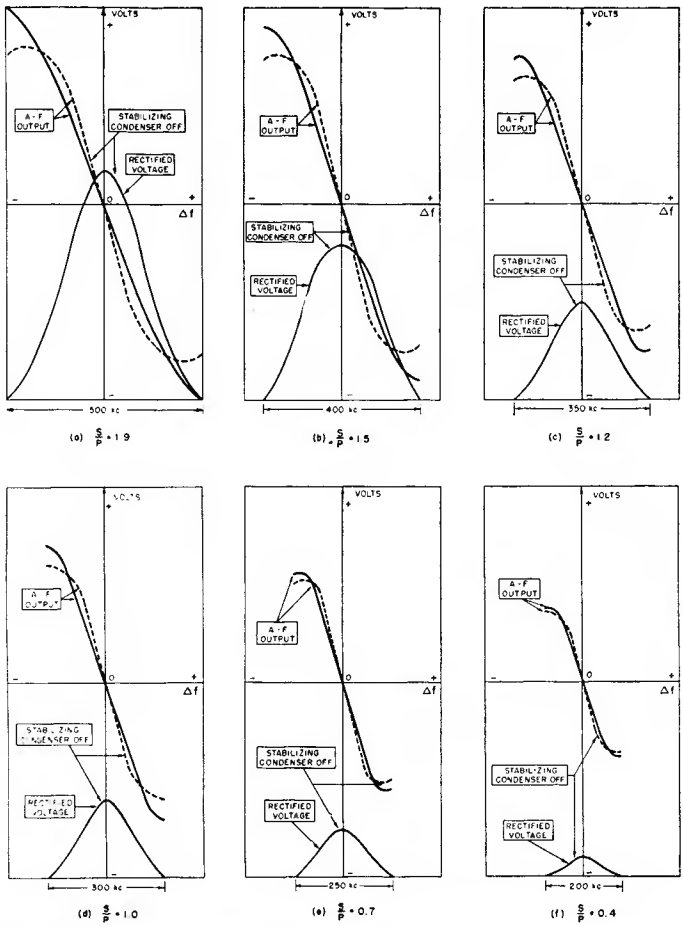


Fig. 16—The output of a ratio detector as a function of the ratio between the secondary and primary voltage, the primary voltage being held constant. The rectified output voltage variation, when the stabilizing condenser is removed, and the corresponding a-f output are also shown for each value of S/P.

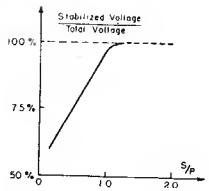


Fig. 17—The fraction of the total rectified voltage which must be stabilized to reduce the residual balanced amplitude modulation is a function of *S/P*. For values of *S/P* greater than about unity, the a-m in the output increases if less than the full rectified voltage is stabilized.

The a-m rejection is a matter of primary interest. To determine the effect of *S/P* on the a-m rejection, amplitude modulation was applied for each of the conditions shown in Figure 16. As shown in Figure 17, a threshold value of *S/P* exists, and for values of *S/P* greater than this it is not possible to obtain good a-m rejection. At this critical value, which is approximately unity, the full rectified voltage must be stabilized. As the value of *S/P* is decreased below this value, good a-m rejection will be obtained provided a particular fraction of the full rectified voltage is stabilized. The lower the value of *S/P*, the smaller the fraction of the rectified voltage which must be held constant. The ratio of unloaded to loaded secondary *Q* is four to one for the curve shown in Figure 17. A 6AL5 double diode was used.

Stabilizing the Rectified Voltage

Two general methods for controlling the desired percentage of the rectified voltage which is stabilized are shown in Figure 18. In (a), a resistor is added in series with the stabilizing condenser. In (b), the same effect is obtained by adding resistors R_a and R_b in series with the load resistance which is shunted by the stabilizing condenser.

If these circuits are to be equivalent, then a change $\triangle I$ in the d-c rectified current must result in the same change in rectified voltage for each of the two circuits.

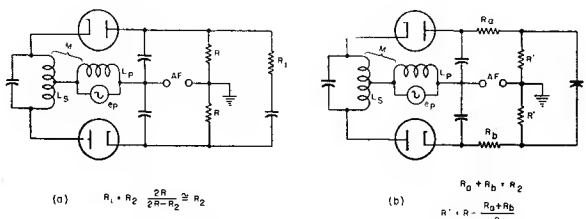


Fig. 18—Two methods of stabilizing any desired fraction of the rectified output voltage. In each case the fraction of the output voltage stabilized is approximately equal to $2R/(2R + R_1)$.

For circuit (a), $$\triangle E = \triangle I \frac{2R}{2R + R_1} \cdot R_1$$

For circuit (b), $$\triangle E = \triangle I \ (R_a + R_b) = \triangle I \cdot R_2$$

$$\therefore \ \triangle i \cdot R_2 = \triangle I \frac{2RR_1}{2R + R_1}$$

$$\text{or} \quad R_1 = R_2 \frac{2R}{2R - R_2} \cong R_2 \quad \text{provided } R_2 << 2R$$

Both circuits provide equal diode loading since the total load resistance is the same. When the desired variation in the rectified output voltage is small, then $R_1 = R_2$ as shown above.

Factors Affecting S/P

Consider the expression for S/P given by Equation (1). If Q_p^0 is large enough so that the primary Q is determined solely by the diode loading, it is clear that

$$\frac{n^2 Q_s^1}{4a\, Q_p^0} << 1 \qquad \text{and} \qquad \therefore \frac{S}{P} = \alpha$$

This gives the useful result that for this condition the ratio S/P is independent of the ratio between the primary and secondary inductance, and also independent of the number of turns in the tertiary winding. For this condition, then, S/P is dependent only on the per cent of critical coupling and is equal to α.

When the diode damping of the primary is comparable to the primary losses, S/P is greater than α as indicated by Equation 1. The extent to which the primary is damped by the diode rectifiers is of interest since the amount of downward amplitude modulation that can be handled depends, among other factors, on the primary diode damping. The greater the amount of damping, the greater will be the rejection of downward amplitude modulation because downward amplitude modulation is accompanied by an increase in the effective load resistance and hence an increase in the effective primary Q. This action increases the gain and helps to offset the drop in signal level.

Referring again to Equation (1), it is desirable to determine the extent to which the damping on the primary can be increased without decreasing the value of S/P which must be maintained at approxi-

mately unity. Since α will be fixed within relatively narrow limits by considerations which will be discussed later, the only variable part in the expression for S/P is $n^2 Q_s^1/4a Q_p^0$. Assuming that Q_s^1 is as large as practicable, consistent with the unloaded secondary Q being several times the loaded secondary Q, this leaves the factor $n^2/4a Q_p^0$. This factor must have a particular value, depending on α, in order to make S/P approximately unity. The question arises as to whether there is any advantage in making n^2 large and $a Q_p^0$ proportionately large, or in keeping n^2 small and making $a Q_p^0$ proportionately small.

To determine the manner in which the primary damping varies for particular values of n^2 and corresponding values of $a Q_p^0$, consider the ratio of the shunt damping resistance introduced by the diodes and the equivalent shunt resistance of the undamped primary. This ratio is equal to

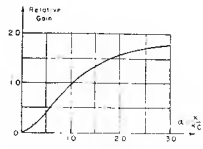
$$\frac{n^2 R}{4Q_p^0 X_p} \propto \frac{n^2}{Q_p^0 a X_s} \propto \frac{n^2}{a Q_p^0}$$

since X_s is not a variable here. Comparing this expression with Equation (1), it follows that the relative amount of diode damping on the primary does not change with any change in n, Q_p^0, or a, provided these changes are made in such a way as to keep constant both the ratio S/P and the percent of critical coupling. Of course, the sensitivity will change, increasing with increases in n, a, and Q_p^0. The grid-plate gain of the driver tube will also increase as n is increased.

It has just been demonstrated that the primary damping cannot be controlled by varying L_p/L_s, the unloaded primary Q, or the relative number of tertiary turns, if the ratio S/P is to be maintained. From Equation (1) it is clear that if the primary damping is to be increased, it must be accompanied by an increase in α, in order to maintain the ratio S/P constant. This increase in α is frequently undesirable, as discussed below.

Consider Equation (2) which gives the value of S. Again if P/S is held constant at approximately unity for optimum a-m rejection, the gain to the secondary increases with increasing values of Q_s, n and α. For values of α between 0.4 and 1.0, S increases rapidly with α, as shown in Figure 19. The implication of this, so far as design for handling maximum downward amplitude modulation is concerned, is that other things being equal, the initial per cent of critical coupling should be approximately 50. This means that as Q_p and Q_s increase with downward modulation, the resultant increase in the per cent of critical coupling will cause the greatest changes in the gain to the secondary.

Fig. 19—Relative amplification from the driver grid to the secondary as a function of the per cent of critical coupling, where the change in the per cent of critical coupling takes place as a result of changes in the primary or secondary Q.



Circuit Impedance Ratio

The ratio detector circuit may be put in the form shown in Figure 20. Normally the input signal voltages e_1 and e_2 will follow the approximate variation shown in Figure 5. These voltages may be derived from a phase-shift circuit, from side-tuned circuits, or from other suitable means. A large number of circuit variations are possible.

If the circuit is analyzed from an impedance standpoint, it is found that the rectified voltages are in the same ratio as the input voltages, provided the circuit impedances z_1 and z_2 are in the same ratio as the input voltages e_1 and e_2.

This can be demonstrated as follows. The diode currents flow in short pulses, assuming ideal diodes, and therefore the fundamental a-c component of each diode current is nearly equal to twice the d-c component. Since the d-c components are equal, it follows that the fundamental frequency a-c components will also be equal.

$$i_1 = i_2 = i; \qquad E_1 = k\,(e_1 - iz_1) \qquad \text{and} \qquad E_2 = k\,(e_2 - iz_2)$$

$$\therefore \frac{E_1}{E_2} = \frac{e_1 - iz_1}{e_2 - iz_2} = \frac{e_1\left(1 - i\dfrac{z_1}{e_1}\right)}{e_2\left(1 - i\dfrac{z_2}{e_2}\right)}$$

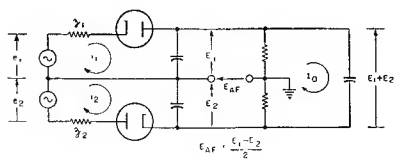


Fig. 20—The ratio-detector circuit in a more general form. The voltages e_1 and e_2 may be produced in various ways, including the phase-shift circuit and side-tuned circuits.

$$\frac{E_1}{E_2} = \frac{e_1}{e_2} \quad \text{provided that} \quad \frac{z_1}{e_1} = \frac{z_2}{e_2}$$

Since an arbitrary value of direct current was assumed, it follows that if $z_1/e_1 = z_2/e_2$, then E_1/E_2 will equal e_1/e_2 independently of the value of I_0.

If now $E_1 + E_2$ is stabilized, then it is evident that a-m rejection is obtained because E_1 and E_2 remain constant as the input signal amplitude is varied.

If only a fraction of $E_1 + E_2$ is stabilized, then the condition for zero a-m output will depart from the relation $e_1/e_2 = z_1/z_2$.

In the case of the balanced phase-shift circuit, useful relationships between the circuit parameters can be more easily obtained by the methods used in the preceding paragraphs. In other circuit variations, an attack based on the impedance relationships may be expedient.

Design Considerations

In this section the general principles governing the design of a ratio detector are described. These are based on the analysis in the preceding section. Although illustrations and specific discussions of circuit constants are based on a center frequency on the order of 10 megacycles and a peak separation on the order of 350 kilocycles, the principles are quite general and apply at other frequencies.

Diode Characteristics

Good ratio detector performance can be obtained with either high-perveance diodes such as the 6AL5 or medium-perveance diodes such as the 6H6. Diodes having a perveance lower than the 6H6 give less satisfactory performance in the ratio detector because the residual balanced component of amplitude modulation becomes appreciably large. In general, better performance has been obtained with ratio detectors using the 6AL5 than with those using the 6H6.

The circuits used with low and high perveance diodes will differ in the extent to which the rectified output voltage is held constant. They will also differ with respect to the compensation used to minimize the residual unbalanced component of amplitude modulation in the output. This unbalance is the result of the input reactance variation of the diodes exerting a detuning effect on the secondary. The magnitude of this effect tends to be less for low-perveance diodes.

The rectification efficiency and hence the diode circuit loading vary with signal level. For this reason, optimum a-m rejection is obtained

at a particular signal level. Above and below this level, a balanced a-m component is present. As indicated in Figure 12, the magnitude of the residual amplitude modulation is not large except at low levels where the diode perveance is low. The level at which optimum a-m rejection is obtained may be varied by altering the circuit constants, particularly the fraction of the total rectified voltage which is stabilized, or the ratio of secondary and tertiary voltages.

Secondary Inductance

It is desirable to make the secondary inductance-capacitance (L/C) ratio as high as possible, consistent with keeping the secondary tuning capacitance high enough so that stray capacitances and variations in tube capacitances do not have an excessive effect on the tuning. Using a high L/C ratio has the advantage of increasing the sensitivity and at the same time tends to reduce the peak diode currents. At a frequency on the order of ten megacycles these considerations point toward the use of a secondary tuning capacitance on the order of 25 to 75 micromicrofarads. Higher values result in lower sensitivity for the same amount of a-m rejection without any significant advantage as far as stability of tuning and independence of tube replacement are concerned.

The advantage of making the secondary a bifilar winding has been described. If a single center-tapped winding is used instead of a bifilar winding, it is desirable to use a fixed slug to obtain the same coupling between the two halves of the secondary and the primary. Another circuit variation involves the use of two equal fixed condensers across the secondary to obtain an i-f center-tap.

The secondary Q should, in general, be as high as possible, consistent with the desired peak separation. At frequencies on the order of ten megacycles, this will mean a Q on the order of 75 to 150. Use of the higher values of Q will result in improved sensitivity for a given a-m rejection. The peak separation, however, will be lower than if smaller values of secondary Q were used.

Load Resistors

The value R of the diode load resistors will depend on the amount of downward amplitude modulation to be handled which in turn depends upon the i-f selectivity characteristic, the amount of multipath transmission anticipated, and the importance of obtaining the maximum sensitivity. In general, R will be selected so as to reduce the operating secondary Q to a value approximately one-fourth or less of its unloaded value.

The lower the value of R, the greater will be the ratio between the unloaded and loaded values of the secondary Q. This means that when the diode current decreases during downward amplitude modulation, the Q of the secondary is able to rise considerably and to compensate for the decrease in input signal by effectively raising the sensitivity of the detector. The greater the extent to which the Q can rise when the input signal drops, the greater will be the downward modulation capability of the detector. In addition, the value of R affects not only the operating Q of the secondary, but also the operating Q of the primary. As a result of both of these factors, using lower values of load resistance increases the extent to which the signal can drop below its average value without the diodes being biased to cutoff by the stabilizing voltage.

If the i-f passband preceding the ratio-detector driver tube is not flat over a range equal to the maximum frequency swing, the amplitude modulation introduced by this selectivity must be removed by the ratio detector. As a result of this, the more the i-f selectivity characteristic deviates from a relatively flat top, the lower are the load resistance values which must be used.

If it is expected that a receiver using a ratio detector will be used under conditions of multipath reception, it follows that the detector must be capable of removing the amplitude modulation which results from the partial cancellation of the direct signal by the delayed signal which reaches the receiver over a longer path. Experience with multipath reception indicates that it is desirable for the detector to be able to reject more than 50 per cent downward amplitude modulation to reduce distortion as much as possible. Where very severe multipath is encountered, the phase modulation distortion which accompanies it becomes appreciable even if all of the a-m distortion is removed.

Sensitivity is also involved in the choice of load resistors. The greater the value of load resistance R, the greater will be the sensitivity and the less the downward modulation-rejecting capability.

Primary L/C Ratio and Q

The primary L/C ratio should be as large as possible in order to obtain the greatest sensitivity. The limiting factor is the maximum stable gain between the grid and plate of the ratio detector driver tube. When determining this maximum gain, it should be kept in mind that the gain may increase during downward amplitude modulation, particularly on either side of the center frequency.

If the primary Q is high enough so that the operating Q of the primary is determined essentially by the diode loading, the grid-plate

gain will rise to a higher value during the a-m cycle. This increase in grid-plate gain is desirable from the viewpoint of increasing the ability of the detector to reject downward amplitude modulation, so that in practice the primary Q is made as high as possible, consistent with obtaining the required peak separation and not exceeding the maximum allowable grid-plate gain during downward amplitude modulation.

Primary Impedance Variation

During the a-m cycle two opposing effects take place which alter the primary impedance and hence affect the gain between grid and plate of the driver tube. When the input signal falls, for example, the impedance coupled into the primary by the secondary increases because the Q of the secondary rises. This tends to lower the gain from the grid to the plate of the driver tube. On the other hand, the diode loading across the primary decreases and this effect tends to raise the gain from the grid to the plate. From the viewpoint of rejecting the maximum downward amplitude modulation, it is desirable that the circuit parameters be so related that the net change in gain be such as not to result in a decrease in gain to the primary when the signal level falls.

Coupling

The value of coupling used, along with the relative number of turns used in the tertiary winding, is the principal factor in determining the ratio S/P. The importance of maintaining this ratio in the vicinity of unity, and the effect of varying the ratio by changing the coupling, or by any other means, has been described. In general, the coupling should be adjusted to be approximately one-half critical. Since the absolute coefficient of coupling which corresponds to critical coupling is itself dependent upon the load resistance and on the tertiary inductance, the final coupling adjustment must be made after the other parameters are fixed because the per cent of critical coupling is affected by changes in these.

Tertiary Inductance

In the notation used here, the tertiary inductance is specified in terms of the parameter n where the tertiary voltage is equal to $1/n$ times the primary voltage. With a specified value of coupling in mind, usually about 0.5 critical, the number of turns on the tertiary is adjusted so that the required S/P ratio is obtained. If the required S/P ratio is obtained with too few turns on the tertiary so that the

grid-plate gain is high, this condition can be remedied by reducing the primary inductance.

Residual Amplitude Modulation in Output

In general the ratio detector will not completely reject all of the applied amplitude modulation. Where the effect of the applied amplitude modulation is to cause the slope of the detector characteristic to vary, the output of the detector will appear as shown in Figure 21. The output amplitude modulation may be either in phase with the applied modulation, or out of phase, i.e., the slope of the detector characteristic may either increase or decrease slightly with increasing values of input signal, depending upon the circuit constants and the signal level. Where the output characteristic appears as shown in Figure 21, when simultaneous amplitude and frequency modulation are applied, the output is said to be balanced in that, just as in the

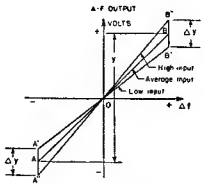


Fig. 21—Effect of amplitude modulation on the output of a balanced detector. If 100M = % amplitude modulation of the input signal, then $\Delta y = yM$, and $\Delta y/y = M$. This is referred to as a balanced a-m component.

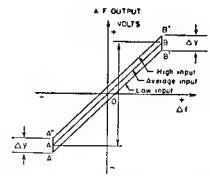


Fig. 22—Effect on the output of a balanced detector when amplitude modulation causes a detuning of the secondary. This is referred to as an unbalanced a-m component.

balanced discriminator, there is a center frequency at which the output is insensitive to applied amplitude modulation. With proper design, the amplitude modulation in the output, as indicated by the slope variation in Figure 21, will be very much less than the amplitude modulation in the input signal. This is indicated by the oscillograms in Figure 12.

The output of a ratio detector will not appear completely balanced as shown in Figure 21, unless the circuit is compensated to remove amplitude variations which result from the fact that the effective input capacitance of the diodes varies with the average diode current and therefore varies during the a-m cycle. The detector output characteristic in Figure 22 shows the output characteristic which is obtained when an unbalance is present in the circuit; this may be due not only to variations in the input capacitance of the diodes, but

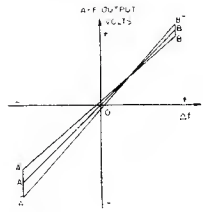


Fig. 23—When unbalance is present, the effect is to cause the crossover or point of zero a-m response to be at other than the center frequency. This type of characteristic has both a balanced component of a-m (Figure 21) and an unbalanced component (Figure 22).

may also be due to unbalance in the transformer, diodes, or circuit constants.

Where the output of a detector contains both balanced and unbalanced components, it will appear in general as shown in Figure 23. As would be expected, the effect of the unbalance is to shift the crossover from the center frequency.

Reducing Residual Amplitude Modulation

It is convenient to consider separately the means for reducing the residual balanced component and the residual unbalanced component of amplitude modulation in the output. Basically, the magnitude of the balanced component depends upon the ratio S/P at the mean carrier level. This may be visualized if the effect of a change in the secondary Q on the ratio E_1/E_2 in Figure 4 is considered; not only does the secondary Q and hence the phase angle of the secondary current change with diode loading during the a-m cycle, but the relative amplitudes of the primary and secondary voltages also change. There is thus a complex relationship which governs the extent to which E_1/E_2 is independent of diode loading, and this relationship is clearly a function of S/P. Thus, it is to be expected, as shown in Figure 24, that there should exist an optimum value of S/P for which E_1/E_2 is

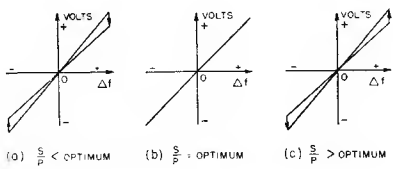


Fig. 24—On either side of the optimum value of S/P for which the balanced component of a-m is zero, there is a reversal in the phase of the a-m in the output. Since, for a given transformer, the value of S/P changes with coupling, the figure shows that for a loose coupling the output is 180 degrees out of phase with the input a-m, while the two are in phase for close coupling.

independent of diode loading and hence for which the balanced component of amplitude modulation in the output is a minimum. It will be noted in Figure 24 that the phase of the output amplitude modulation with respect to the input amplitude modulation undergoes a 180-degree reversal at the critical value of S/P. Since the principal effect of varying the coupling is to change the ratio S/P, Figure 24 also shows that for a given set of circuit constants, there exists a particular value of coupling for which the balanced component of amplitude modulation is zero.

Reducing Unbalanced A-M Component

The residual unbalanced component in the output may be reduced by several different means. The ratio detector circuit shown in Figure 8 contains two methods for reducing the unbalanced component. One of these consists of making the two resistors R_3 and R_4 unequal. By this means, it is possible to cancel the existing unbalance. This cancellation takes place because the differential drop across these resistors, which varies with a-m modulation, feeds into the a-f output. In practice, R_3 and R_4 are adjusted so that the unbalanced component as observed when frequency and amplitude modulation are applied simultaneously, goes to zero. During this adjustment the sum of R_3 and R_4 remains essentially constant since this sum depends upon the ratio S/P rather than upon any unbalance which may be present. In addition, a resistor R_5 is used in series with the tertiary winding common to both diode circuits. This modifies the peak diode currents and, particularly at high signal levels, has the effect of appreciably reducing the unbalanced a-m component.

Another method which may be used to reduce the unbalanced component is to use diode load condensers which have appreciable reactance at the intermediate frequency. At 10 megacycles, the values of load capacitance which are effective will range from about 50 micromicrofarads to several hundred micromicrofarads. In general, the greater the perveance of the diodes used in the ratio detector, the smaller will be the value of diode load capacitance which is required to eliminate the unbalanced component.

Still another method for reducing the unbalanced component is to vary the effective center-tap on the secondary winding slightly from the nominal center position.

Single-ended or Unbalanced Circuit

All of the circuits which have been described thus far have been of the balanced type, with the a-f take-off point at ground potential

when the input signal is applied at the center frequency. It is possible, however, to ground one side of the rectified output, as shown in Figure 25. In this case, the voltage available for a-v-c purposes is twice as great as in the balanced circuit case. There is no change in the a-f output.

Where only one diode load condenser is used, as in Figure 25(b), its capacitance should be twice as large as that used in Figure 25(a). This follows because in (a) the two condensers are effectively in parallel at the intermediate frequency, while in (b) since there is only one condenser its value must be twice as great. This relationship is of particular interest where the unbalanced a-m component is reduced by using a critical value for the diode load capacitance, as previously described.

In connection with unbalanced circuits, it may be noted for the circuit shown in Figure 8, it is not necessary that the center of the

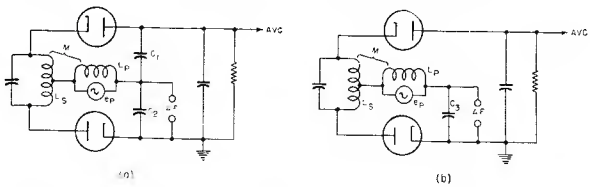


Fig. 25—Either one or two diode load condensers may be used in the ratio-detector load circuit. To obtain the same effective diode load capacitance in (b) as in (a), C_3 must be equal to the sum of C_1 and C_2.

load resistor be grounded; the a-m rejection will be essentially the same if any point along the load resistor bypassed by the stabilizing condenser is grounded. This statement is not true, however, at frequencies which are so low that the stabilizing condenser does not effectively hold the voltage across its terminals constant.

Time Constant of Stabilizing Voltage

At low audio frequencies, the a-m rejection of the ratio detector depends upon the time constant of the stabilized rectified voltage. The same considerations which determine the time constant in an a-v-c circuit can also be applied to the determination of the time constant of the stabilizing voltage. In general, the discharge time constant of the stabilizing voltage, formed by the load resistance and the stabilizing capacitance should be about 0.2 second. Time constants larger than this have an undesirable effect on the tuning characteristic which is similar to the effect of too long a time constant in the a-v-c circuit of a receiver.

The manner in which the a-m rejection of a ratio detector varies with the frequency of the amplitude modulation is shown in Figure 26. This detector (see Figure 8) has a time constant of 0.1 second. The characteristics apply (a) to a balanced detector with the center of the load resistor grounded and (b) to the unbalanced form of the circuit which is the same as Figure 8, but with one end of the stabilizing condenser grounded. It should be noted that for the same time constant, the balanced circuit gives much better rejection of low audio frequencies. This is to be expected because in the case of the balanced circuit, fluctuations in the stabilizing voltage do not cause any disturbance in the output, whereas in the case of the unbalanced

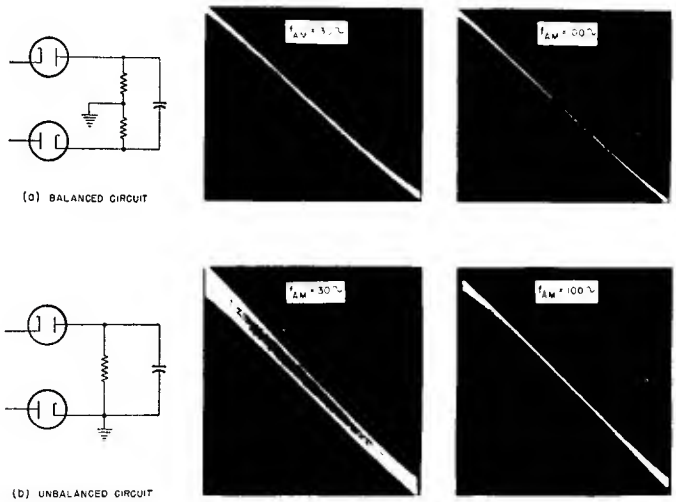


Fig. 26—A comparison of balanced and unbalanced circuits with respect to their effectiveness in rejecting low-frequency amplitude modulation. The time constant of the load circuit is the same for both.

circuit they do contribute to the output. For a given value of time constant in the stabilizing voltage, then, it follows that the balanced circuit will give better rejection of low-frequency amplitude fluctuations.

Tolerances

In general, the ratio detector circuit will require more attention to tolerances than is necessary for a balanced discriminator. The principal factor which must be controlled closely is the effective coupling between the primary and secondary winding, since this controls the ratio between the secondary and tertiary voltages, which in turn affects the a-m rejection. This coupling is determined not only

by the separation between the primary and secondary, but also by the capacitance unbalance from either end of the secondary to ground.

The ratio detector is essentially a balanced circuit and from its method of operation it is evident that the degree of a-m rejection will depend upon the two diodes having substantially the same characteristics. If the diodes differ materially from each other, the principal effect will be the production of an unbalanced a-m component in the output. The effect on the peak separation and f-m sensitivity is usually negligible. If the detector is tested with an f-m signal which is also amplitude modulated, so that oscillograms similar to those in Figure 12 are obtained, unbalanced diodes will have the effect of broadening the output characteristic which is indicative of an unbalanced a-m component in the output.

Measurements

The measurements necessary in the design of a ratio detector and in checking its performance are described in this section. These include the unloaded and loaded values of primary and secondary Q, the per cent critical coupling, the ratio of the secondary and tertiary voltages, and amplitude modulation rejection.

Primary and Secondary Q's

To measure the unloaded primary Q, the double diode is removed from its socket and the secondary is detuned. The primary Q is then determined from the selectivity of the primary tuned circuit. The loaded primary Q is measured in the same way but with the diodes in the circuit so that the normal diode loading is reflected across the primary through the tertiary winding.

To measure the secondary Q, the primary winding is heavily loaded by shunting a resistance across it. The secondary Q is then determined from the selectivity of the secondary winding as indicated by the rectified voltage variation with frequency. When measuring the unloaded secondary Q, the diode load resistance should be replaced with high values of resistance, on the order of a megohm. When measuring the loaded secondary Q, the normal load resistors should be used. For these measurements the centertap of the secondary should be disconnected from the high side of the tertiary winding, so as to prevent the tertiary voltage from contributing to the rectified voltage.

Coupling

The per cent of critical coupling can be measured by noting the

change in primary voltage as the secondary is varied from a tuned to a detuned condition. With all circuit conditions normal, and the signal applied at the center frequency, the primary voltage is noted. With the same input signal level, the secondary is detuned so that the primary voltage rises. The per cent critical coupling can then be expressed in terms of the ratio between the primary voltage with the secondary tuned and detuned. For example, if the signal at the plate rises 25 per cent when the secondary is detuned, the coupling is 50 per cent of critical.

Ratio S/P

The ratio between the secondary and tertiary voltages can be measured indirectly in terms of the rectified output voltage which is obtained (a) with the secondary tuned and (b) with the secondary detuned. In both cases the input signal is adjusted so that the primary voltage remains constant. If the ratio between the voltages read in (a) and (b) is r, that is, if

$$\frac{\text{Rectified voltage (secondary tuned)}}{\text{Rectified voltage (secondary detuned)}} = \frac{\sqrt{P^2 + S^2}}{P} = r$$

then

$$\frac{S}{P} = \sqrt{r^2 - 1}$$

Signal Generator

In checking the performance of ratio detectors it is convenient to have available a signal generator which is capable of being simultaneously frequency and amplitude modulated. The advantage of using this type of modulation is that it provides graphically a complete picture (Figure 12, for example) of the f-m response of the detector and at the same time shows the a-m rejection under dynamic conditions at every frequency within the range of the detector.

It is important that the signal generator used to check ratio detectors be free of spurious modulation. In particular, it is essential that when the signal generator is amplitude modulated, no frequency modulation be produced as a result of the amplitude modulation. If frequency modulation takes place when the generator is amplitude modulated, the resultant output will be construed as being due to faulty a-m rejection whereas it is actually the normal f-m response of the detector.

Alignment Procedure

The ratio detector may be aligned by using either an unmodulated signal at the center frequency and a d-c vacuum-tube voltmeter, or by using an f-m signal generator and an oscilloscope.

If an unmodulated signal generator is used, the procedure is to set the signal generator to the center frequency. With the d-c vacuum-tube voltmeter connected to measure the rectified output voltage (this is usually at the a-v-c take-off point), the primary tuning is adjusted for maximum voltage.

The procedure used for adjusting the secondary tuning depends upon whether the center-tap of the stabilizing voltage is grounded or whether one end is grounded. If the center-tap is grounded, the secondary tuning is adjusted so that the d-c voltage at the audio take-off point is equal to zero; the d-c vacuum-tube voltmeter is shifted to the audio take-off point for this measurement. If the stabilizing voltage is grounded at one end (see Figure 25) rather than at the center, the secondary tuning is adjusted so that the d-c voltage at the audio take-off point is equal to half the rectified voltage.

If sweep alignment is used, the primary can be accurately aligned by using a comparatively low deviation and adjusting it for maximum amplitude on the screen. The secondary may be adjusted by using a deviation such that the total frequency swing (twice the deviation) is equal to the peak separation. This procedure makes it possible to adjust the secondary tuning so that a symmetrical detector characteristic is obtained.

Peak Separation

The separation between the peaks on the f-m output characteristics of a ratio detector may be measured most simply by applying an f-m signal and increasing the deviation until the response is just observed to flatten at the peaks. When this is done, the peak separation is equal to twice the deviation.

If an attempt is made to measure the peak separation by plotting the output characteristic point by point, the peak separation obtained will usually be considerably less than is obtained under dynamic conditions with the output voltage stabilized. The f-m detector characteristic may be plotted point by point provided a battery of the proper voltage is connected across the stabilizing condenser. The voltage of this battery must be equal to the rectified voltage which exists at the center frequency. In practice it is convenient to use a 7.5-volt C-battery and to adjust the signal input so that the condenser-stabilized voltage at the center frequency is equal to the battery voltage.

Measurement of A-M Rejection

The measurement of a-m rejection can be carried out by means of a signal generator which can be simultaneously frequency and amplitude modulated. The type of pattern obtained when there are residual balanced and unbalanced components of amplitude modulation in the output has been described in connection with Figure 23. Typical oscillograms are shown in Figure 12.

The measurement of the a-m rejection of a ratio detector can be described in terms of the pattern obtained for a given frequency deviation, i.e., for a given per cent of frequency modulation, and for a given per cent of amplitude modulation. Ideally the pattern obtained should remain a diagonal line regardless of the presence of the amplitude modulation. The extent to which this pattern departs from a straight line indicates to what extent the detector fails to reject amplitude modulation. While it would be desirable to be able to quote a single figure of merit to indicate the a-m rejection, it is essential to check the performance at various levels and to interpret the wedge-shaped patterns which are obtained so as to determine the extent of the residual balanced and unbalanced components of amplitude modulation in the output.

In addition to the method of using a signal which is simultaneously frequency and amplitude modulated, it is possible to make a-m rejection measurements by using a generator which is only frequency modulated. To measure a-m rejection using this method, an f-m signal is applied at full deviation corresponding to 100 per cent modulation, and a battery equal in value to the rectified output is shunted across the stabilizing condenser. This battery will have negligible effect on the f-m output. Now the input signal is reduced in value until the f-m output becomes distorted as a result of the diodes being biased by the stabilizing voltage, and the f-m output drops to zero. The ratio between the initial input signal and the minimum input signal for which the f-m output becomes distorted is then a measure of the amount of downward modulation that the detector can reject. For example, if this ratio is r, then the per cent downward modulation which can be handled is $100\ (r-1)/r$. The performance of a typical ratio detector as measured by this method has been shown in Figure 11. To determine the rejection of the detector for upward amplitude modulation, the same setup is used, and the change in the output is noted as the input signal is increased.

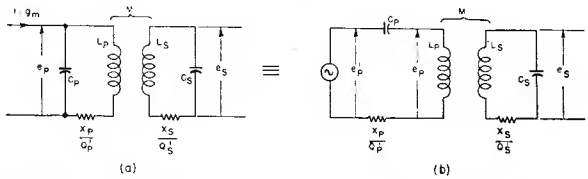
Since the amplitude of the input signal is not varied dynamically, this method will not indicate any unbalanced component which may

be present in the output. The latter is more conveniently measured by using simultaneous frequency and amplitude modulation as previously described.

Appendix

Formulas for S/P and S

The fundamental circuit under consideration is shown below in (a).



By Thévenin's theorem it is equivalent to (b), where

$$e_p^1 = i X_p = g_m X_p$$

At the center frequency,

$$i_p = \frac{e_p^1}{\frac{X_p}{Q_p^1} + \frac{\omega^2 M^2 Q_s^1}{X_s}}$$

But $\omega^2 M^2 = K^2 X_p X_s = \frac{K^2}{K_c^2} K_c^2 X_p X_s$

$$\omega^2 M^2 = \alpha^2 \frac{X_p X_s}{Q_p^1 Q_s^1}$$

$$\therefore\ i_p = \frac{g_m X_p}{\frac{X_p}{Q_p^1} + \alpha^2 \frac{X_p}{Q_p^1}}$$

$$i_p = \frac{g_m Q_p^1}{1 + \alpha^2}$$

$$e_s^1 = \omega M\, i_p$$

$$e_s^1 = \alpha \sqrt{\frac{X_p X_s}{Q_p^1 Q_s^1}} \quad \frac{g_m Q_p^1}{1+\alpha^2}$$

$$e_s = Q_s^1 e_s^1$$

$$e_s = g_m \sqrt{X_p X_s} \sqrt{Q_p^1 Q_s^1} \frac{\alpha}{1+\alpha^2}$$

$$e_p = i_p X_p = \frac{g_m Q_p^1 X_p}{1+\alpha^2}$$

$$\therefore \frac{e_s}{e_p} = \alpha \sqrt{\frac{X_s Q_s^1}{X_p Q_p^1}}$$

To find Q_p^1, the operating Q of the primary, refer to Figure 14.

$$Q_p^1 = \frac{Q_p^0 \cdot \frac{n^2 R}{4}}{Q_p^0 X_p + \frac{n^2 R}{4}}$$

$$Q_p^1 = \frac{\frac{n^2 R}{4a \cdot \frac{X_p}{a}}}{1 + \frac{n^2 R}{4a Q_p^0 \cdot \frac{X_p}{a}}} \cdot \frac{1}{} \qquad \text{where } a = \frac{X_p}{X_s}$$

$$Q_p^1 = \frac{n^2 Q_s^1}{4a} \cdot \frac{1}{1 + \frac{n^2 Q_s^1}{4a Q_p^0}}$$

$$\therefore \frac{e_s}{e_p} = \alpha \left\{ \frac{L_s}{a L_s} \cdot \frac{4a}{n^2} \left(1 + \frac{n^2 Q_s^1}{4a Q_p^0} \right) \right\}^{1/2}$$

Since $S = \frac{e_s}{2}$ and $P = \frac{e_p}{n}$

$$\frac{S}{P} = \alpha \sqrt{1 + \frac{n^2 Q_s^1}{4a\, Q_p^0}}$$

To find S, substitute in the expression for e_s

$$S = \frac{g_m}{2} \left\{ a X_s \cdot X_s \frac{n^2 Q_s^1}{4a} \cdot \frac{1}{1 + \frac{n^2 Q_s^1}{4a\, Q_p^0}} \right\}^{1/2} \frac{\alpha}{1 + \alpha^2}$$

$$= \frac{g_m X_s Q_s^1 n}{4} \cdot \frac{1}{\sqrt{1 + \frac{n^2 Q_s^0}{4a\, Q_p^0}}} \cdot \frac{\alpha}{1 + \alpha^2}$$

But $$\frac{S}{P} = \alpha \sqrt{1 + \frac{n^2 Q_s^1}{4a\, Q_p^0}}$$

$$\therefore S = \frac{g_m X_s Q_s^1 n}{4} \left(\frac{P}{S} \right) \frac{\alpha^2}{1 + \alpha^2}$$

THE DEVELOPMENT OF A FREQUENCY-MODULATED POLICE RECEIVER FOR ULTRA-HIGH-FREQUENCY USE*†

By

H. E. Thomas‡

RCA Manufacturing Co., Inc., Camden, N. J.

Summary

This paper first describes the general considerations bearing upon design of a frequency-modulated mobile police receiver for use in the 30- to 40-megacycle band where the channel width is restricted to 40 Kilocycles. The details of developing the various circuits around these considerations are dealt with, and the paper then proceeds with a discussion of over-all performance and field testing. Comparative and quantitative results show very favorable performance using a double-superheterodyne circuit with automatic-frequency control of the second oscillator.

(12 pages; 7 figures; 2 tables)

* Decimal Classification: R630.3.

† *RCA REVIEW*, October, 1941.

‡ Now with Engineering Products Department, RCA Victor Division, Camden, N. J.

DEVELOPMENT OF AN ULTRA LOW LOSS TRANSMISSION LINE FOR TELEVISION*†

By

E. O. Johnson

Engineering Products Department, RCA Victor Division, Camden, N. J.

Summary

The development of a low loss 300-ohm parallel wire polyethylene dielectric transmission line is described. Loss curves, as well as a photograph of a production run sample of the line, are included. The disadvantages of twisted pair and coaxial lines are discussed and the developmental specifications for a line to eliminate these disadvantages are set forth. The remainder of the paper deals with actual development of the line and treats with the dielectric, line conductors, and line impedance. Results of field test are included. This line is also used for FM reception.

(9 pages; 4 figures)

* Decimal Classification: R320.41.

† *RCA REVIEW*, June, 1946.

INPUT CIRCUIT NOISE CALCULATIONS FOR F-M AND TELEVISION RECEIVERS*†

By

William J. Stolze

Industry Service Laboratory, RCA Laboratories Division,
New York, N. Y.

Summary

Efficient design of input stages, a critical requisite in FM and television, involves a careful consideration of three important factors: total noise, sensitivity and signal-to-noise ratio. This paper discusses these factors and offers a physical concept of the three noise components with formulae and data necessary for input-circuit calculations.

(5 pages; 10 figures)

---

* Decimal Classification: R361.211 x R630.1.

† *Communications*, February, 1947.

DUPLEX TRANSMISSION OF FREQUENCY-MODULATED SOUND AND FACSIMILE*†

By

MAURICE ARTZT
RCA Manufacturing Company, Inc., Camden, N. J.

AND DUDLEY E. FOSTER
RCA License Laboratory, New York

Summary—Laboratory and field tests of multiplexing aural and facsimile programs on a frequency-modulated wave have been made. The results obtained show that while technically possible, difficulties are encountered that will probably make such operation undesirable. Enough of the theory of multitone modulation is discussed to show that circuit linearity requirements are severe, and that the side bands can easily exceed the 200-kilocycle channel width if the aural deviation is not reduced. Quantitative data are given on the types and percentages of cross-talk obtained.

1. INTRODUCTION

THE F.C.C. rule defining requirements of high-frequency, aural and facsimile broadcasting stations is as follows:

"S 3.228 Facsimile Broadcasting and Multiplex Transmission. The Commission may grant authority to a high-frequency broadcast station for the multiplex transmission of facsimile and aural broadcast programs provided the facsimile transmission is incidental to the aural broadcast and does not reduce the quality of or the frequency swing required for the transmission of the aural program. The frequency swing for the modulation of the aural program should be maintained at 75 kilocycles and the facsimile signal added thereto. No transmission outside the authorized band of 200 kilocycles shall result from such multiplex operation nor shall interference be caused to other stations operating on adjacent channels."

Little practical experience has yet been obtained with multiplex operation of the kind called for in this regulation. For this reason it has seemed worthwhile to report the results of some laboratory and field tests carried out by RCA in Camden and New York. These tests were made with the object of determining first, if any limitations and practical difficulties were likely to be encountered, and second, to what extent the introduction of multiplex facsimile signals might effect the design of receivers for aural service only.

It is generally conceded that in such a multiplex system the signal-to-noise level in the aural program should be favored as much as possible. In order to maintain and protect adequately the high signal-to-noise ratio of the aural service, the facsimile signal must use a much

* Decimal Classification: R581 × R630
† Reprinted from *RCA REVIEW*, July, 1941.

lower deviation. The ability of the facsimile equipment to operate on a much less favorable signal-to-noise ratio then becomes increasingly important. The facsimile equipment used in these tests was designed to operate under conditions heretofore considered too severe for good reproduction, and the final results obtained in the multiplex operation should therefore be considered on the basis that all possible favoring has been given to the aural channel. When used with a facsimile system requiring a higher signal-to-noise ratio, the facsimile would in general become less reliable and reproduction poorer.

Practical difficulties were encountered in the field tests described later that limited the frequency deviations used for the aural portion of the multiplexing to 60 kilocycles. The facsimile service was added at 15 kilocycles deviation, thus sharing the maximum deviation of 75 kilocycles between the two services. Many things contributed to this limitation of aural deviation and each will be described in turn.

Before going to the tests and their results a description of the facsimile signals used in these tests will be given. A general discussion of the frequency-modulation theory for multitone use is also in order if the proper interpretation of the results is to be made.

2. The Facsimile Signals

In the usual broadcast facsimile services contemplated or now in use, copy areas of from four to twenty square inches per minute are transmitted. A good average at present would be 10 square inches per minute, the figure chosen for these tests. This was transmitted at 120 scanning lines per inch, using a picture signal frequency range of from zero to 1200 cycles. This signal can be transmitted in a variety of ways, the usual one being to carry the signals as an amplitude modulation of an audio tone. However, previous facsimile work has shown that the use of a frequency-modulated sub-carrier for the facsimile signals has many advantages over the amplitude-modulated sub-carrier.[1] The sub-carrier chosen in this case was 18 to 24 kilocycles, the picture signals occupying the band of 20 to 24 kilocycles and the frequency being shifted down to 18 kilocycles for a short time once per scanning line to indicate phasing. The spectrum then is as follows:

18 kilocycles—phase signal
20 kilocycles—white picture signal
22 kilocycles—50/50 gray picture signal
24 kilocycles—black picture signal

The use of a phase signal to hold synchronism of the recorder is usually insufficient if very fine detail of recording is expected, and a

[1] Mathes and Whitaker, "Radio Facsimile by Sub-Carrier Frequency Modulation," RCA Review, Oct. 1939.

secondary synchronizing signal to control motor speed between phase pulses is desirable. A simple method of accomplishing this is to transmit a control tone (60 cycles in this case) as an amplitude modulation super-imposed on the already frequency-modulated sub-carrier. An amplitude-modulation of 30 per cent was used in this case, and the 60-cycle wave is thus transmitted without requiring a third multiplex channel and attendant filtering problems.

This complex wave of 18- to 24-kilocycle frequency variation and 30 per cent amplitude modulation then carries the picture, phasing, and synchronizing signals. When used in a simplex service and with the proper limiters and filters, these signals have been found to be practically immune to any normal signal fluctuations, and to reject all

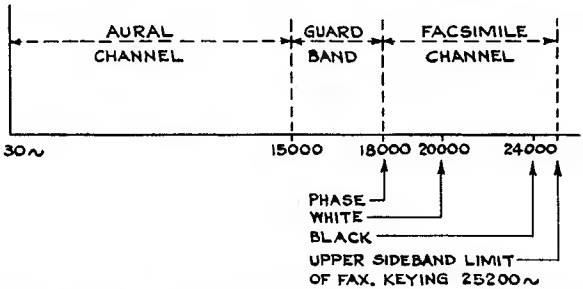


Fig. 1—Audio spectrum for multiplexed aural and facsimile.

noise 6 decibels or more below the signal itself. Occasional interferences equal to or even greater than the signal usually cause little marring of the copy.

This composite signal was then used to deviate the r-f carrier 15 kilocycles, the aural channel being added at 60 cycles deviation. The overall audio spectrum transmitted is then represented in Figure 1.

3. Multiplex Theory

In tone multiplexing on an amplitude modulated system, the addition of the second tone produces an additional pair of side bands displaced from the central carrier by the tone frequency itself. All side bands are proportional to the carrier amplitude and depth of modulation. In multitone frequency modulation the relationship of side bands is not as simple as this, for each tone in itself produces an infinite series of side frequencies. The case for two tones involves two such series of side frequencies, and their cross beat products. A derivation of the side bands produced in such a two-tone system is given by Crosby[2] and can be readily applied to this type of multiplexing.

[2] Murray G. Crosby, "Carrier and Side Frequency Relations with Multitone Frequency or Phase Modulation," RCA Review, July 1938.

Applying the Bessel function expression to the facsimile signal only, with no aural signal, gives the plot in Figure 2 of the relative amplitudes of carrier and side bands to transmit this signal with deviations up to 25 kilocycles. It can be seen that for deviations below 15 kilocycles the second-harmonic side band (J_2) is small and can probably be neglected without seriously affecting the facsimile signal itself.

The radio-frequency spectrum for the facsimile signal when operating simplex is shown in Figure 3. Suppose now that an aural tone is introduced and with such a low-frequency period that any small section of the wave may be considered as constant in frequency com-

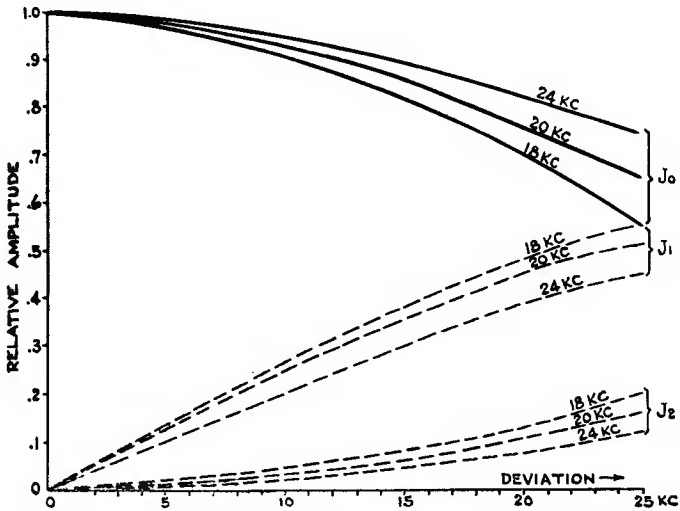


Fig. 2—Facsimile sideband amplitudes (No aural).

pared to the rapid deviation caused by the facsimile signal. Figure 4 will then represent the wave. This is strictly true only when the aural tone introduced has zero frequency, for frequency modulations cannot be added directly; but for frequencies in the aural band up to several hundred cycles there is very little error. If only the first-harmonic side bands of the facsimile signal (± 20 kilocycles) are considered, the band width under this condition will be ± 60 ± 20 or ± 80 kilocycles. If the second-harmonic side bands of ± 40 cycles are included, the width is ± 100 kilocycles.

With an audio frequency of 3125 cycles having a frequency deviation of 75 kilocycles applied alone to the frequency modulator the spectrum appears as in Figure 5 for frequencies on one side of the carrier, the side bands for frequencies on the opposite side of the carrier being similar.

At an audio frequency of 2500 cycles the resultant argument (Deviation divided by modulating frequency) comes within the range of the usual Bessel tabulations, and the sidebands of the multiplex modulation can be determined. Thus with a 2500-cycle tone at 60 kilocycles deviation and the 20,000-cycle tone at 15 kilocycles deviation the amplitudes can be plotted as in Figure 6. The first chart gives the side bands of the audio tone only, and the second the side bands for the combination of both tones. In both cases all side bands and side frequencies are plotted as positive quantities, but due account of sign of the coefficients has been taken where several terms of the expansion are involved for a particular frequency. In these charts any coefficient greater than 0.1 per cent was included.

It can be seen that the bandwidth in Figure 6 is somewhat greater than in Figure 5, but that the side bands outside the 200-kilocycle channel have less than 1 per cent amplitude.

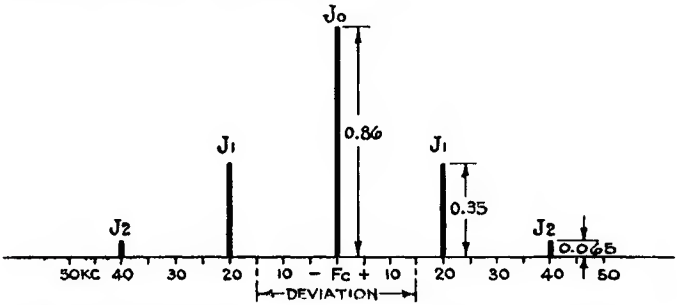


Fig. 3—Sideband chart of 20-kc signal at 15-kc deviation.

By comparing Figure 4 with Figures 5 and 6, it may be seen that the low aural frequencies cause a greater side band extension than the higher aural frequencies. For the higher frequencies, the product of the Bessel coefficients reduces the amplitude of the higher order side bands to a degree where the bandwidth of consequence becomes less than for a single modulating tone.

The 60-kilocycle deviation for aural signals was used in the multiplex calculations rather than 75 kilocycles, in order to keep the signal within the 200-kilocycle band width. As can be readily seen from Figure 5, maintaining the aural deviation at 75 kilocycles and using the 15-kilocycle facsimile deviation would produce side bands outside the 200-kilocycle channel. Reducing the facsimile deviation to 10 kilocycles would lower the amplitude of the out-of-channel side bands to 64 per cent of their former value. This reduction would be insufficient to clear up the out-of-channel radiation, and furthermore would handicap the facsimile in signal-to-noise ratio. Thus Figure 5 shows the first practical reason for reducing the aural deviation to 60 kilocycles when multiplexing with facsimile.

Before going to the experimental results it would be in order to point out the types of cross-talk that can occur if the equipment fails to pass this entire band as represented in Figure 6. If the bandwidth of the single-tone transmission is restricted at the extreme ends, only harmonic distortion of the resulting detected tone will be present, and a few per cent of distortion of this type will be little noticed. This same amount of restriction on the two-tone transmission will result quite

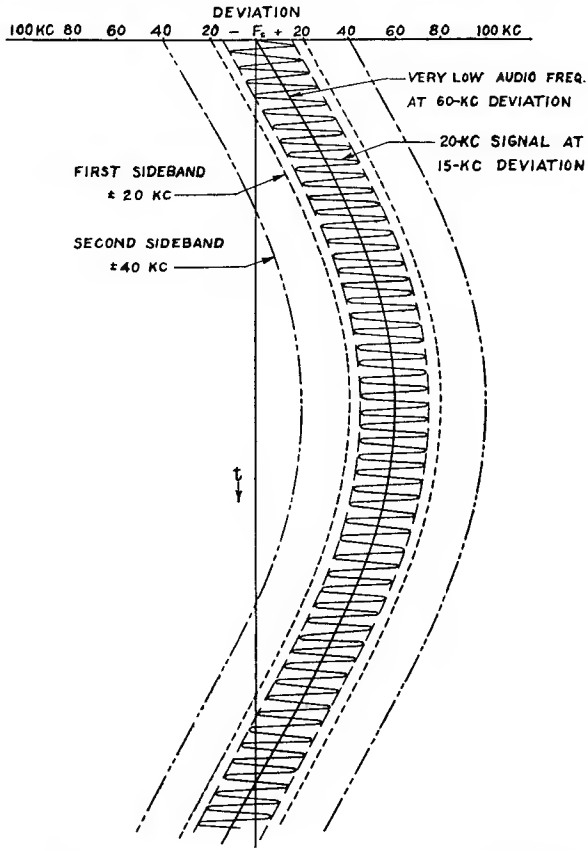


Fig. 4—Facsimile signal of Figure 3 added to very low frequency signal at 60-kc deviation.

differently, for such band limiting will act as a preliminary frequency discrimination and change various side-band frequency modulations to amplitude modulations. This will introduce cross-talk in a manner similar to the action of a non-linear tube in amplitude multiplexing. It should be pointed out that a small percentage of cross-talk is far more objectionable than the same percentage of harmonic distortion.

The process of detecting a frequency-modulated wave requires dif-

ferentiating[3] the wave to change the frequency variations into an amplitude-modulated wave. Amplitude detection follows this differentiation (or discrimination) to obtain the amplitude envelope. If the waves represented by Figures 4 and 6 are differentiated, the original modulation frequencies are obtained intact. However, if one or more of the terms in these side frequencies are eliminated before differentiating, the differentiated (or discriminated wave) tone will contain terms proportional to the derivative of these missing terms. Any two of these waves can then beat together in the amplitude detector that follows this discriminator to give cross talk.

Bandwidth limitation could occur in the transmitter itself, but the more likely place will be in the receiver intermediate-frequency amplifier or discriminator circuits. This is especially true in receivers hav-

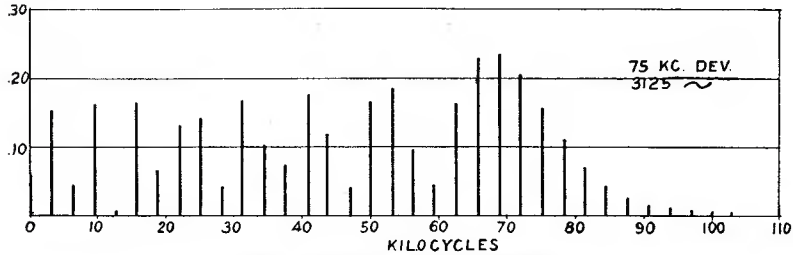


Fig. 5—Sideband amplitude.

ing good adjacent channel selectivity and consequent narrow band-pass action in the i-f transformers.

After detecting, the requirements of linearity in any tube that both aural and facsimile signals pass through together are the same as for amplitude multiplexing.

In the case of two modulating frequencies which differ in character as greatly as do those of sound and facsimile programs the linearity requirements imposed on the system are much more severe than in the case of single program modulation. With a single program, any non-linearity of the modulator, the detector, or the audio-frequency portion of the receiver causes harmonic distortion. Distortion may also be caused by too narrow a receiver passband. This latter distortion will be decreased appreciably, but not entirely eliminated by a limiter. All of these causes of distortion may be classed as non-linearity.

The distortion which results from non-linearity in the case of a single program, if not too great, is not readily distinguished by the ear. For example, there are many radio receivers having 3 per cent or even 5 per cent distortion which are not objectionable to the major-

[3] Carson and Fry, "Variable Frequency Electric Circuit Theory With Application to the Theory of Frequency Modulation," *Bell System Tech. Jour.*, Oct. 1937.

ity of listeners. In the case of modulation by two different types of program, however, non-linearity anywhere in the system results in combination tones or cross modulation. The cross modulation of one program by the other has the characteristics of interference which is readily noticed by the listener. An interfering program 30 db below the desired program is easily apparent, whereas distortion of a single program 30 db below (approximately 3 per cent of) the fundamental, is seldom noticed.

Consequently, duplex frequency-modulation transmission of sound and facsimile requires a somewhat greater bandwidth than a sound

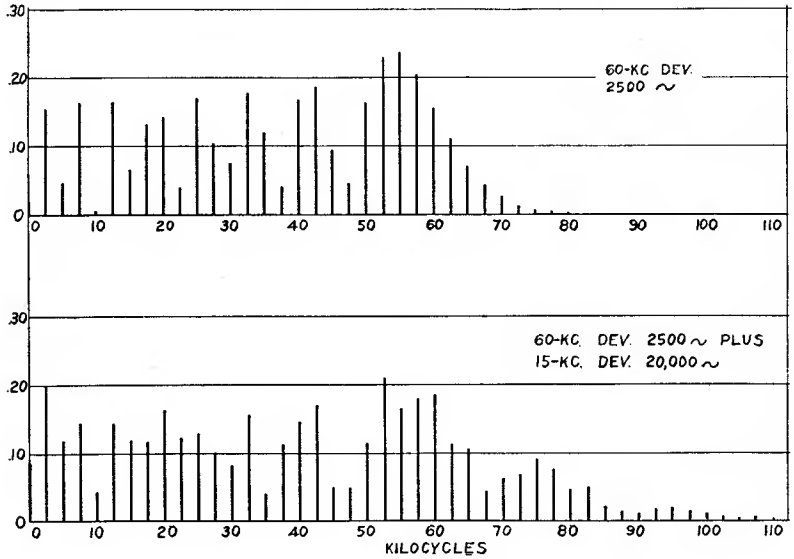


Fig. 6—Sideband amplitude.

program alone of comparable deviation, particularly for low modulating frequencies, and imposes much more stringent requirements on system linearity.

4. Duplex Receiver

The receiver for duplex transmission of sound and facsimile, in addition to the requirement of exceedingly linear circuits, must be provided with filters for separating the two programs following the detector.

A block diagram of a receiver constructed for tests of duplex transmission is shown in Figure 7. Two selective circuits were used ahead of the converter and an i.f. of 8.25 Mc was employed to minimize likelihood of spurious responses. The i-f passband was made somewhat wider than normal to guard against non-linearity from that source.

It had a response characteristic down 14 db at 100 kc each side of the center and down 37 db at 200 kc each side of the center.

The detector characteristic was more linear than is usually the case, having very small departure from a straight line up to 100 kc on each side of the center.

It was found necessary to separate the sound and facsimile programs immediately following the detector, as passing both through a single common amplifier stage, even a fully degenerated stage, resulted in excessive cross modulation. The de-emphasis filter in the output of the detector attenuated the facsimile frequencies sufficiently so that cross-modulation did not occur in the succeeding stage. An additional low-pass filter was used following the first audio-frequency

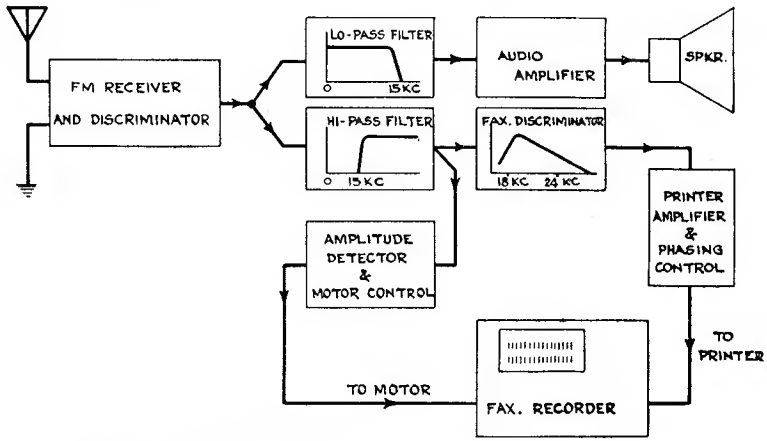


Fig. 7—Receiver block diagram for multiplex operation of aural and facsimile.

amplifier since the impedance at that point could be made low by use of a cathode load and a better filter designed for the resultant low impedance than for the high impedance at the output of the detector. The audio program was thus subjected to proper compensation for the high-frequency pre-emphasis at the transmitter and also had an additional filter to attenuate further any facsimile frequencies in the sound channel. The resultant attenuation for facsimile frequencies in the sound channel was greater than 45 db for even the lowest (18 kc) facsimile frequency.

High-pass filters to attenuate the sound program in the facsimile channel were provided in the facsimile equipment.

A high-quality, Type 64B speaker was used with the receiver and was fed by a pair of 6L6G tubes in push-pull so that the reproduction of the sound program was of high fidelity.

5. Laboratory Tests of Duplex Transmission

The first laboratory tests were of a preliminary nature and used a special receiver (with audio circuits similar to those in Figure 7) and a signal generator set at 43.5 megacycles. The facsimile signal was as previously described, and the aural channel was fed by a beat oscillator or from a phonograph pickup. In these tests the receiver bandwidth was slightly higher than 200 kilocycles and good results were expected with the full aural deviation of 75 kilocycles and a facsimile deviation of 15 kilocycles. A tabulation of the results obtained follows:

Audio Freq.	*Audio Deviation*	*Facsimile Deviation*	*Interference*	
			On Aural	*On Facsimile*
10 kc	75 kc	15 kc	— 25 db	Prints Interference Synch. Erratic
5 kc	75 kc	15 kc	— 69 db	Prints Interference Synch. Erratic
400 cy	75 kc	15 kc	— 70 db	None Noticeable
*400 cy	75 kc	15 kc	— 58 db	Prints Interference Synch. Erratic

* Receiver detuned 18 kc.

These results show that only at the low frequencies of aural tone (400 cycles in this case) were the cross-beat products low enough substantially to eliminate all cross-talk. When high aural frequencies were used (5 and 10 kilocycles), the combination frequencies increased enough to cause deterioration of the signal-to-noise ratio in both channels. It was to be expected that the facsimile channel would become unusable first, for the cross-beat interference tones were of a high frequency and would come through the high-pass filter network feeding the facsimile amplifiers.

Two general observations can be made from these tests: First, the use of pre-emphasis of higher aural frequencies will occasionally bring enough interference into the facsimile channel to mar the recorded copy if the full 75-kilocycle deviation is used for aural programs. Second, only a very wide band receiver could receive such a signal combination without cross-talk. Any reduction in bandwidth in the receiver will aggravate the interference in either of the above cases and a reduced bandwidth would probably be used in the average receiver to obtain a fair degree of adjacent channel selectivity.

The most apparent cure for these difficulties was to reduce the aural deviation to 60 kilocycles, and this was done in all subsequent tests.

A further test was made in the laboratory to simulate duplex trans-

mission of sound and facsimile. In this case a 4000-cycle tone was used for the audio frequency and a 17,000-cycle tone for the facsimile frequency.

Simultaneous application of two tones to a system having any non-linearity results in combination tones that are the sum and difference of multiples of the individual tones. For example, if two tones having frequencies a and b are applied to a non-linear element, the result will be frequencies $2a$, $2b$, $3a$, $3b$, $2a - b$, $2a + b$, $3a - b$, $3a + b$, $2b - a$, $2b + a$, $3b - a$, $3b + a$, etc. In general, in a non-linear system the coefficients of the higher-order terms are much smaller than those of the lower-order terms so that the greatest amplitude of resultant combination tones is likely to occur with a high audio frequency and a low facsimile frequency. However, audio frequencies much above 4000 or 5000 cycles seldom have sufficient energy, even when pre-emphasis is used, to modulate the transmitter fully. Consequently interference between sound and facsimile is most likely for audio frequencies in the 2000 to 5000-cycle range.

The test was performed by modulating a laboratory frequency-modulation signal generator by 4000 cycles and 17,000 cycles simultaneously and measuring the resultant frequencies in a dummy load replacing the voice coil in the output of the receiver by means of a wave analyzer. Throughout the series of tests the input to the receiver was 50 microvolts and the output was held constant by means of the volume control at 0.5 watt. In all cases, except where otherwise noted, tuning was performed by applying approximately 150-kc deviation and the receiver tuned to symmetrical distortion of the top and bottom of the wave as observed on an output oscilloscope.

As an initial test, the 4000-cycle tone alone was applied and the harmonics read as the deviation was varied.

TABLE I
4000-cycle modulation

Deviation	*2nd*	*3rd*	*4th*
75 kc	7.5%	3.1%	0.25%
60	3.6	1.2	0.14
40	1.98	0.62	0.13
20	1.85	1.1	0.10

With the exception of the second harmonic with 75-kc deviation which was believed due to the signal generator curvature, the harmonic content was low. In any event the second harmonic would be due to a second-order term and with 4000- and 17,000-cycle tones, the second-order term would give a 13,000-cycle resultant, and the ear is much less responsive to that frequency than to frequencies of

the order of 1000 or 2000 cycles. If we designate the 4000-cycle tone as a and the 17,000-cycle tone as b, the resultant tones up to the fifth order, would be $b - a$, $b - 2a$, $b - 3a$, $b - 4a$, or 13,000 cycles, 9000 cycles, 5000 cycles and 1000 cycles respectively. The amplitude of these resultant frequencies was measured in the output of the receiver with the two tones modulating the signal generator.

In the second test, shown in Table II, the 4000-cycle tone had a constant deviation of 60 kc and the deviation of the 17,000-cycle tone was varied from 5 to 15 kc. The amplitude of 4000-cycle output was taken as 100 per cent.

TABLE II

4000 cycles; constant deviation, 60 kc

17,000-cycle Deviation	*4000*	*1000*	*5000*	*9000*	*13,000*
15 kc	100%	4.0%	5.1%	18.5%	1.0%
10	100	2.1	5.0	15.0	1.0
5	100	3.1	5.1	13.7	1.5

Variation of deviation of the 17,000-cycle component is seen to have little effect on most resultant frequencies.

The results when the 17,000-cycle tone deviation was kept constant at 15 kc and the deviation of the 4000-cycle tone varied are shown in Table III.

TABLE III

17,000 cycles; constant deviation, 15 kc

4000-cycle Deviation	*4000*	*Output frequencies, cycles*			
		1000	*5000*	*9000*	*13,000*
60 kc	100%	3.1%	5.1%	13.7%	1.5%
40	100	1.9	4.6	12.2	1.1
20	100	1.2	1.2	8.3	0.46

Decreasing the deviation of the 4000-cycle component is seen to decrease materially the magnitude of the resultant components.

With 60-kc deviation of 4000 cycles and 15-kc deviation of 17,000 cycles applied, the receiver was then tuned to minimize the audible resultant tones by listening. This tuning was found to be very critical and to depart by about 20 kc from the tune position determined with a single modulating tone. The resultant outputs were then:

4000 cycles	*1000*	*5000*	*9000*	*13,000*
100%	0.62%	3.0%	9.7%	4.2%

This method of tuning the receiver is seen to reduce greatly the

resultant components, especially 1000 cycles, which was the most prominent in listening. Even with the receiver so tuned, the output to the ear was readily evident as being different from pure 4000 cycles.

The results of these laboratory tests confirm the theoretical considerations that two tones applied to a system having any non-linearity whatsoever will result in combination frequencies exceeding in amplitude the harmonics which occur with a single modulating tone. When these two tones differ widely in frequency, the low modulating frequencies, while producing greatest side-band extent, produce combination tones outside the audible range, hence are less disturbing to the sound program. The higher-frequency audio tones, combining with the facsimile frequencies, however, do produce serious interference in the sound channel.

6. Field Tests

A series of tests was run on the system of duplex sound and facsimile on W2XWG (W51NY) frequency-modulation transmitter located on the Empire State Building in New York City during January 1941. The receiver used in the tests was the one described above and was tried at three different locations in Radio City, about one mile from the transmitter.

The facsimile and sound programs were applied in parallel to the grid of the reactance tube modulator of the f-m transmitter and were independently monitored as to deviation. The response of the facsimile system was flat from scanner to transmitter within 0.5 db up to 24 kc.

With careful tuning of the receiver and limitation of the deviation to 60 kc for the sound and 15 kc for the facsimile, substantial freedom from interference between the two was possible.

Even with these conditions, a sound program with prominent sibilants produced a short skip (white space) in the facsimile copy for each "s" sound, and during pauses in sound programs the facsimile could be heard in the background.

If the receiver was not precisely tuned, little increase in interference with facsimile was apparent, but the facsimile background became markedly evident in the sound reproduction.

During the progress of the test one practical operating factor was noticed that illustrated the interdependence of sound and facsimile in duplex transmission. There was a tendency for unwarned individuals to tune to a different sound program for check purposes while receiving facsimile. This caused the facsimile to drop out of synchronism which had to be regained upon returning to W2XWG, so that about ¼ inch of copy was lost for even momentary tuning to a different station.

7. Conclusions

The laboratory and field tests indicated that the frequency-shift facsimile method was capable of producing good copy and could be

synchronized by a simultaneous amplitude modulation of the picture frequency modulation. This resultant signal could also be used to frequency-modulate a transmitter at the same time as a sound program to give duplex transmission. The required spectrum space for duplex operation is greatest for very low modulating frequencies, and for such frequencies, is likely to exceed the channel width.

The test indicated several practical difficulties which may be expected in duplex sound and facsimile operation.

1. Necessity for adequate filtering to separate signals before passing through any audio amplifier. This could be done on receivers designed for duplex operation, but would be necessary on all f-m receivers whether designed for facsimile reproduction or not, if cross modulation of the sound by the facsimile is to be avoided. Thus, receivers designed for sound reproduction only would require design provisions not now needed.

2. Necessity for continuous uninterrupted reception of an undesired sound program to obtain a desired facsimile program and vice versa.

3. Necessity for tuning accuracy beyond the capabilities of the average user for acceptable freedom from cross-talk.

4. Necessity for unusually high degree of oscillator frequency stability to maintain this tuning accuracy.

5. Necessity for unusually high degree of linearity in any circuit carrying both sound and facsimile signals.

These considerations result in the opinion that duplex operation of f-m sound and f-m facsimile is technically possible, but has certain difficulties both technical and commercial which would appear to make it undesirable.

The two services have too many conflicting requirements in receiver design and sale, in program design and in user's desire to change one, but not the other, to be tied together thus intimately. It would appear more desirable to operate with independent modulation of sound and facsimile on a time division basis, the facsimile being transmitted at a period when the station is not in use for sound transmission.

The authors are indebted to W. A. R. Brown and T. J. Buzalski of the National Broadcasting Company for furnishing and operating the transmitter facilities for the field tests, and to W. L. Carlson, V. D. Landon, and H. Kihn of the RCA Manufacturing Company for a part of the laboratory tests described. Credit is also due J. A. Rankin of the RCA License Laboratories for able assistance in the field and laboratory tests, and to M. G. Crosby of R.C.A. Communications for his helpful discussions of the theoretical aspects of the problem.

USE OF SUBCARRIER FREQUENCY MODULATION IN COMMUNICATION SYSTEMS*†

By

Warren H. Bliss‡

RCA Communications, Inc., Terminal Facilities Research Group, New York, N. Y.

Summary—*When subcarrier frequency modulation having a frequency range of 1600 to 2000 cycles was used for transoceanic facsimile transmission, pictures were obtained with finer detail and better half-tone quality than those transmitted by previous systems; the speed of transmission could also be increased. An extension of the system to a two-way multiple-channel radio-relay circuit providing teletype service between New York and Philadelphia gave improved stability of operation when variations in signal strength occurred.*

General

As THE art of frequency modulation has developed, it has naturally produced many by-products and has found applications and uses not at first apparent to radio engineers. One of these is in the field of point-to-point communication where an application known as subcarrier frequency modulation has been investigated and found to have important advantages. When a subcarrier wave is frequency-modulated by signal material and is then used to modulate a primary carrier, the system is referred to as subcarrier frequency modulation.

Two applications of subcarrier frequency modulation which are highly successful or show definite promise are (a) the transmission of facsimile material over long radio circuits and (b) the transmission of telegraph characters (printer-type signals in particular) over short-wave radio-relay circuits. The former has been used by RCA Communications, Inc., for over two years with results far surpassing those of any other method.[1] Experimental tests recently conducted with the latter prove there is a definite advantage in stabilizing teletype signal levels when the different tones from multiple-channel operation are combined together to frequency modulate a common subcarrier wave. In both of these cases the major advantage is due to the great reduction

---

* Decimal Classification: R581 × R630.

† Reprinted from *Proc. I.R.E.*, August, 1943.

‡ Now with the Research Department, RCA Laboratories Division, Princeton, N. J.

[1] R. E. Mathes and J. N. Whitaker, "Radio Facsimile by Subcarrier Frequency Modulation", *RCA REVIEW*, Vol. 4, pp. 131-154; October, 1939.

in signal-amplitude variation rather than to any increase in signal-to-noise ratio resulting from the frequency-modulation method.

Radiophoto Practice

All facsimile systems use a scanning device at the sending terminal for converting graphic material into a variable amplitude current or voltage. In the subcarrier-frequency-modulation process this variable quantity representing the point-by-point density of the copy is used to swing or shift a subcarrier tone over a definite predetermined range. This constant-amplitude, frequency-modulated tone is then sent out by means of a standard amplitude-modulation radiotelephone transmitter. At the distant terminal all amplitude variations resulting from fading and multipath effects are removed by means of a limiter, and a simple frequency-discriminator circuit converts the received signals back to the conventional varying amplitude wave. The copy is then recorded photographically, or by other means.

Since May, 1939, when subcarrier frequency modulation first went into public service between London and New York, additional terminals of this type have been put into commercial use. Berlin did and Buenos Aires does provide this service now and just recently transmitting facilities have been put into operation in Moscow. Figure 1 shows one of the first photographs sent by radio from Moscow over a 5000-mile circuit.

The two outstanding advantages of subcarrier frequency modulation over dot-pattern and other systems are better quality of received copy and increased speed. Much finer picture detail can be reproduced and contrast in the darker half tones is better. Eight- and ten-point type are now transmitted successfully whereas fourteen-point type formerly was the lower limit in size. Streaks and irregularities due to noise are not so noticeable in the radiophotos. Because of this better quality and freedom from a screen, the radio pictures can be handled the same as conventional photographs by publishers. An eight- by ten-inch picture can be transmitter in twenty minutes under radio-circuit conditions which three years ago would require a period of one hour. Improved operating technique and advancements in radio communication indicate that the speed may be increased still more.

The primary reason for this great improvement is due to the fact that subcarrier frequency modulation offers the equivalent of a linear amplitude-variation system which functions independently of signal-level fluctuations caused by fading and multipath phenomena which always occur on long-distance radio circuits. Since the picture intelligence is conveyed by frequency modulation of the subcarrier tone,

variations in amplitude can be completely removed by means of a limiting device. Since changes in amplitude as great as 70 decibels can be eliminated, signal fading can have very little influence on the reproduced copy. Only rarely does the signal drop below the usable level.

Although frequency modulation is capable of giving a definite gain in signal-to-noise ratio, this advantage increases with increasing deviation ratio. In the present picture-transmission system this advantage is not fully used because a comparatively low deviation is utilized; actually, the bandwidth is no greater than that for amplitude modulation. Some time ago a series of tests was made in one of the research

Fig. 1—One of the first test radiophotos sent from Moscow to New York, via RCA Communications, Inc.

laboratories of RCA Communications, Inc., to determine the bandwidth requirement for picture transmission by subcarrier frequency modulation. Various depths of modulation and bandwidths were tried and oscillograms and transmitted copy were compared and closely examined. The comparative importance of various sideband components was also investigated.

The results proved that if the index of modulation m, which is the ratio of frequency deviation of the carrier to the modulating frequency,

did not exceed 0.47 for the highest important frequency f_m from the picture material, then the total bandwidth needed is equal to 2 f_m. This range $(f_c - f_m)$ to $(f_c + f_m)$, where f_c is the carrier frequency, is the same as that for amplitude modulation and reproduces good copy. The highest important frequency f_m is determined by the speed of scanning and by the "definition" of the graphic material to be reproduced according to the I.R.E. standard method. With this bandwidth of $2f_m$ only the first-order sidebands of f_m are transmitted, the higher-order sidebands being sufficiently unimportant to be neglected.

Under this condition of bandwidth and $M = 0.47$ for the top important frequency, it can be shown that for any other lower signal frequency no sideband components having amplitudes greater than 10 per cent of that of the unmodulated carrier are lost although the frequency swing or deviation is held constant. For example, with $f_c = 1800$ cycles per second, $f_m = 600$ cycles per second, and $m = 0.47$, the first-order sidebands have amplitudes of 15 per cent of that of the unmodulated carrier and the second-order sidebands have amplitudes of 2 per cent. The former will lie just within the $2f_m$ band width while the latter will be lost. For a lower signal frequency of 300 cycles per second, m would be 0.94 for the same deviation and the first-order sidebands having amplitudes of 42 per cent lie well within the transmission band. The second-order sidebands in this case have amplitudes of 10 per cent but they are also inside the $2f_m$ band. Hence no components having amplitudes greater than 10 per cent are lost. Picture quality is just as good as that obtained with subcarrier amplitude modulation employing the same bandwidth.

Although a subcarrier frequency range 1600 to 2000 cycles has been used for most of the transoceanic picture traffic, other ranges would be equally satisfactory. It is good practice to have the mid- or carrier-frequency several times the highest signal frequency but good results can be obtained with a wide variety of values of frequency swing. One communications system uses the frequency deviation range 2600 to 3400 cycles. In a higher-speed, experimental, wire-line test the range 7.5 to 10.5 kilocycles gave satisfactory results. From the standpoint of successful radio transmission the phenomena of selective fading would indicate that a comparatively low value of the subcarrier would be desirable since this would keep the radio-frequency bandwidth relatively narrow. However, there was no noticeable difference between results with $f_c = 1800$ cycles per second and $f_c = 3000$ cycles per second.

Another advantage of subcarrier frequency modulation over the old constant-frequency, variable-dot-length method is a more simplified, less critical operating technique. When a picture is to be sent, a band

on the transmitting drum containing alternate black and white sections is first scanned. The scanner and modulator are then adjusted to produce 2000 cycles and 1600 cycles for black and white, respectively. At the radio transmitter the percentage amplitude modulation is set at the desired value and this remains constant since there is no amplitude variation in the subcarrier frequency-modulated picture wave. At the receiving terminal the recording galvanometer light valve or a similar device is adjusted to the proper settings for the incoming black-and-white controlled signals. Several pictures may then be transmitted without readjustment at either terminal.

The value and success of the present system of transmitting pictures over long distances by radio is indicated by the large number of such pictures which appear in our newspapers. The quality of the transmitted photos is frequently so good that they are indistinguishable from the regular news pictures.

A means for partially overcoming some of the effects of selective fading has been outlined.[1] This consisted in using the second harmonic of the subcarrier frequency which was developed in the receivers when the radio-frequency carrier completely dropped out. The harmonic wave which also contains the picture modulation was switched into use whenever the level of the normal frequencies fell below the usable point.

Telegraph Channeling with Subcarrier Frequency Modulation

For several years there has been in operation between New York and Philadelphia a two-way, multiple-channel, telegraph radio-relay circuit consisting of three links.[2] There are eight narrow-band, voice-frequency, teletype channels and also some wide-band, higher-frequency channels for facsimile. Each teletype channel is 100 cycles wide and the channels are spaced 170 cycles apart beginning with 425 cycles for the lowest frequency. The signal of each channel consists of its frequency keyed on and off according to the message material and the outputs of all eight channels are combined directly together through filters. The composite signal is then used to amplitude modulate the radio-frequency carrier, but in order to prevent overmodulation from peaks due to random combination the signal levels in the individual channels must be kept comparatively low.

Although this system has been successful in operation, difficulty is experienced from time to time because in this system of operation the teletype printers in narrow-band channels are sensitive to variations

[2] J. Ernest Smith, Fred H. Kroger, and R. W. George, "Practical Application of an Ultra-High-Frequency Radio-relay Circuit", *Proc. I.R.E.*, Vol. 26, pp. 1311-1327, November, 1938.

in incoming signal level. This requires a very careful adjustment of levels at both terminals and at the relay points. If for any reason there is a signal-level variation of a few decibels, then printer failures occur.

Because of the great success of subcarrier frequency modulation in picture transmission, it seemed reasonable to expect an improvement if it were applied to subcarrier telegraph operation. Tests were accordingly made and a marked improvement was noted. Figure 2 is a block diagram showing how a subcarrier frequency-modulation unit was used between the group of outgoing channels and the radio-frequency transmitter and a combined limiter and frequency discriminator were included between the receiver and incoming channeling filters. By direct comparison it was found that where a 4-decibel drop in equivalent

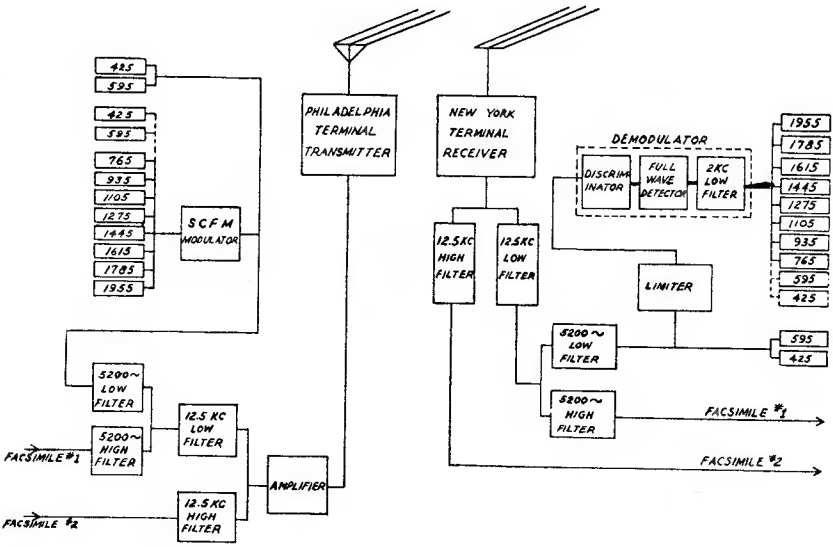


Fig. 2—Diagram of test circuit showing subcarrier frequency modulation applied to multichannel teletype operation.

signal level caused failure in the amplitude-modulation method, a 22-decibel drop was required to cause failure when subcarrier frequency modulation was used. This shows an appreciable advantage.

In both cases the signal-to-noise ratio was of secondary importance. The advantageous gain in allowable signal-level fluctuation before printer dropouts occur is due primarily to elimination of most of the effect of this fluctuation before it gets to the actual printer machine. As in the case of pictures, amplitude variations encountered in transmission are reduced to nil by the limiter device included at the receiving terminal.

Another suggested means for overcoming the variation in signal level in the present telegraph channeling system is the use of automatic volume control in each channel. However, the use of subcarrier frequency modulation is preferable because of its inherent nature of conveying intelligence irrespective of amplitude fluctuations.

Still another means of accomplishing the same desired result is to use subcarrier frequency modulation in each of the individual channels. This, in reality, would be a case of frequency-shift keying. Although it provides for satisfactory operation, frequency-shift keyers and limiters and discriminators are needed for each channel.

If either amplitude or frequency modulation of a subcarrier is used, the matter of a low ratio of subcarrier frequency to maximum signal frequency is of importance. If this subcarrier is to modulate a higher-frequency primary carrier it is desirable to use as low a value of subcarrier as possible so as to conserve bandwidth or to use a given bandwidth most efficiently. Results for the subcarrier amplitude-modulation

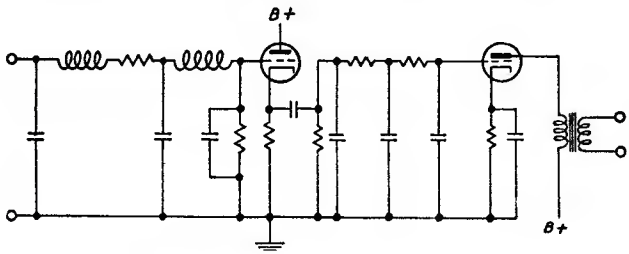


Fig. 3—Wide-range frequency discriminator.

case with a facsimile type of signal have been published.[3] With too low a ratio of carrier to signal, poor results were obtained but a value of 2.5 to 1 was considered "entirely adequate for commercial service." However, in the teletype channeling tests with subcarrier frequency modulation, satisfactory results were obtained with a maximum signal frequency of 2 kilocycles modulating a 3-kilocycle carrier, a ratio of 1.5 to 1. This was with a low index of modulation and is unusual because the lower first-order sideband has a frequency value less than the modulating frequency.

The reader who is familiar with various types of frequency discriminators will probably be interested in the unit used in the above mentioned tests. It was made up with two parts as illustrated in Figure 3. The first part consisted of two sections of constant-*K*, low-pass

[3] R. E. Mathes and J. E. Smith, "Propagation Requirements for Facsimile", RADIO FACSIMILE, RCA Institutes Technical Press, New York, N. Y., 1938.

filter, with coils having a comparatively low value of Q. The frequency response is given by curve A of Figure 4 and shows discrimination such as that used in the radiophoto terminal equipment. The range of linearity of this filter unit was not as great as desired for another application so two sections of a resistance-capacitance-type low-pass filter, having the characteristic of curve B, were added. Isolating vacuum tubes are used for coupling and impedance matching. Curve C shows the over-all frequency response of the complete unit which has an exceedingly wide range of linearity.

Such a characteristic is necessary when the ratio of subcarrier to signal frequency is low, but in some of the applications of subcarrier frequency modulation, this mode of operation is highly advantageous in the conservation of frequency bandwidth. An actual saving is made

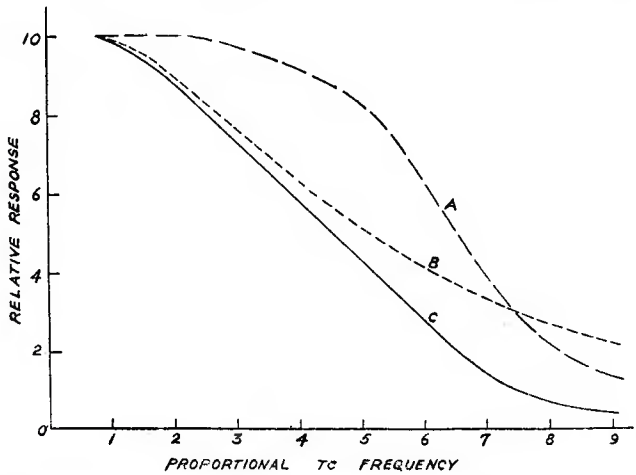


Fig. 4—Frequency-response curves of wide-range frequency discriminator.

because if a high value of subcarrier is used in the modulation of a higher primary carrier then there will be a "waste" space of frequencies between the primary carrier and its nearest sidebands.

Conclusions

In summing up the results and observations with subcarrier frequency modulation, the outstanding advantage of this system in point-to-point communications is in overcoming the ill effects of unwanted amplitude variations which are always present to some extent in all transmission circuits. In any system in which the intelligence is conveyed by amplitude variation, any noise wave introduced into the path

of transmission will add directly to the signal and produce distortion, masking, or other undesirable effects. In frequency modulation, all amplitude irregularities are removed by limiting or clipping. The inherent gain in signal-to-noise ratio obtained with this type of modulation is also utilized to an extent although the index of modulation is usually less than unity. In general the same advantages of frequency modulation are obtained in subcarrier frequency modulation as in the more common application of straight frequency modulation of a radio-frequency carrier.

THE TRANSMISSION OF A FREQUENCY-MODULATED WAVE THROUGH A NETWORK*†

By

Walter J. Frantz

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*A practical method for calculating the effect of a four-terminal network upon a frequency-modulated wave being transmitted through it is developed and demonstrated. The form of the solution is simple enough to be applied by anyone familiar with electric circuits. No knowledge of calculus or higher mathematics is required; nor is the solution restricted in any way, being equally accurate and practical for large and small values of modulation index, and for any physical network.*

In order to determine whether or not a particular network problem involving a frequency-modulated input wave may be analyzed from the "instantaneous frequency" viewpoint, a test, or validity condition, has been developed. This test quickly classifies the problem either as one for which only a complete, straightforward analysis determines the response, or as one for which a quasi-steady state exists. The quasi-steady state is a condition under which the amplitude—instantaneous-input-frequency envelope and the phase-shift—instantaneous-input-frequency envelope of the output wave approach closely enough to the steady-state-amplification—frequency and phase-shift—frequency characteristics of the network.

Introduction

WITH THE increasing use of a frequency-modulated wave for circuit alignment and test, spectrum analysis, altitude or time-delay measurement, and the transmission of intelligence, there arises the need for a practical and dependable method of calculating the response of a network to a frequency-modulated input wave.

Suppose that it is desired to picture accurately the steady-state amplification—frequency characteristic of a tuned one-megacycle amplifier with an effective Q of 100. The layout of Figure 1 is often used for visual circuit test and alignment in the communications industry. Since the modulating voltage is applied also to the horizontal plates of the oscilloscope, the horizontal axis is linear in instantaneous frequency of the input frequency-modulated wave. In Figure 2, (a) is the oscillogram obtained when the frequency-modulated signal generator is oper-

* Decimal classification: R630.
† Reprinted from *Proc. I.R.E.*, March, 1946.

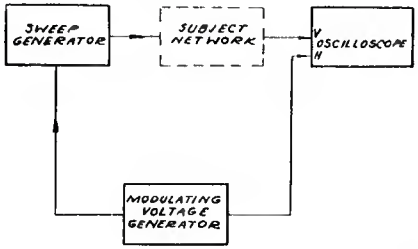


Fig. 1—Block layout of a visual network test and alignment position.

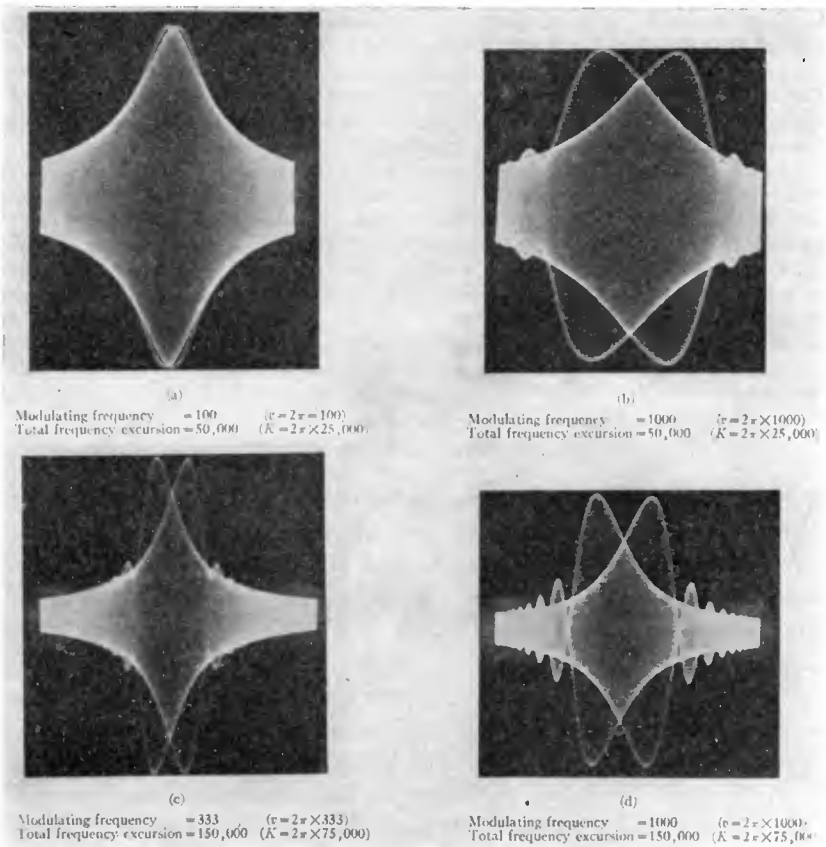


Fig. 2—Experimental amplitude—instantaneous-input-frequency and phase-shift—instantaneous-input-frequency response of a tuned amplifier to a frequency-modulated wave.

Effective Q of amplifier $= 100$ $(p = 0.01)$
Center frequency $= 10^6$ $(\omega_0 = 2\pi \times 10^6)$

ating at a center frequency of one megacycle, a total sweep width of 50 kilocycles, and a modulating frequency of 100 cycles per second; the amplification—frequency characteristic of the amplifier has been fairly well pictured. If the modulating frequency is increased to 1000 cycles per second, the pattern (b) departs considerably from the steady-state amplification—frequency characteristic of the network. The two traces are the frequency upsweep and the downsweep. The upsweep trace leans to the right and the downsweep trace leans to the left. In Figure 2, (c) is the oscillogram obtained for a total sweep width of 150 kilocycles and a modulating frequency of 333 cycles per second, while (d) is the oscillogram obtained for a total sweep width of 150 kilocycles and a modulating frequency of 1000 cycles per second.

These oscillograms will be referred to quantitatively later in the article. They are presented here only to persuade those unfamiliar with the problem that it is real, important, and frequently encountered. The problem is probably given the most attention by the designers of high-quality frequency-modulated transmitters and receivers.

It is the purpose of this paper to devise a simple test by which one may determine whether or not the effects of a network upon a frequency-modulated wave being transmitted through it differ appreciably from the steady-state amplification—frequency and phase-shift—frequency characteristics of the network, and, if so, to offer a general and practical method of calculating these effects as a function of either time or instantaneous input frequency.

The Response-Envelope Equation

The following procedure is a rigorous and general method of obtaining the output-voltage envelope for a network driven by a frequency-modulated wave:

1. Measure or calculate the steady-state amplification—frequency and phase-shift—frequency characteristics of the network in question.
2. Express the frequency-modulated input wave in terms of its steady-state spectrum of sinusoidal side frequencies.
3. Pass the individual side frequencies through the network, altering the amplitude and phase of each according to the steady-state amplification and phase characteristics of the network at the particular frequency of the side frequency.
4. Plot the sum of the altered side frequencies point by point to obtain the output wave as a function of time.
5. Draw a smooth curve through the carrier-voltage peaks to obtain the response envelope of the output voltage.

Such a procedure, however, is not a solution to the frequency-

modulated transmission problem. The method is rigorous and general, but not practical. Calculating the response pattern of (a) in Figure 2 by the method just outlined, for example, would require the labor of a staff of clerks for several years, because it would be necessary to calculate about 100,000 points for a smooth curve of the radio-frequency voltage oscillations during one modulation cycle, each point being the sum of the instantaneous values of more than 500 side-frequency voltages.

It is fortunate that the operations outlined by steps 2, 3, 4, and 5 can be done analytically, initiating a practical equation for the output or response-voltage envelope in terms of the network parameters and the frequency-modulated input-wave parameters. This equation will be derived with time as the independent variable and transformed to a function of instantaneous input frequency to correspond with the abscissae of Figure 2. The method used in this report for bridging steps 2, 3, 4, and 5 is very similar to the pattern followed by Cherry and Rivlin.[1] The problem has also been considered by Kulp, who developed a solution which is practical when the modulation index of the frequency-modulated input wave is small.[2]

First a unit frequency-modulated wave is defined. The equations derived in this paper are based upon unit magnitude of the input signal.

$$\omega_0 = 2\pi \times \text{the center frequency} \tag{1}$$

$$K = 2\pi \times \text{the frequency deviation} \tag{2}$$

$$v = 2\pi \times \text{the modulating frequency} \tag{3}$$

$$m_f = \frac{K}{v} = \text{the modulation index} \tag{4}$$

$$\hat{\omega} = \omega_0 + K \cos vt = 2\pi \times \text{the instantaneous input frequency.} \tag{5}$$

It has been shown that[3]

$$e_{in} = \sin\ (\omega_0 t + m_f \sin vt)$$

[1] E. C. Cherry and R. S. Rivlin, "Nonlinear Distortion, with Particular Reference to the Theory of Frequency-modulated Waves," Part II, *Phil. Mag.*, vol. 33, pp. 272-293; April, 1942.

[2] M. Kulp, "Spektra und Klirrfaktoren Frequenz- und Amplitudenmodulierter Schwingungen," Part II, *Elek. Nach. Tech.*, vol. 19, pp. 96-109; June, 1942.

[3] John R. Carson, "Notes on the Theory of Modulation," *Proc. I.R.E.*, vol. 10, pp. 57-64; February, 1922.

$$= J_0 (m_f) \sin \omega_0 t$$

$$+ \sum_{n=1}^{\infty} (J_{2n-1} (m_f) \{\sin [\omega_0 + (2n-1) v] t$$

$$- \sin [\omega_0 - (2n-1) v] t\}$$

$$+ J_{2n} (m_f) \{\sin [\omega_0 + 2nv] t + \sin [\omega_0 - 2nv] t\}). \quad (6)$$

The series (6) is absolutely convergent, allowing the convenience of a partial summation with any desired accuracy.

$$e_{in} \approx J_0 (m_f) \sin \omega_0 t$$

$$+ \sum_{n=1}^{N} (J_{2n-1} (m_f) \{\sin [\omega_0 + (2n-1) v] t$$

$$- \sin [\omega_0 - (2n-1) v] t\}$$

$$+ J_{2n} (m_f) \{\sin [\omega_0 + 2nv] t + \sin [\omega_0 - 2nv] t\}) \quad (7)$$

The choice of N depends upon m_f and the accuracy required of the analysis. The choice of N is not ordinarily difficult, because the Bessel coefficients of the series become negligible rather abruptly for a given argument as the order is increased beyond a certain value. For example, J_n (10,000)'s are relatively important up to $n = 10{,}000$ and slightly above, but are insignificant beyond $n = 10{,}050$.

For pure sinusoidal input of angular velocity ω to a linear network, the output wave can differ from the input wave only in amplitude and phase. By the rules of algebraic steady-state circuit analysis E_{out} can be expressed in general as the product of E_{in} and a complex function of ω. (See Appendix.)

$$E_{out} \{=\} E_{in} f(\omega) \quad (8)$$

where

$$| f(\omega) | = \frac{\text{amplitude of } E_{out}}{\text{amplitude of } E_{in}} \quad (9)$$

and where

$$\text{arc tangent} \frac{\text{imaginary } [f(\omega)]}{\text{real } [f(\omega)]} = \text{angular phase } [E_{out}] - \text{angular phase } [E_{in}]. \quad (10)$$

The $f(\omega)$ is a dimensionless transfer-voltage ratio. If I_{out} were used instead of E_{out}, $f(\omega)$ would be the transfer admittance.

The next step is to approximate $f(\omega)$ by a finite Fourier series about ω_0 in a period Ω equal to or greater than the interval containing the side-frequency voltage of (7). From (7) it is obvious that this interval extends from $(\omega_0 - 2Nv)$ to $(\omega_0 + 2Nv)$, a total width of $4Nv$.

$$f(\omega) \approx \sum_{m=0}^{M} \left(A_m \cos \left[\frac{2\pi m}{\Omega} (\omega - \omega_0) \right] + B_m \sin \left[\frac{2\pi m}{\Omega} (\omega - \omega_0) \right] \right). \tag{11}$$

The notation $\approx$ used in (11) means "approximately equals only in the interval $(\omega_0 - 2Nv) \leqq \omega \leqq (\omega_0 + 2Nv)$." It is sufficient that the partial sum represent the function only in the interval $(\omega_0 2 - 2Nv) \leqq \omega \leqq (\omega_0 + 2Nv)$ since N has previously been chosen large enough that only relatively insignificant side frequencies of the frequency-modulated input voltage lie outside that interval. The choice of M depends upon the accuracy with which it is desired to approximate $f(\omega)$ and the nature of $f(\omega)$.

The reason for expressing $f(\omega)$ as a series approximation is to make the method general. If an analytical solution to steps 2, 3, 4, and 5 can be obtained for the terms of the series approximation, the problem is then solved for any $f(\omega)$, since any physical network characteristic can be approximated by such a finite series as (11). The Fourier series was chosen in preference to other series approximations because the analytical solution to steps 2, 3, 4, and 5 is not difficult for sinusoidal network terms, because engineering personnel are most familiar with the use of the Fourier series, and because the Fourier series, being periodic, is well behaved and known everywhere outside the interval of approximation. The power and real exponential series were avoided because such series increase without bound outside the chosen interval as the argument becomes large. Although N has been chosen large enough so that the side frequencies outside the interval $(\omega_0 - 2Nv) \leqq \omega \leqq (\omega_0 + 2Nv)$ are small enough to be neglected in comparison with the amplitude of the side frequencies within the interval, the products of the side frequencies and the value of the series outside the interval of approximation might not be negligible in comparison with the products of the side frequencies and the series approximation within the interval.

The A_m's and B_m's of (11) are the complex Fourier coefficients defined by the following integrals.

$$A_0 = \frac{1}{\Omega} \int_{\omega_0 - \Omega/2}^{\omega_0 + \Omega/2} f(\omega)\, d\omega \tag{12}$$

$$A_{m>0} = \frac{2}{\Omega} \int_{\omega_0 - \Omega/2}^{\omega_0 + \Omega/2} f(\omega) \cos\left[\frac{2\pi m}{\Omega}(\omega - \omega_0)\right] d\omega \tag{13}$$

$$B_m = \frac{2}{\Omega} \int_{\omega_0 - \Omega/2}^{\omega_0 + \Omega/2} f(\omega) \sin\left[\frac{2\pi m}{\Omega}(\omega - \omega_0)\right] d\omega. \tag{14}$$

These integrals can seldom be evaluated analytically, because either the $f(\omega)$ is known only numerically, or, if known analytically, the integrals are too difficult to handle. Numerical methods of obtaining the values of the integrals are nearly always preferable. These methods have been organized well enough so that, for example, a 24-point numerical analysis of a function takes only a few hours.[4,5]

Bridging steps 3, 4, and 5 analytically yields the following general equation for the output frequency-modulated voltage envelope. (See Appendix.)

$$\boxed{E_{out}\ \{=\}\ \sum_{m=0}^{M} D_m (A_m \cos\psi_m + B_m \sin\psi_m)} \tag{15}$$

where D_m and ψ_m are calculated as functions of either time t or instantaneous angular input velocity $\hat{\omega}$ as follows (see Appendix):

$$D_m(t) = \cos\left[\frac{K}{v}\sin vt\left(1 - \cos\frac{2\pi m v}{\Omega}\right)\right] - j\sin\left[\frac{K}{v}\sin vt\left(1 - \cos\frac{2\pi m v}{\Omega}\right)\right] \tag{16}$$

$$\psi_m(t) = \left[\frac{K}{v}\cos vt \sin\frac{2\pi m v}{\Omega}\right] \tag{17}$$

[4] J. B. Scarborough, "NUMERICAL MATHEMATICAL ANALYSIS," Johns Hopkins University Press, Baltimore, Md., 1930, chap. 17, pp. 388-404.

[5] R. P. G. Denman, "Thirty-six and Seventy-two Ordinate Schedules for General Harmonic Analysis," *Electronics*, vol. 15, pp. 44-47; September 1942. Corrections by F. W. Grover, *Electronics*, vol. 16, pp. 214-215; April, 1943.

$$D_m(\hat{\omega}) = \cos\left[\frac{\sqrt{K^2-(\hat{\omega}-\omega_0)^2}}{v}\left(1-\cos\frac{2\pi mv}{\Omega}\right)\right] \pm j\sin\left[\frac{\sqrt{K^2-(\hat{\omega}-\omega_0)^2}}{v}\left(1-\cos\frac{2\pi mv}{\Omega}\right)\right] \quad (18)$$

$$\psi_m(\hat{\omega}) = \frac{\hat{\omega}-\omega_0}{v}\sin\frac{2\pi mv}{\Omega}. \quad (19)$$

The amplitude and phase shift of E_{out} can be plotted as a function of time by calculating the magnitude and phase of (15) for several equally spaced values of t between $vt = 0$ and $vt = 2\pi$. As a function of instantaneous input frequency the amplitude and phase shift of E_{out} can be determined by calculating the magnitude and phase of (15) for several equally spaced values of $\hat{\omega}$ between $(\hat{\omega} - \omega_0) = -K$ and $(\hat{\omega} - \omega_0) = K$. As a function of instantaneous input frequency E_{out} has two values, one corresponding to the frequency downsweep and one corresponding to the frequency upsweep. The sign of the imaginary part of $D_m(\hat{\omega})$ is negative for the downsweep of frequency and positive for the upsweep. (See Appendix.)

It should be noted that E_{out} (15) is unusual in that it describes the amplitude and phase of a frequency-modulated oscillation. The arc tangent of the quotient of the imaginary and real parts of E_{out} is equal to the deviation between the instantaneous angular phase of the output wave and the instantaneous angular phase of the frequency-modulated input wave. When using (15) to compute points for a response envelope it must be remembered that D_m, A_m, and B_m are complex quantities, and that the summation is a complex summation. The magnitude of E_{out} is the absolute value of the complex summation rather than the sum of the absolute values of $D_m(A_m \cos\psi_m + B_m \sin\psi_m)$.

The Validity Condition Upon the "Instantaneous-Frequency" Method of Analysis

It is not surprising that the response of the networks of a final working frequency-modulation design can usually be justified from the "instantaneous-frequency" premise. A quasi-steady state usually results from careful design for a practical application; the quasi-steady state is often necessary. When the "instantaneous-frequency" method of analyzing the effect of a network upon a frequency-modulated wave is valid (that is, when a quasi-steady state exists) the response envelope approaches the steady-state amplification—frequency and phase-shift

—frequency characteristics of the network. At any instant of time the frequency-modulated output wave deviates from the frequency-modulated input wave in amplitude and phase in a manner no different from that of a steady-state output wave whose constant frequency is equal to the instantaneous frequency of the frequency-modulated wave. Letting the quasi-steady-state response envelope $E_{\text{out}}{}^{1}(\hat{\omega})$ approach $f(\omega)$, equation (11),

$$E_{\text{out}}{}^{1}(\hat{\omega}) \rightarrow f(\omega) \approx \sum_{m=0}^{M} (A_m \cos \psi_m{}^{1} + B_m \sin \psi_m{}^{1}) \tag{20}$$

where

$$\psi_m{}^{1}(\hat{\omega}) = \frac{2\pi m}{\Omega} (\hat{\omega} - \omega_0). \tag{21}$$

Equation (20) is equivalent to (15) if $D_m \rightarrow 1$ and $\psi_m \rightarrow \psi_m{}^{1}$. The discrepancy between the angles ψ_m and $\psi_m{}^{1}$ is given by

$$\psi_m{}^{1} - \psi_m = \frac{2\pi m}{\Omega} (\hat{\omega} - \omega_0) - \frac{\hat{\omega} - \omega_0}{v} \sin \frac{2\pi m v}{\Omega}. \tag{22}$$

For small discrepancies this becomes

$$\psi_m{}^{1} - \psi_m \approx \frac{\hat{\omega} - \omega_0}{6v} \left(\frac{2\pi m v}{\Omega} \right)^3. \tag{23}$$

The maximum discrepancy between $\psi_m{}^{1}$ and ψ_m occurs at the ends of the sweep where $\hat{\omega} - \omega_0 = K$, the maximum frequency deviation.

$$\psi^{1} - \psi_m \approx \frac{K}{6v} \left(\frac{2\pi m v}{\Omega} \right)^3 \text{maximum.} \tag{24}$$

The discrepancy between D_m and 1 is given by

$$D_m - 1 = \cos \left[\frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(1 - \cos \frac{2\pi m v}{\Omega} \right) \right]$$
$$\pm j \sin \left[\frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(1 - \cos \frac{2\pi m v}{\Omega} \right) \right] - 1. \tag{25}$$

For small discrepancies this becomes

$$D_m - 1 \approx \pm \frac{j}{2} \frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(\frac{2\pi m v}{\Omega} \right)^2. \tag{26}$$

This discrepancy is maximum at the center of the frequency sweep $\hat{\omega} - \omega_0 = 0$, where the time rate of change of frequency is maximum. Neglecting the factor $\pm j$, this maximum discrepancy becomes

$$D_m - 1 \approx \frac{Kv}{2} \left(\frac{2\pi m}{\Omega} \right)^2 \text{maximum} \tag{27}$$

The discrepancy given by (24) is nearly always negligible if the discrepancy given by (27) is negligible. Therefore, if the discrepancy given by (27) for a particular problem is small enough to be considered negligible, the validity condition upon the "instantaneous-frequency" method of analysis has been fulfilled. One can establish quantitative ideas about how small the value of (27) must be to be considered negligible by studying the examples in the next section. Most engineers have already established their own personal landmarks concerning the allowable distortion of the terms of a Fourier approximation to a nonsinusoidal function. Fortunately, (27) gives the phase discrepancy of the terms of the Fourier approximation as proportional to m^2, so only the shortest period terms ($m = M$) need usually be considered. The terms for m less than M are distorted much less than the Mth.

A practical procedure which has proven quite satisfactory for using (27) follows. First, decide how far the magnitude of the response envelope can differ from the network characteristic in its most hard-to-follow region before a quasi-steady state ceases to exist. This magnitude may vary from one tenth of one per cent of the peak value of the network characteristic to 50 per cent of the peak, depending entirely upon the requirements and specifications of the equipment being designed. M is then determined such that the sum of the absolute values of all the coefficients of the Fourier approximation beyond the Mth terms is equal to or less than the allowable difference decided upon. It should be pointed out that the sum of the absolute values of the coefficients of the Fourier approximation always converges (and always converges quite rapidly beyond a certain coefficient) since there can be no discontinuities in the function or its derivatives for a physical network characteristic. How to avoid

creating extrinsic discontinuities at the ends of the period of the approximation will be discussed in the example in the next section.

Thirdly, the phase displacement of this Mth term of the approximation is investigated by (27). A quasi-steady state is assumed to exist if the discrepancy is less than one third radian or about 20 degrees. This sort of procedure is quite similar to common video-frequency-amplifier design practice. Amplifiers for non-sinusoidal waves are commonly designed to be down less than three decibels, that is, to have less than 45 degrees phase shift, at the frequency of the last significant harmonic of the signal to be amplified. That design criterion is somewhat less conservative than the allowable 20 degrees phase shift assumed here for the last significant harmonic of the network characteristic.

The procedure outlined above is, of course, not foolproof. It is certainly feasible to imagine a Fourier approximation of functions for which the above procedure would not insure the desired fidelity of the output envelope. Such functions, however, are seldom, if ever, encountered in physical networks.

Step-by-Step Procedure for Testing the Validity of the "Instantaneous-Frequency" Method of Analysis

The steps necessary for determining whether or not a quasi-steady state exists are organized and demonstrated in this section. The test is applied to the problems which, when tried experimentally, yielded (a) and (b) in Figure 2.

A. Investigate the spectrum of the frequency-modulated wave applied to the network, deciding how wide a frequency band need be considered in order to include all the non-negligible side frequencies of the wave.

Example: The frequency-modulated wave applied to a tuned amplifier to produce (b) in Figure 2 experimentally is described by the following parameters, which are repeated here for convenience:

$$\omega_0 = 2\pi \times 10^6 \qquad (2\pi \times \text{the center frequency})$$
$$K = 2\pi \times 25{,}000 \qquad (2\pi \times \text{the frequency deviation})$$
$$v = 2\pi \times 1000 \qquad (2\pi \times \text{the modulating frequency})$$
$$m_f = \frac{K}{v} = 25 \qquad (\text{the modulation index}).$$

The steady-state spectrum of the frequency-modulated wave is given

by (6). The Bessel amplitude coefficients $J_n(m_f)$ of the side frequencies are as follows:[6]

$J_0(25) = 0.0963$	$J_{21}(25) = 0.1646$
$J_1(25) = -0.1254$	$J_{22}(25) = 0.2246$
$J_2(25) = -0.1063$	$J_{23}(25) = 0.2306$
$J_3(25) = 0.1083$	$J_{24}(25) = 0.1998$
$J_4(25) = 0.1323$	$J_{25}(25) = 0.1529$
$J_5(25) = -0.0660$	$J_{26}(25) = 0.1061$
$J_6(25) = -0.1587$	$J_{27}(25) = 0.06778$
$J_7(25) = -0.0102$	$J_{28}(25) = 0.04028$
$J_8(25) = 0.1530$	$J_{29}(25) = 0.02245$
$J_9(25) = 0.1081$	$J_{30}(25) = 0.01181$
$J_{10}(25) = -0.0752$	$J_{31}(25) = 0.005889$
$J_{11}(25) = -0.1682$	$J_{32}(25) = 0.002795$
$J_{12}(25) = -0.0729$	$J_{33}(25) = 0.001267$
$J_{13}(25) = 0.0983$	$J_{34}(25) = 0.000550$
$J_{14}(25) = 0.1751$	$J_{35}(25) = 0.000229$
$J_{15}(25) = 0.0978$	$J_{36}(25) = 0.000092$
$J_{16}(25) = -0.0577$	$J_{37}(25) = 0.000036$
$J_{17}(25) = -0.1717$	$J_{38}(25) = 0.000013$
$J_{18}(25) = -0.1758$	$J_{39}(25) = 0.000005$
$J_{19}(25) = -0.0814$	$J_{40}(25) = 0.000002$
$J_{20}(25) = 0.0520$	$J_{41}(25) = 0.000001$

The amplitudes of the side frequencies are certainly negligible beyond the 35th side frequency for any practical application. Therefore, the interval $2\pi \times 0.965 \times 10^6 \leqq \omega \leqq 2\pi \times 1.035 \times 10^6$ contains all the non-negligible side frequencies.

B. Measure or calculate the steady-state complex-network characteristic $f(\omega)$.

Example: The amplification—frequency and phase-shift—frequency characteristics of a network can be measured. The real part of $f(\omega)$ is then obtained by multiplying the amplification by the cosine of the phase-shift angle. The imaginary part of $f(\omega)$ is obtained by multiplying the amplification by the sine of the phase-shift angle.

For several networks, however, it is practical to calculate the network characteristic from its circuit. Figure 3 is the simplified schematic of a tuned amplifier, for which

$$X_L = -X_C = jX_0 \text{ at } \omega_0 \tag{28}$$

[6] Eugene Jahnke and Fritz Emde, "TABLE OF FUNCTIONS WITH FORMULAE AND CURVES," Dover Publications, New York, N. Y., 1943, p. 179.

$$p = \frac{1}{Q_L} + \frac{X_0}{R_p} = \text{power factor of plate circuit} \tag{29}$$

$$g_m = \frac{-I_p}{E_{in}} \tag{30}$$

$$Z = \frac{1}{\dfrac{p}{X_0} + \dfrac{1}{j\dfrac{\omega}{\omega_0} - X_0} - \dfrac{1}{j\dfrac{\omega_0}{\omega} - X_0}}$$

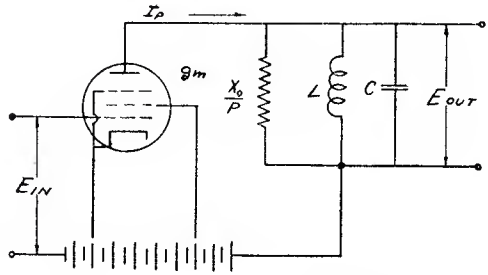


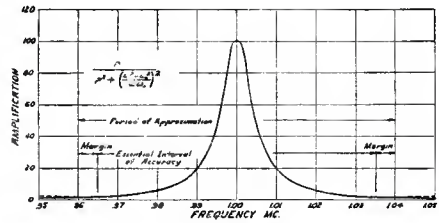
Fig. 3—Simplified schematic of a tuned amplifier.

$$= X_0 \left[\frac{p}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0} \right)^2} - j \frac{\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0}}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0} \right)^2} \right] \tag{31}$$

$$E_{out} = E_{in} g_m Z = E_{in} f(\omega) \tag{32}$$

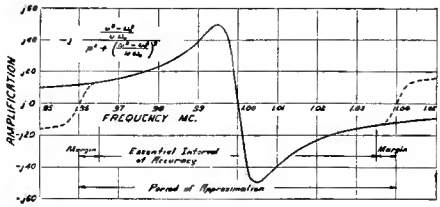
$$\frac{f(\omega)}{g_m X_0} = \frac{p}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0} \right)^2} - j \frac{\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0}}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0} \right)^2}. \tag{33}$$

In Figure 4, (a) is a plot of the real part of $f(\omega)/g_m X_0$ and (b) is the imaginary part. The square root of the sum of the squares of the real



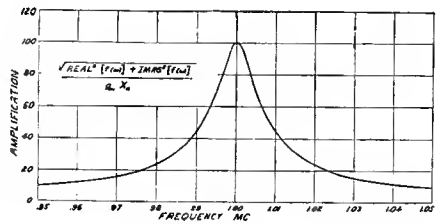
(a)

(a)—Real part of $\frac{f(\omega)}{g_m X_0}$.



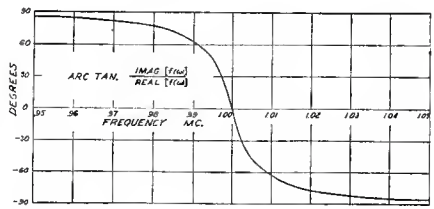
(b)

(b)—Imaginary part of $\frac{f(\omega)}{g_m X_0}$



(c)

(c)—Magnitude of $\frac{f(\omega)}{g_m X_0}$



(d)

(d)—Phase of $f(\omega)$.

Fig. 4—Complex-network characteristic of a tuned amplifier.

$$\frac{f(\omega)}{g_m X_0} = \frac{p}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0}\right)^2} - j\,\frac{\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0}}{p^2 + \left(\dfrac{\omega^2 - \omega_0^2}{\omega\omega_0}\right)^2}.$$

and imaginary parts is the amplification—frequency characteristic (c). The arc tangent of the quotient of the imaginary and real parts of $f(\omega)/g_m X_0$ is the phase-shift—frequency characteristic (d).

C. Approximate the network characteristic by a finite Fourier series over an interval equal to or greater than the essential interval determined in step A.

Example: Unless the network-characteristic function and its derivatives at one end point of the essential interval chosen in step A happen to be equal to the function and its derivatives, respectively, at the opposite end point of the interval, it is necessary to take the following steps to avoid extrinsic discontinuities which would needlessly burden and reduce the accuracy of the Fourier approximation. First, the period for the Fourier approximation should be chosen slightly larger than the interval over which the approximation must accurately represent the network characteristic. Second, the function should be altered in the two margins where the chosen period overlaps the essential interval of approximation in such a way that the new function has no discontinuities either within the chosen period or at its end points. The artificial function constructed within the margins should be as smooth as possible to avoid its contributing to the higher-order terms of the Fourier approximation.

Recall that the essential interval of approximation was $2\pi \times 0.965 \times 10^6 \leqq \omega \leqq 2\pi \times 1.035 \times 10^6$; let Ω, the period of approximation, be slightly greater, say, $2\pi \times 0.960 \times 10^6 \leqq \omega \leqq 2\pi \times 1.040 \times 10^6$. Then construct an artificial function in the margins of overlap of the intervals ($2\pi \times 0.960 \times 10^6 \leqq \omega \leqq 2\pi \times 0.965 \times 10^6$ and $2\pi \times 1.035 \times 10^6 \leqq \omega \leqq 2\pi \times 1.040 \times 10^6$) that smoothly joins the ends of the period. This artificial function has been dotted into (b) of Figure 4. In Figure 4, (a) is so nearly continuous at the end points that no alteration is necessary. The essential interval of approximation and the period of approximation are also indicated.

A 24-point numerical analysis[4] of this improved function in the indicated period yields its Fourier coefficients. When using numerical methods to obtain these coefficients, one must be careful that the signs of the coefficients are correct to place the approximation above ω_0 as required by (11).

$$
\begin{aligned}
A_0 &= 17.9 + j0 \\
A_1 &= 26.8 + j0 \\
A_2 &= 17.8 + j0 \\
A_3 &= 12.2 + j0 \\
A_4 &= 8.2 + j0 \\
A_5 &= 5.6 + j0
\end{aligned}
\qquad
\begin{aligned}
A_6 &= 3.8 + j0 \\
A_7 &= 2.6 + j0 \\
A_8 &= 1.8 + j0 \\
A_9 &= 1.3 + j0 \\
A_{10} &= 0.9 + j0 \\
A_{11} &= 0.8 + j0 \\
A_{12} &= 4.0 + j0
\end{aligned}
$$

$B_1 = 0 - j33.1$
$B_2 = 0 - j14.3$
$B_3 = 0 - j14.5$
$B_4 = 0 - j6.4$
$B_5 = 0 - j6.6$
$B_6 = 0 - j2.6$
$B_7 = 0 - j3.1$
$B_8 = 0 - j1.0$
$B_9 = 0 - j1.5$
$B_{10} = 0 - j0.3$
$B_{11} = 0 - j0.4$

Because the real part of this network characteristic is nearly symmetrical in the period chosen, the B_m's have a negligible real part. Likewise, the imaginary part of this network characteristic is very nearly skew symmetrical in the period chosen; therefore, the A_m's have a negligible imaginary part.

D. Find the maximum phase shift of the Mth component of the network-characteristic approximation by (27). If the maximum phase shift of the Mth component is less than 20 degrees, a quasi-steady state exists.

Example: Suppose that, for this particular problem, the requirements of a quasi-steady state shall be that the response envelope shall nowhere deviate from the network characteristic more than fifteen per cent of its peak value. Following the suggested procedure, it will be found that the sum of the absolute values of the coefficients of the A_m's and B_m's beyond the sixth is about 14 per cent of the peak magnitude of the function. Therefore, the terms corresponding to $m = M = 6$ are the last significant terms of the series.

The problem solved experimentally in (b) of Figure 2 will now be tested analytically for a quasi-steady state. The parameters K and v have been given, Ω was chosen in step C, and the largest significant value of m ($M = 6$) has been chosen in the preceding paragraph.

$$D_m - 1 \approx \frac{Kv}{2}\left(\frac{2\pi m}{\Omega}\right)^2 \approx \frac{2\pi \times 25{,}000 \times 2\pi \times 1000}{2}\left(\frac{2\pi \times 6}{2\pi \times 80{,}000}\right)^2 \tag{27}$$

$$\approx 2.8 \text{ radians or } 160 \text{ degrees.}$$

This large phase discrepancy for $m = 6$ indicates that a quasi-steady state does not exist, and that the distortion is quite large. This conclusion is confirmed by (b) in Figure 2.

The series obtained in step C is also adequate for the problem solved experimentally in (a) of Figure 2, because the spectrum is even more closely confined. Substituting the parameters for (a) of Figure 2 into (27),

$$D_m - 1 \approx \frac{2\pi \times 25{,}000 \times 2\pi \times 100}{2} \left(\frac{2\pi \times 6}{2\pi \times 80{,}000} \right)^2$$

$$\approx 0.28 \text{ radian or } 16 \text{ degrees.}$$

Since the phase discrepancy for this Mth term is less than 20 degrees, a quasi-steady state (such as postulated) may be assumed to exist. In Figure 2, (a) bears out this conclusion nicely; the discrepancy between (a) and the static-network response characteristic is obviously everywhere less than 15 per cent of the peak.

Step-by-Step Procedure for Calculating the Response of a Network to a Frequency-Modulated Wave

The necessary steps for evaluating (15) as a function of either t or $\hat{\omega}$ are organized and demonstrated in this section. The amplitude—instantaneous-input-frequency response and phase-shift—instantaneous-input-frequency response of a tuned amplifier will be calculated using the same parameters that were used experimentally to obtain (b) of Figure 2.

I. Perform steps A, B, and C of the validity-test procedure.

II ($\hat{\omega}$). Prepare work sheets similar to Table I for several equally spaced values of $(\hat{\omega} - \omega_0)$ between zero and K.

Example: There should be work sheets for enough points to determine smooth curves of the amplitude—instantaneous-input-frequency and phase-shift—instantaneous-input-frequency response. Work sheets for $(\hat{\omega} - \omega_0) = 2\pi \times 2000,\ 2\pi \times 4000,\ 2\pi \times 6000, \cdots,\ 2\pi \times 24{,}000$ were prepared for the example; Table I is the work sheet for $(\hat{\omega} - \omega_0) = 2\pi \times 6000$.

II (t). Prepare work sheets similar to Table I for several equally spaced values of vt between zero and $\pi/2$ (90 degrees).

III. Into column III enter the orders of the Fourier coefficients obtained in step C.

This column is the same for all the work sheets.

IV ($\hat{\omega}$). Into column IV enter the value of

$$\frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(1 - \cos \frac{2\pi m v}{\Omega} \right) \times 57.3 \text{ degrees.}$$

Example: The parameters K and v are given. The period Ω was chosen in step C. A value of $(\hat{\omega} - \omega_0)$ is assigned to each calculation

sheet; the value assigned to Table I is $(\hat{\omega} - \omega_0) = 2\pi \times 6000$. Thus, for $m = 2$,

$$\text{IV} = \frac{\sqrt{(2\pi \times 25{,}000)^2 - (2\pi \times 6000)^2}}{2\pi \times 1000}$$

$$\left(1 \cos \frac{2\pi \times 2 \times 2\pi \times 1000}{2\pi \times 80{,}000}\right) \times 57.3$$

$$= 17.1 \text{ degrees}$$

TABLE I

SAMPLE WORK SHEET FOR CALCULATING E_{out}

III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
m	Phase of D_m (degrees)	ψ_m (degrees)	A_m	$A_m \cos$ (V)	(VII) cos (IV)	j(VII) sin (IV)	B_m	$B_m \sin$ (V)	(XI) cos (IV)	j(XI) sin (IV)
0	0	0	17.9+j0	17.9+j0	17.9+j0					
1	4.3	27.0	26.8+j0	23.9+j0	23.9+j0	0+j1.8	0−j33.1	0−j14.9	0−j14.9	1.1+j0
2	17.1	53.8	17.8+j0	10.5+j0	10.0+j0	0+j3.1	0−j14.3	0−j11.5	0−j11.1	3.4+j0
3	38.4	80.3	12.2+j0	2.1+j0	1.6+j0	0+j1.3	0−j14.5	0−j14.3	0−j11.2	8.9+j0
4	68	106	8.2+j0	−2.3+j0	−0.9+j0	0−j2.1	0− j6.4	0− j6.1	0− j2.3	5.7+j0
5	106	132	5.6+j0	−3.7+j0	1.0+j0	0−j3.6	0− j6.6	0− j4.9	0+ j1.3	4.7+j0
6	152	156	3.8+j0	−3.5+j0	3.1+j0	0−j1.6	0− j2.6	0− j1.1	0+ j1.0	0.5+j0
7	205	180	2.6+j0	−2.6+j0	2.4+j0	0+j1.1	0− j3.1	0+ j0	0+ j0	0 +j0
8	266	202	1.8+j0	−1.7+j0	0.1+j0	0+j1.7	0− j1.0	0+ j0.4	0+ j0	0.4+j0
9	333	223	1.3+j0	−1.0+j0	−0.9+j0	0+j0.5	0− j1.5	0+ j1.0	0+ j0.9	0.5+j0
10	407	243	0.9+j0	−0.4+j0	−0.3+j0	0−j0.3	0− j0.3	0+ j0.3	0+ j0.2	−0.2+j0
11	487	262	0.8+j0	−0.1+j0	0.1+j0	0−j0.1	0− j0.4	0+ j0.4	0− j0.3	−0.3+j0
12	572	278	0.4+j0	0.1+j0	−0.1+j0	0−j0.1				
$\hat{\omega} - \omega_0 = 2\pi \times 6000$ or $vt =$ ———				TOTALS	57.9+j0	0+j1.7			0−j36.4	24.7+j0

$Q = (57.9+j0)+(0+j1.7)+(0-j36.4)+(24.7+j0) = 82.6-j34.7 = 89.5 \angle -22.8$ degrees

$R = (57.9+j0)-(0+j1.7)+(0-j36.4)-(24.7+j0) = 33.2-j38.1 = 50.5 \angle -48.9$ degrees

$S = (57.9+j0)+(0+j1.7)-(0-j36.4)-(24.7+j0) = 33.2+j38.1 = 50.5 \angle 48.9$ degrees

$T = (57.9+j0)-(0+j1.7)-(0-j36.4)+(24.7+j0) = 82.6+j34.7 = 89.5 \angle 22.8$ degrees

IV (t). Into column IV enter the value of

$$57.3 \times \frac{K}{v} \sin vt \left(1 - \cos \frac{2\pi mv}{\Omega}\right) \text{ degrees.}$$

V $(\hat{\omega})$. Into column V enter the value of

$$57.3 \times \frac{\hat{\omega} - \omega_0}{v} \sin \frac{2\pi mv}{\Omega} \text{ degrees.}$$

Example: For $m = 2$ on the work sheet for $(\hat{\omega} - \omega_0) = 2\pi \times 6000$

$$V = 57.3 \times \frac{2\pi \times 6000}{2\pi \times 1000} \sin \frac{2\pi \times 2 \times 2\pi \times 1000}{2\pi \times 80{,}000}$$

$$= 53.8 \text{ degrees.}$$

V (t). Into column V enter the value of

$$57.3 \times \frac{K}{v} \cos vt \sin \frac{2\pi mv}{\Omega} \text{ degrees.}$$

V. Enter the A_m's obtained in step C into column VI.

Example: From step C, $A_2 = 17.8 + j0$. This column is the same for all the work sheets.

VII. Multiply the A_m's, column VI, by the cosine of the angle in column V and enter the result into column VII.

Example: For $m = 2$,

$$(17.8 + j0) \cos 53.8 \text{ degrees} = 10.5 + j0.$$

VIII. Multiply column VII by the cosine of the angle in column IV and enter the result into column VIII.

Example: For $m = 2$,

$$(10.5 + j0) \cos 17.1 \text{ degrees} = 10.0 + j0.$$

IX. Multiply column VII by j sine of the angle in column IV and enter the result into column IX.

Example: For $m = 2$,

$$j\,(10.5 + j0) \sin 17.1 \text{ degrees} = 0 + j3.1.$$

X. Enter the B_m's obtained in step C into column X.

Example: From step C, $B_2 = 0 - j14.3$. This column is the same for all the work sheets.

XI. Multiply the B_m's, column X, by the sine of the angle in column V and enter the result into column XI.

Example: For $m = 2$,

$$(0 - j14.3) \sin 53.8 \text{ degrees} = 0 - j11.5.$$

XII. Multiply column XI by the cosine of the angle in column IV and enter the result into column XII.

Example: For $m = 2$,

$$(0 - j11.5) \cos 17.1 \text{ degrees} = 0 - j11.1.$$

XIII. Multiply column XI by j sine of the angle in column IV and enter the result into column XIII.

Example: For $m = 2$,

$$j\ (0 - j11.5)\ \sin 17.1 \text{ degrees} = 3.4 + j0.$$

XIV. Total columns VIII, IX, XII, and XIII on each work sheet. (See Table I.)

XV. Calculate the following quantities for each work sheet and express the result in polar form.

$$Q = \text{VIII} + \text{IX} + \text{XII} + \text{XIII}$$
$$R = \text{VIII} - \text{IX} + \text{XII} - \text{XIII}$$
$$S = \text{VIII} + \text{IX} - \text{XII} - \text{XIII}$$
$$T = \text{VIII} - \text{IX} - \text{XII} + \text{XIII}$$

where VIII, IX, XII, and XIII are the totals obtained in step XIV. (See Table I.)

XVI ($\hat{\omega}$). For each of the work sheets prepared for different assigned values of $(\hat{\omega} - \omega_0)$, plot on the response graphs,

$$Q = \text{the upsweep response at } \omega_0 + (\hat{\omega} - \omega_0)$$
$$R = \text{the downsweep response at } \omega_0 + (\hat{\omega} - \omega_0)$$
$$S = \text{the upsweep response at } \omega_0 - (\hat{\omega} - \omega_0)$$
$$T = \text{the downsweep response at } \omega_0 - (\hat{\omega} - \omega_0).$$

The magnitudes of Q, R, S, and T for each work sheet give four points on the desired amplitude — instantaneous-input-frequency response curve. The phases of Q, R, S, and T give four points on the desired phase-shift—instantaneous-input-frequency response curve.

Example: For Table I,

$$Q = 89.5 \angle -22.8 \text{ degrees} = \text{the upsweep response at } \hat{\omega} = 2\pi \times 1.006 \times 10^6$$

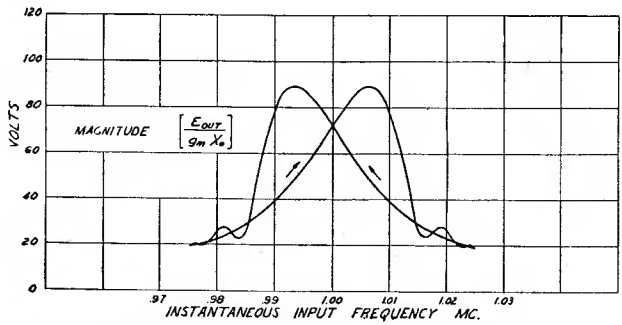
$$R = 50.5 \angle -48.9 \text{ degrees} = \text{the downsweep response at } \hat{\omega} = 2\pi \times 1.006 \times 10^6$$

$$S = 50.5 \angle 48.9 \text{ degrees} = \text{the upsweep response at } \hat{\omega} = 2\pi \times 0.994 \times 10^6$$

$T = 89.5 \angle 22.8$ degrees = the downsweep response at

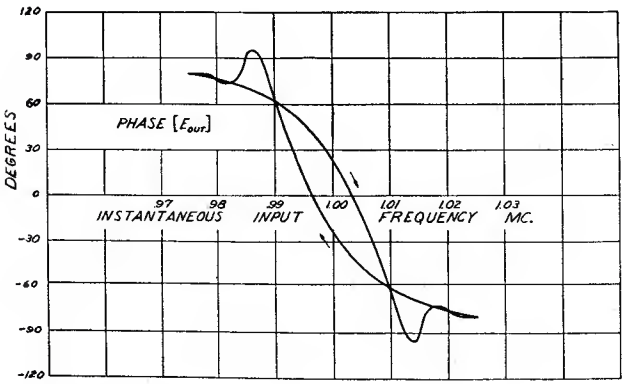
$$\hat{\omega} = 2\pi \times 0.994 \times 10^6$$

Plotting the points obtained on work sheets for $(\hat{\omega} - \omega_0) = 2\pi \times 2000$, $2\pi \times 4000$, $2\pi \times 6000$, $\cdots$, $2\pi \times 24{,}000$ determines curves (a) and (b)



(a)

(a)—Amplitude of $\dfrac{E_{\text{out}}}{g_m X_0}$



(b)

(b)—Phase of E_{out}.

Fig. 5—Calculated amplitude—instantaneous-input-frequency and phase-shift—instantaneous-input-frequency response of a tuned amplifier to a frequency-modulated wave.

Effective Q of amplifier	$= 100$	$(p = 0.01)$
Center frequency	$= 10^6$	$(\omega_0 = 2\pi \times 10^6)$
Modulating frequency	$= 1000$	$(v = 2\pi \times 1000)$
Total frequency excursion	$= 50{,}000$	$(K = 2\pi \times 25{,}000)$

of Figure 5. In this particular problem the upsweep response is symmetrical with the downsweep response because the steady-state network characteristic is symmetrical about ω_0. Note that curve (a) of Figure

5, the calculated amplitude—instantaneous-input-frequency response, agrees very closely with (b) of Figure 2, the experimental amplitude —instantaneous-input-frequency response. The reader may be interested in comparing the results of the calculation with the results obtained by Roder on a very similar problem, which he solved by the aid of a differential analyzer. His response curve was plotted as a function of time.[7]

XVI (t). For each of the work sheets prepared for different assigned values of vt, plot on the response graphs.

$$Q = \text{the response at } 2\pi - vt$$

$$R = \text{the response at } vt$$

$$S = \text{the response at } \pi + vt$$

$$T = \text{the response at } \pi - vt.$$

The magnitudes of Q, R, S, and T for each work sheet give four points on the desired amplitude—time response curve. The phases of Q, R, S, and T give four points on the desired phase-shift—time-response curve.

Appendix

If a steady-state driving voltage $\epsilon^{j\omega t}$ is applied to the input terminals of a linear network, it can be shown that the steady-state solution for the voltage at the output terminals[8] is of the form $R(\omega)\ \epsilon^{j[\omega t+\theta(\omega)]}$. $R(\omega)$ is the ratio of the peak amplitudes of the output and input voltages, and $\theta(\omega)$ is the phase difference. Abbreviating $\epsilon^{j\omega t}$ as $E_{\text{in}}{}^{+}$ and $R(\omega)\ \epsilon^{j[\omega t+\theta(\omega)]}$ as $E_{\text{out}}{}^{+}$, it follows that

$$E_{\text{out}}{}^{+} = E_{\text{in}}{}^{+} R(\omega)\ \epsilon^{j\theta(\omega)}. \tag{34}$$

Furthermore, if a steady-state driving voltage $\epsilon^{-j\omega t}$ is applied to the input terminals of the same linear network, it can be shown that the voltage at the output terminals is of the form $R(\omega)\ \epsilon^{-j[\omega t+\theta(\omega)]}$, where R and θ are equal to R and θ of (34). Setting $\epsilon^{-j\omega t}$ as $E_{\text{in}}{}^{-}$ and $R(\omega)\ \epsilon^{-j[\omega t+\theta(\omega)]}$ as $E_{\text{out}}{}^{-}$ it follows that

$$E_{\text{out}}{}^{-} = E_{\text{in}}{}^{-} R(\omega)\ \epsilon^{-j\theta(\omega)}. \tag{35}$$

---

[7] Hans Roder, "Effects of Tuned Circuits upon a Frequency-modulated Signal," *Proc. I.R.E.*, vol. 25, pp. 1617-1648; December, 1937.

[8] H. Pender and S. Warren, ELECTRIC CIRCUITS AND FIELDS, McGraw-Hill Book Company, New York, N. Y., 1943, pp. 161-163.

Therefore, if a steady-state driving voltage sin ωt is applied to the input terminals of the network, the output voltage can be determined as follows:

$$E_{\text{in}} = \sin \omega t = \frac{\epsilon^{j\omega t} - \epsilon^{-j\omega t}}{2j} \cdot \tag{36}$$

By (34) and (35),

$$E_{\text{out}} = \frac{R(\omega)\, \epsilon^{j[\omega t + \theta(\omega)]} - R(\omega)\, \epsilon^{-j[\omega t + \theta(\omega)]}}{2j} = R(\omega) \sin [\omega t + \theta(\omega)]. \tag{37}$$

$R(\omega)\, \epsilon^{j\theta(\omega)}$ is now defined as the complex network characteristic, $f(\omega)$. It can be seen from (36) and (37) that

$$E_{\text{out}} \{=\} E_{\text{in}} f(\omega) \tag{8}$$

where

$$| f(\omega) | = \frac{\text{amplitude of } E_{\text{out}}}{\text{amplitude of } E_{\text{in}}} \tag{9}$$

and where

$$\text{arc tan} \frac{\text{imaginary } [f(\omega)]}{\text{real } [f(\omega)]} = \text{angular phase } [E_{\text{out}}] - \text{angular phase } [E_{\text{in}}]. \tag{10}$$

The notation $\{=\}$ in (8) means "equals by convention," because the right- and left-hand members of the equation are not mathematically equivalent except when $\theta(\omega) = n\pi$.

Note also from (34) and (35) that

$$E_{\text{out}}{}^{+} = E_{\text{in}}{}^{+} f(\omega) \tag{38}$$

and

$$E_{\text{out}}{}^{-} = E_{\text{in}}{}^{-} f^{*}(\omega) \tag{39}$$

where $f^{*}(\omega) =$ the complex conjugate of $f(\omega)$. (40)

The $=$ signs of (38) and (39) indicate exact mathematical equivalence. Only the $\{=\}$ sign of (8) is restricted.

Equation (7), which expresses the input frequency-modulation voltage as a spectrum of steady-state side frequencies, is repeated here for convenience.

$$e_{\text{in}} = J_0(m_f) \sin \omega_0 t + \sum_{n=1}^{N} (J_{2n-1}(m_f) \{\sin [\omega_0 + (2n-1)\, v\,]\, t - \sin [\omega_0 - (2n-1)\, v\,]\, t\} + J_{2n}(m_f) \{\sin [\omega_0 + 2nv]\, t + \sin [\omega_0 - 2nv]\, t\}). \quad (7)$$

Since $J_n(z) = (-1)^n J_{-n}(z)$, equation (7) can be compressed as follows:

$$e_{\text{in}} = \sum_{\eta=-2N}^{2N} J_\eta(m_f) \sin (\omega_0 + \eta v)t. \quad (41)$$

Changing $\sin (\omega_0 + \eta v)t$ to exponential form,

$$2je_{\text{in}} = \sum_{\eta=-2N}^{2N} J_\eta(m_f) [\epsilon^{j(\omega_0+\eta v)t} - \epsilon^{-j(\omega_0+\eta v)t}]. \quad (42)$$

The steady-state output oscillation corresponding to each steady-state oscillation of the frequency-modulated input spectrum is given by (38) and (39).

$$E_{\text{out}}{}^+[\text{for } E_{\text{in}}{}^+ = \epsilon^{j(\omega_0+\eta v)t}] = \epsilon^{j(\omega_0+\eta v)t} f(\omega_0 + \eta v) \quad (43)$$

and

$$E_{\text{out}}{}^-[\text{for } E_{\text{in}}{}^- = \epsilon^{-j(\omega_0+\eta v)}] = \epsilon^{-j(\omega_0+\eta v)t} f^*(\omega_0 + \eta v) \quad (44)$$

Equation (11) gives the values of $f(\omega_0 + \eta v)$ and $f^*(\omega_0 + \eta v)$.

$$f(\omega_0 + \eta v) \approx \sum_{m=0}^{M} (A_m \cos km\eta v + B_m \sin km\eta v) \quad (45)$$

and

$$f^*(\omega_0 + \eta v) \approx \sum_{m=0}^{M} (A_m{}^* \cos km\eta v + B_m{}^* \sin km\eta v) \quad (46)$$

where

$$k = \frac{2\pi}{\Omega}. \quad (47)$$

The output-voltage equation can now be written from (42), (43), and (44).

$$2je_{\text{out}} = \sum_{m=0}^{M} \sum_{\eta=-2N}^{2N} J_\eta(m_f)\,[\epsilon^{j(\omega_0+\eta v)\,t}(A_m \cos km\eta v + B_m \sin km\eta v) - \epsilon^{-j(\omega_0+\eta v)\,t}(A_m^* \cos km\eta v + B_m^* \sin km\eta v)]. \tag{48}$$

Changing $\cos km\eta v$ and $\sin km\eta v$ to exponential form,

$$2je_{\text{out}} = \sum_{m=0}^{M} \sum_{\eta=-2N}^{2N} J_n(m_f)$$

$$\left[\frac{A_m}{2}(\epsilon^{j[(\omega_0+\eta v)\,t+km\eta v]} + \epsilon^{j[(\omega_0+\eta v)\,t-km\eta v]}) + \frac{B_m}{2j}(\epsilon^{j[(\omega_0+\eta v)\,t+km\eta v]} - \epsilon^{j[(\omega_0+\eta v)\,t-km\eta v]}) - \frac{A_m^*}{2}(\epsilon^{j[-(\omega_0+\eta v)\,t+km\eta v]} + \epsilon^{j[(-\omega_0+\eta v)\,t-km\eta v]}) - \frac{B_m^*}{2j}(\epsilon^{j-[(\omega_0+\eta v)\,t+km\eta v]} - \epsilon^{j-[(\omega_0+\eta v)\,t-km\eta v]})\right]. \tag{49}$$

The Bessel coefficients are combined with the exponentials by the use of the relation,

$$\sum_{\eta=-2N}^{2N} J_\eta(m_f)\,\epsilon^{j\eta\theta} \approx \epsilon^{jmf \sin\theta},$$

which is accurate because the $J_\eta(m_f)$'s below $\eta = -2N$ and above $\eta = 2N$ are negligible by the definition of N.

$$2je_{\text{out}} = \sum_{m=0}^{M}\left[\frac{A_m}{2}(\epsilon^{j[\omega_0 t\,+\,mf \sin(vt\,+\,kmv)]} + \epsilon^{j[\omega_0 t+mf \sin(vt-kmv)]}) + \frac{B_m}{2j}(\epsilon^{j[\omega_0 t+mf \sin(vt+kmv)]} - \epsilon^{j[\omega_0 t+mf \sin(vt-kmv)]}) - \frac{A_m^*}{2}(\epsilon^{j[-\omega_0 t+mf \sin(-vt+kmv)]} + \epsilon^{j[-\omega_0 t+mf \sin(-vt-kmv)]}) - \frac{B_m^*}{2j}(\epsilon^{j[-\omega_0 t+mf \sin(-vt+kmv)]} - \epsilon^{j[-\omega_0 t+mf \sin(-vt-kmv)]})\right]. \tag{50}$$

Expanding $\sin(vt \pm kmv)$ and factoring the exponentials,

$$2je_{\text{out}} = \sum_{m=0}^{M} \left[\frac{A_m}{2} \left(\epsilon^{j(\omega_0 t + mf \sin vt \cos kmv)}\right) \left(\epsilon^{jmf \cos vt \sin kmv}\right.\right.$$

$$\left. + \epsilon^{-jmf \cos vt \sin kmv}\right) + \frac{B_m}{2j} \left(\epsilon^{j(\omega_0 t + mf \sin vt \cos kmv)}\right)$$

$$\left(\epsilon^{jmf \cos vt \sin kmv} - \epsilon^{-jmf \cos vt \sin kmv}\right)$$

$$- \frac{A_m^*}{2} \left(\epsilon^{-j(\omega_0 t + mf \sin vt \cos kmv)}\right) \left(\epsilon^{jmf \cos vt \sin kmv}\right.$$

$$\left. + \epsilon^{-jmf \cos vt \sin kmv}\right) - \frac{B_m^*}{2j} \left(\epsilon^{-j(\omega_0 t + mf \sin vt \cos kmv)}\right)$$

$$\left.\left(\epsilon^{jmf \cos vt \sin kmv} - \epsilon^{-jmf \cos vt \sin kmv}\right)\right] . \tag{51}$$

Factoring $\epsilon^{\pm j(\omega_0 t + mf \sin vt \cos kmv)}$ and changing the other exponentials to trigonometric form,

$$\begin{aligned} 2je_{\text{out}} = \sum_{m=0}^{M} [& A_m \cos(m_f \cos vt \sin kmv) \epsilon^{jmf \sin vt(\cos kmv - 1)} \epsilon^{j(\omega_0 t + mf \sin vt)} \\ & + B_m \sin(m_f \cos vt \sin kmv) \epsilon^{jmf \sin vt(\cos kmv - 1)} \epsilon^{j(\omega_0 t + mf \sin vt)} \\ & - A_m^* \cos(m_f \cos vt \sin kmv) \epsilon^{-jmf \sin vt(\cos kmv - 1)} \epsilon^{-j(\omega_0 t + mf \sin vt)} \\ & - B_m^* \sin(m_f \cos vt \sin kmv) \epsilon^{-jmf \sin vt(\cos kmv - 1)} \epsilon^{-j(\omega_0 t + mf \sin vt)}] . \end{aligned} \tag{52}$$

$$\begin{aligned} 2je_{\text{out}} = & \epsilon^{j(\omega_0 t + mf \sin vt)} \sum_{m=0}^{M} D_m [A_m \cos \psi_m + B_m \sin \psi_m] \\ & - \epsilon^{-j(\omega_0 t + mf \sin vt)} \sum_{m=0}^{M} D_m^* [A_m^* \cos \psi_m + B_m^* \sin \psi_m] \end{aligned} \tag{53}$$

where

$$D_m(t) = \cos\left[\frac{K}{v} \sin vt \left(1 - \cos \frac{2\pi mv}{\Omega}\right)\right] - j \sin \left[\frac{K}{v} \sin vt \left(1 - \cos \frac{2\pi mv}{\Omega}\right)\right] \tag{16}$$

and where $$\psi_m(t) = \frac{K}{v} \cos vt \sin \frac{2\pi mv}{\Omega}. \tag{17}$$

It is easy to derive $D_m(\hat{\omega})$ and $\psi_m(\hat{\omega})$ from (16) and (17) by starting with (5), which relates $\hat{\omega}$ and t.

$$\hat{\omega} = \omega_0 + K \cos vt \tag{5}$$

$$\cos vt = \frac{\hat{\omega} - \omega_0}{K} \tag{54}$$

$$\sin vt = \frac{\pm \sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{K}. \tag{55}$$

Substituting for sin vt in (16),

$$D_m(\hat{\omega}) = \cos \left[\frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(1 - \cos \frac{2\pi mv}{\Omega} \right) \right] \pm j \sin \left[\frac{\sqrt{K^2 - (\hat{\omega} - \omega_0)^2}}{v} \left(1 - \cos \frac{2\pi mv}{\Omega} \right) \right]. \tag{18}$$

Substituting for cos vt in (17),

$$\psi_m(\hat{\omega}) = \frac{\hat{\omega} - \omega_0}{v} \sin \frac{2\pi mv}{\Omega}. \tag{19}$$

Equation (53) can be changed from exponential form to sinusoidal form as follows:

Let $$P = \left| \sum_{m=0}^{M} D_m \left[A_m \cos \psi_m + B_m \sin \psi_m \right] \right| \tag{56}$$

and

$$\phi = \arctan \frac{\text{imaginary} \left\{ \sum_{m=0}^{M} D_m \left[A_m \cos \psi_m + B_m \sin \psi_m \right] \right\}}{\text{real} \left\{ \sum_{m=0}^{M} D_m \left[A_m \cos \psi_m + B_m \sin \psi_m \right] \right\}}. \tag{57}$$

then

$$2je_{\text{out}} = \epsilon^{j(\omega_0 t + mf \sin vt)} P\epsilon^{j\phi} - \epsilon^{-j(\omega_0 t + mf \sin vt)} P\epsilon^{-j\phi} \tag{58}$$

$$e_{\text{out}} = \frac{P}{2j} [\epsilon^{j(\omega_0 t + mf \sin vt + \phi)} - \epsilon^{-j(\omega_0 t + mf \sin vt + \phi)}] \tag{59}$$

$$e_{\text{out}} = P \sin (\omega_0 t + m_f \sin vt + \phi). \tag{60}$$

The output-voltage-envelope equation, therefore, written in conventional form is

$$E_{\text{out}} \{=\} \sum_{m=0}^{M} D_m [A_m \cos \psi_m + B_m \sin \psi_m]. \tag{15}$$

The notation $\{=\}$ again means "equals by convention"; the amplitude of the envelope is the absolute value of the complex summation, and the angular phase difference between the output and input frequency-modulated waves is equal to the phase of the complex summation.

ACKNOWLEDGMENT

The author wishes to express his appreciation for the many suggestions and constructive criticisms of this paper in the draft stage offered by members of the Test Engineering Section of RCA Victor, and to the department management, who supplied laboratory and clerical facilities; Mr. L. C. Smith suggested that this material be prepared for publication, and supervised the work. Mr. M. S. Corrington of the Advanced Development Section recommended changes that strengthened the text, suggested identities which greatly simplified the appendix, and cited additional pertinent references.

* * * * * * * * *

NETWORK TRANSMISSION OF A FREQUENCY MODULATED WAVE†

BY

L. J. GIACOLETTO

Research Department, RCA Laboratories Division,
Princeton, N. J.

IN A RECENT paper[1] Frantz described a method for obtaining the output signal from a network when the input signal is frequency modulated. (For an excellent generalized treatment of this prob-

† Reprinted from *Proc. I.R.E.*, October, 1947.

[1] W. J. Frantz, "The transmission of a frequency-modulated wave through a network," *Proc. I.R.E.*, vol. 34, pp. 114P-125P; March, 1946.

lem, when the network characteristics are expressed by means of a power series, or by means of a trigonometric series, the reader is referred to a paper by Block.[2]) In the application of Frantz's method it is nccessary that the network response be expressed as a Fourier series or, as was also indicated, by other series approximations. However, it is possible to obtain the amplitude and phase of the output signal by using point-by-point data from the network response curve; contrary to Frantz's statement, this can be accomplished without plotting the output r.f. signal point by point as a function of time. The additional work of obtaining a series expression, as well as questions as to the proper convergence of this expression, are thus avoided. It is believed that the method of calculation to be described is, therefore, simpler and more straightforward, and in any event may be of interest as an alternate method of solution.

To rephrase the problem: a network having a gain-versus-frequency characteristic given by $R(\omega)$ and a phase-versus-frequency characteristic given by $\theta(\omega)$ is driven by a single-tone f.m. voltage (assumed to be of unit amplitude for convenience, as network is assumed to be linear with amplitude) given by

$$e_{\text{in}} = \sin(\omega_0 t + m_f \sin vt) = \sum_{n=-\infty}^{\infty} J_n(m_f) \sin(\omega_0 + nv)\, t. \qquad (1)$$

It is desired to determine the output voltage from the network.

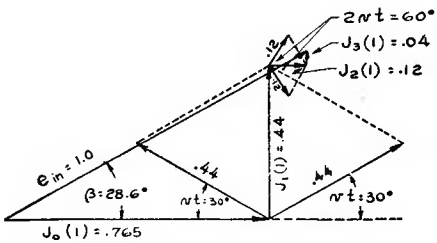
The input signal indicated in (1) consists of a multiplicity of individual sinusoidal signals comprising the carrier and the upper- and lower-side frequencies. These individual sinusoidal signals can be pictured as vectors of proper length given by $J_n(m_f)$ rotating counterclockwise about the center at an angular frequency given by $(\omega_0 + nv)$. The upper- and lower-side frequencies can be paired off by rewriting (1) as

$$\begin{aligned} e_{\text{in}} = {} & J_0(m_f) \sin \omega_0 t \\ & + J_1(m_f)\,[\sin(\omega_0 + v)\, t - \sin(\omega_0 - v)\, t] \\ & + J_2(m_f)\,[\sin(\omega_0 + 2v)\, t + \sin(\omega_0 - 2v)\, t] \\ & + J_3(m_f)\,[\sin(\omega_0 + 3v)\, t - \sin(\omega_0 - 3v)\, t] + \cdots . \end{aligned} \qquad (2)$$

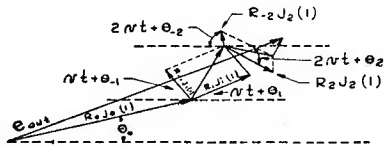
To better visualize the action that is taking place, it is convenient to assume that the frame of reference rotates counterclockwise with an angular frequency ω_0 so that the carrier vector $J_0(m_f)$ remains stationary. If this is done, then the vectors with the higher angular fre-

[2] A. Block, "Modulation theory," *Jour. I.E.E.* (London), Pt. III, vol. 91, pp. 31-42; March, 1944.

quencies will rotate counterclockwise with angular frequencies of nv, and the vectors with the lower angular frequencies will rotate clockwise with the same angular frequencies nv. This is pictured in Figure 1(a) for the case where $m_f = 1.0$ and $vt = 30°$. It is readily seen from this figure (as well as from (2) by subtracting ω_0 from each angular frequency) that the even-order sidebands pair off to give a vector sum in phase with the carrier vector and the odd-order sidebands pair off to give a vector sum in phase-quadrature with the carrier vector. Thus it is seen either from Figure 1(a) or from (2) that



(a)



(b)

Fig. 1—Vector summation of (a) input signal, and (b) output signal for an arbitrary network (assuming modulation index = 1.0, $vt = 30°$).

$$|e_{in}| = \sqrt{\begin{aligned}&[J_0(m_f) + 2J_2(m_f)\cos 2vt + 2J_4(m_f)\cos 4vt + \cdots]^2 \\ &+ [2J_1(m_f)\sin vt + 2J_3(m_f)\sin 3vt + \cdots]^2\end{aligned}} \qquad (3)$$

and

$$\beta = \text{phase of } e_{in} \text{ with respect to signal of phase } \omega_0 t = m_f \sin vt = \arctan \left\{ \frac{[2J_1(m_f)\sin vt + 2J_3(m_f)\sin 3vt + \cdots]}{[J_0(m_f) + 2J_2(m_f)\cos 2vt + 2J_4(m_f)\cos 4vt + \cdots]} \right\} \qquad (4)$$

The network output voltage is obtained, as indicated by Frantz, by treating each of the sideband terms as an individual signal and altering its amplitude by $R_n(\omega)$ and its phase by $\theta_n(\omega)$. Thus, the output voltage derived from (2) is

$$e_{\text{out}} = R_0 J_0(m_f) \sin(\omega_0 t + \theta_0) + J_1(m_f)[R_1 \sin\{(\omega_0 + v)t + \theta_1\} - R_{-1} \sin\{(\omega_0 - v)t + \theta_{-1}\}] + J_2(m_f)[R_2 \sin\{(\omega_0 + 2v)t + \theta_2\} + R_{-2} \sin\{(\omega_0 - 2v)t + \theta_{-2}\}] + \cdots . \quad (5)$$

If (5) is pictured on the counterclockwise rotating reference frame, it is seen (as indicated in Figure 1(b) for same conditions as 1(a) with hypothetical network characteristics) that

$$|e_{\text{out}}| = \sqrt{\{R_0 J_0(m_f) \cos\theta_0 + J_1(m_f)[R_1 \cos(vt + \theta_1) - R_{-1} \cos(vt - \theta_{-1})] + J_2(m_f)[R_2 \cos(2vt + \theta_2) + R_{-2} \cos(v - \theta_{-2})] + \cdots\}^2 + \{R_0 J_0(m_f) \sin\theta_0 + J_1(m_f)[R_1 \sin(vt + \theta_1) + R_{-1} \sin(vt - \theta_1)] + J_2(m_f)[R_2 \sin(2vt + \theta_2) - R_{-2} \sin(2vt - \theta_{-2})] + \cdots\}^2} \quad (6)$$

and

β phase of e_{out} with respect to signal of phase $\omega_0 t$ = arc tan

$$\left\{\frac{R_0 J_0(m_f) \sin\theta_0 + J_1(m_f)[R_1 \sin(vt + \theta_1) + R_{-1} \sin(vt - \theta_{-1})] + \cdots}{R_0 J_0(m_f) \cos\theta_0 + J_1(m_f)[R_1 \cos(vt + \theta_1) - R_{-1} \cos(vt - \theta_{-1})] + \cdots}\right\} \quad (7)$$

Using (6) and (7) the point-by-point amplitude and phase of the output voltage as a function of vt can be obtained either by algebraic or graphical calculation. Sufficient terms are retained to give the desired accuracy with more terms being necessary when m_f is large. If the series indicated in (6) can be summed up analytically, then a closed analytic expression for the output is possible.

If the network-gain characteristic is symmetrical so that $R_{-n} = R_n$, and the network phase characteristic is skew symmetrical so that $\theta_{-n} = -\theta_n$ (which is the case treated in detail by Frantz) then (6) and (7) simplify to

$$|e_{\text{out}}| = \sqrt{\{R_0 J_0(m_f) + 2R_2 J_2(m_f) \cos(2vt + \theta_2) + 2R_4 J_4(m_f) \cos(4vt + \theta_4) + \cdots\}^2 + \{2R_1 J_1(m_f) \sin(vt + \theta_1) + 2R_3 J_3(m_f) \sin(3vt + \theta_3) + \cdots\}^2}$$

and (8)

β = arc tan

$$\left\{\frac{[2R_1 J_1(m_f) \sin(vt + \theta_1) + 2R_3 J_3(m_f) \sin(3vt + \theta_3) + \cdots]}{[R_0 J_0(m_f) + 2R_2 J_2(m_f) \cos(2vt + \theta_2) + 2R_4 J_4(m_f) \cos(4vt + \theta_4) + \cdots]}\right\} . \quad (9)$$

If $vt = 180° + \gamma$ it is seen that $|e_{out}|$ is the same as for $vt = \gamma$, and that β is negative the value for $vt = \gamma$, so that for (8) and (9) only the region of $vt = \theta$ to $vt = 180°$ need be investigated. In order to compare (8) and (9) with similar results by Frantz, one should keep in mind that the phase angle of e_{out} used by Frantz is the phase of e_{out} with respect to the pase of the FM wave. Thus

Table I
Computation of Points on Curves
Shown in Figure 5 in Article by Frantz[1] for $vt = 60°$
Instantaneous Frequency = 1.0125 Mc.

n	C_n	$nvt - \theta_n$	$C_n \cos (nvt - \theta_n)$
0	+10	0°	+10
2	−20	95	+ 2
4	+21	−163	−20
6	−21	− 51	−13
8	+16	63	+ 7
10	− 7	177	+ 7
12	− 6	− 67	− 2
14	+12	49	+ 8
16	− 4	166	+ 4
18	− 9	− 76	− 2
20	+ 2	42	+ 2
22	+ 9	161	− 9
24	+ 8	− 80	+ 1
26	+ 4	39	+ 3
28	+ 1	158	− 1
			− 3

n	C_n	$nvt - \theta_n$	$C_n \sin (nvt - \theta_n)$
1	−25	48°	−18
3	+19	143	+11
5	− 9	−108	+ 9
7	− 1	6	0
9	+10	120	+ 9
11	−14	−125	+12
13	+ 7	− 9	− 1
15	+ 6	107	+ 6
17	−10	−135	+ 7
19	− 4	− 17	+ 1
21	+ 7	102	+ 7
23	+ 9	140	− 6
25	+ 5	− 20	− 2
27	+ 2	99	+ 2
29	+ 1	−143	0
			+37

$C_n = Kf(\omega)/g_m x\, J_n(25)$; $K = 1$ when $n = 0$ and $K = 2$ when $n \neq 0$; $f(\omega)/g_m x_0 = R(\omega)$ is interpolated from Figure 4(c) of the article;[1] θ_n is interpolated from Figure 4(d) of the article. Then $|E_{out}| = \sqrt{(-3)^2 + (37)^2} = 37$; $\beta = \tan^{-1}(+37/-3) = 95°$, and since $25 \left(\frac{180}{\pi}\right)^\circ \sin 60° = 160°$, $\phi = \beta - 160° = -65°$. For $vt = 180° + 60°$, instantaneous frequency = 0.9875 Mc. $|e_{out}| = 37$, $\beta = -95°$, and, $\phi = \beta + 160° = 65°$.

θ = phase of e_{out} with respect to the phase of the input

$$\text{FM wave} = \beta - m_f \sin vt. \tag{10}$$

A sample calculation of e_{out} is shown in Table I for $vt = 60°$ for the same e_{in} used by Frantz, and using values of $R(\omega)$ and $\theta(\omega)$, interpolated from curves given by him.

It is perhaps worth noting that the complete value of e_{out} is known, since its amplitude is given by $|e_{out}|$ and its phase is $\omega_0 t + \beta$ or $\omega_0 t + m_f \sin vt + \theta$, i.e., $e_{out} = |e_{out}| \sin(\omega_0 t + \beta)$. In the event one is interested only in the frequency distortion caused by transmission through a network (as, for instance, when limiters remove all amplitude variation), the angular frequency distortion term is given by the time derivative of θ with respect to time.

PUSH-PULL FREQUENCY MODULATED CIRCUIT AND ITS APPLICATION TO VIBRATORY SYSTEMS*†

By

Alexis Badmaieff

RCA Victor Division, Hollywood, Calif.

Summary—*The purpose of this paper is to describe a new push-pull frequency modulated circuit in which the push-pull action is accomplished by varying the resonant frequencies of both the oscillator and the discriminator in opposite phase relation to each other. Modulation of the oscillator and discriminator is achieved through the use of 2 capacitors with a common plate and arranged in a push-pull circuit. Thus, if the common plate is the moving element of a vibratory system, this circuit can be used for measuring vibrations or for monitoring purposes.*

One of the applications of the push-pull FM circuit is to calibrate recording heads while cutting phonograph records. Using this calibrator, it is possible to measure, without mechanical or inductive coupling, the actual frequency response of the cutting head under normal load, and to observe the input-output characteristics as well as the amount of distortion. It is also possible to use it for monitoring while recording.

1—Basic Push-Pull Frequency Modulated Circuit

CONSIDERABLE interest is being shown in the development of frequency modulated circuits in which the oscillator and the discriminator are combined in one unit. This type of circuit can be used to convert mechanical vibrations to electrical voltage variations and is applicable to measuring instruments, pickup devices, and other forms of transducers.[1] The purpose of this paper is to discuss a new type of FM circuit in which the advantages of push-pull action are utilized.

In a single-ended FM circuit,[1] when used in conjunction with a very small capacitor as the frequency controlling element, a nonlinear relation exists between the condenser plates spacing and the frequency which the condenser controls. The magnitude of nonlinearity depends on the ratio between the spacing of the plates and the amount of variation of that spacing when the plates are vibrated mechanically. To illustrate, consider an LC circuit, in which one condenser plate is moved away from the other plate during tuning. Progressively increased separation has less effect on the resonant frequency because

* Decimal Classification: R148.2 × R621.375.622
† Reprinted from *Jour. Soc. Mot. Pic. Eng.*, January, 1946.

the latter varies approximately as the square root of the distance of separation. This means that a nonlinear relation between the frequency and the displacement results in even harmonic distortion. It is well known that by combining the voltages from 2 similar nonlinear components in any push-pull system, the distortion can be greatly reduced, since the even harmonic components are practically eliminated.

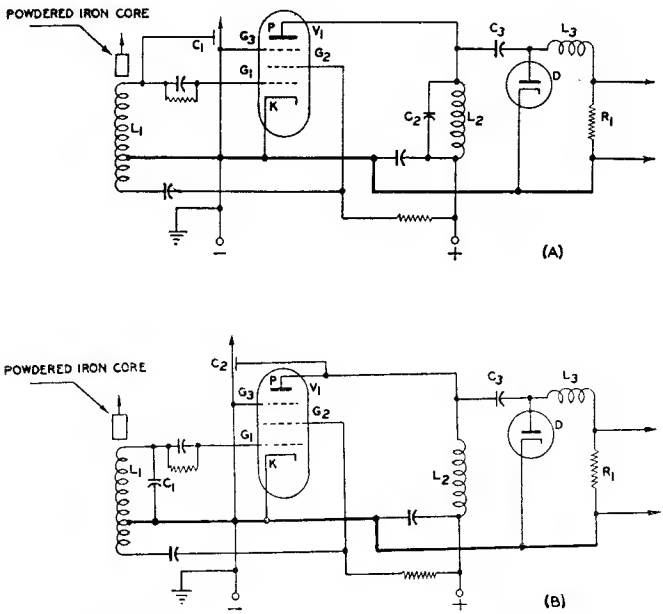


Fig. 1—(A) Modulated oscillator FM circuit.
(B) Modulated discriminator FM circuit.

The distortion owing to the nonlinear relation between condenser spacing and frequency, however, can be canceled in a single-sided FM circuit by a complementary distortion in the discriminator circuit. This is not easy to attain, however, because the ratio of the spacing and the change of that spacing, which produces a given distortion, must match inversely the complementary nonlinear slope of the discriminator curve. This condition can only be satisfied when the distortion produced by the condenser plates does not exceed the complementary nonlinear action of the discriminator. This indicates that the change in condenser plate spacing must be small in comparison with the average spacing (in the order of 5 per cent). To meet this require-

ment, and to produce the swing in frequency necessary to attain an appreciable amount of output, comparatively large plate area must be used. If, however, the plates are made small in comparison and their movement is large (in the order of 30 per cent) to secure a usable output, the second harmonic distortion produced is then quite large, and was measured to be about 17 per cent in one instance. This large amount of second harmonic content is too great to be canceled by the complementary nonlinear discriminator curve, except very close to the resonance point, where the range is limited and adjustment is critical,

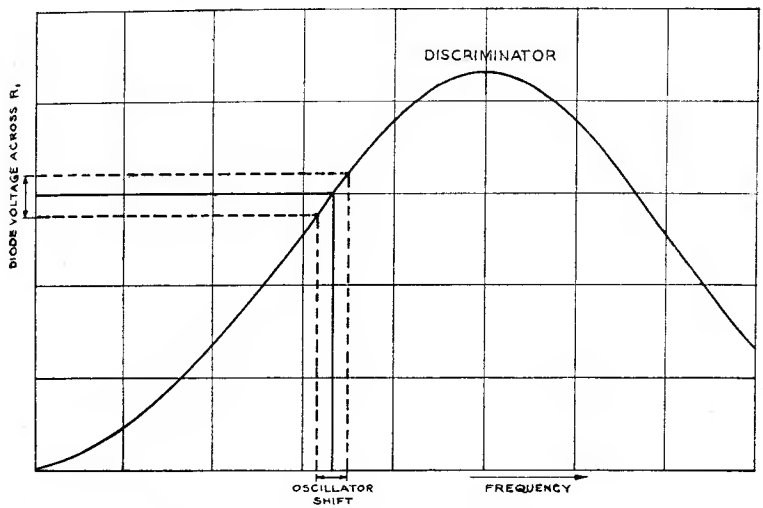


Fig. 2—Graphical representation of the modulated oscillator FM action.

or far off resonance where the sensitivity is low, as shown later in this paper under Section 2.

Another consideration is that, in a single-sided FM circuit, only one side of the movable plate is active. The other is inactive since it is not faced by another plate. In the push-pull FM circuit, both sides of the common center plate are active, since each side forms one plate of 2 condensers. This fact reduces the effective size of the movable plate to a half of what is necessary in the case of a single-ended circuit. Further reduction in size is attained by a larger movement in a small spacing, as mentioned before, which produces an adequate frequency swing by using 30 per cent of the spacing for the total maximum movement. These 2 factors can be advantageously employed in

designing a translating device having a minimum mass and therefore contributing to a minimum mechanical impedance.

A typical single-ended FM circuit in its elementary form is illustrated in Figure 1A. A pentode (V_1) with its cathode (K), grid (G_1), and screen (G_2) is combined with the tuned circuit (L_1) and a variable condenser (C_1) to form an oscillatory circuit electron-coupled[2] to the plate (P) and electrostatically shielded by grid (G_3) (which is normally the suppressor element of the tube). The oscillatory energy from the plate is applied to a circuit which consists of another tuned coil (L_2) and condenser (C_2). This circuit is slightly off resonance

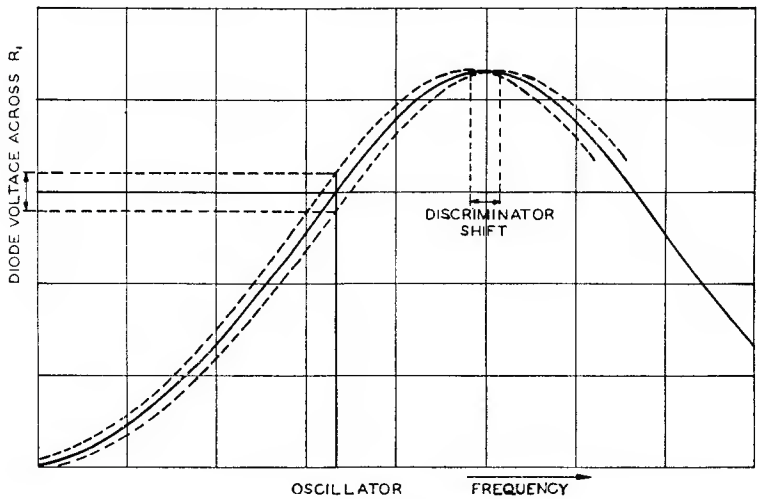


Fig. 3—Graphical representation of the modulated discriminator FM action.

relative to the mean frequency of the oscillator and forms the discriminator part of the FM system. The condenser (C_1) across the oscillator coil (L_1) is the controlling element of the FM circuit. The output from the discriminator tuned inductance is coupled through a condenser (C_3) and rectified by a diode (D), after which it is filtered of its rf component by a choke (L_3) and appears as a varying voltage across the diode load resistor (R_1). The function of this system is graphically represented in Figure 2, in which the oscillator mean frequency is positioned close to the discriminator's resonant peak. The dotted lines represent the variation of the oscillator frequency and its equivalent change in the voltage amplitude of the output of the rectifier and filter circuits.

Suppose the circuit just described were again used, but instead of modulating the oscillator, the discriminator were modulated by shunting a variable condenser across the resonant coil (L_2). As illustrated in Figure 1B, the frequency of the resonant peak of the discriminator would move back and forth along the frequency axis. The output voltage of the discriminator would be identical to that produced in the case in which the oscillator frequency was modulated. The oscillator frequency will still cover the same part of the discriminator peak as before. This is graphically illustrated in Figure 3 in which

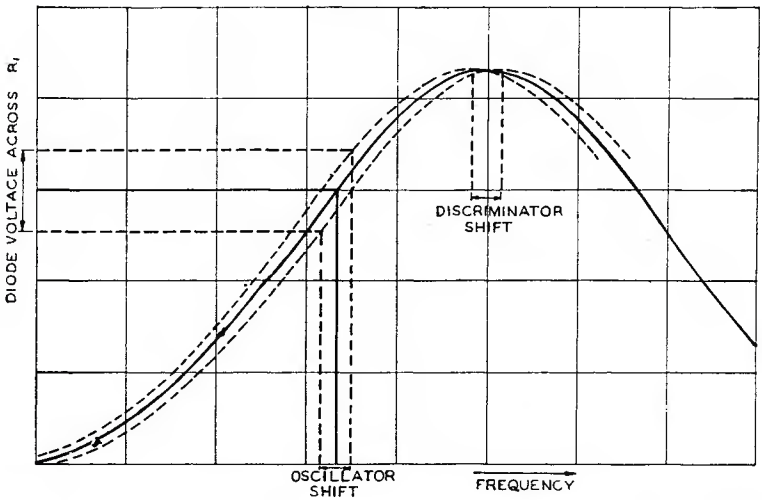


Fig. 4—Graphical representation of both oscillator and discriminator combined modulations in the push-pull FM system.

the dotted lines represent the displacement of the discriminator's resonant peak. If we combine both methods of modulation in the same circuit in such a manner that the oscillator frequency shifts in the opposite direction to the discriminator's resonant frequency, the voltage output from the discriminator would be twice as great as would be the case if only one of the factors had been varied through the same frequency range. If, for instance, the oscillator frequency moved down along the frequency axis at the same time the discriminator resonant peak moved up, the relative shift along the frequency axis would be doubled, producing twice the amplitude in the discriminator voltage output. This is illustrated in Figure 4. Since both modulations have to be in opposite phase to produce the push-pull action, and since a simple controlling factor is a capacity change in the 2 tuned circuits,

the push-pull FM circuit is particularly adapted to the use of 2 capacitors arranged for push-pull action.

The combining of 2 frequency modulations is accomplished by employing a small condenser, one side of which is connected across the oscillator coil and the other across the discriminator coil. The common center plate is grounded. If the center grounded plate is moved in either one of 2 directions, the frequency shifts of the oscillator and the discriminator circuits will be in opposite directions to each other resulting in push-pull action. Such a circuit is shown in Figure 5. In this circuit, coil (L_1) and one half of the push-pull condenser (C_1), together

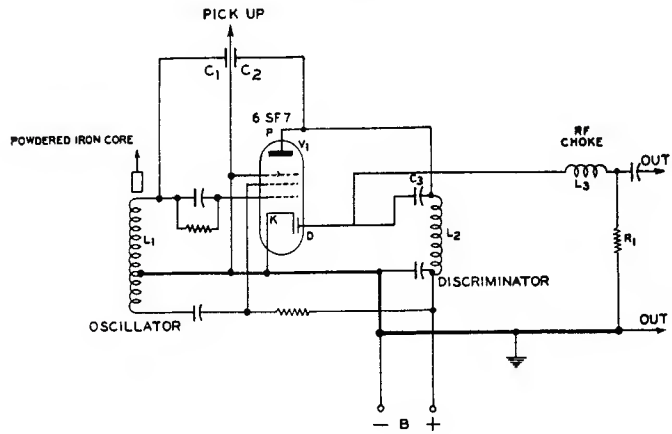


Fig. 5—Push-pull FM circuit combining both the oscillator and discriminator modulations.

with the grid, cathode, and screen of the pentode-diode tube, such as a 6SF7 (V_1), form the oscillator, and the coil (L_2), the other half of the push-pull condenser (C_2), and the plate of the tube form the discriminator circuit. These 2 circuits are electron-coupled, and shielded from each other within the tube by its suppressor which is grounded. The output is then rectified and filtered by the diode in the same tube, and the choke (L_3).

It should be mentioned, at this point, that the push-pull feature applies only to the action of the variable capacities. No claim is made to any cancellation of distortion because of the nonlinearity in the discriminator characteristic. Maintaining low distortion in the discriminator circuit depends upon the relatively small range of frequency change. The full benefit of the push-pull action in reducing distortion

will not be realized unless the 2 parts of the system are closely balanced with respect to each other. To realize this, inductances of the coils must be alike and the mechanical construction of both sides of the push-pull capacities must be alike so to give substantially equal capacity, equally varied in opposite phase.

The balance between the 2 modulated tuned circuits was checked by modulating the oscillator and the discriminator separately, and noting the outputs in each case. This was done by substituting an equivalent fixed capacitor for the discriminator side of the push-pull capacitor and with a constant modulation applied mechanically to the oscillator part of the push-pull capacitor, the output was then noted. When substituting a fixed condenser of equal value in the oscillator circuit and modulating the discriminator, the output was found to be of the same

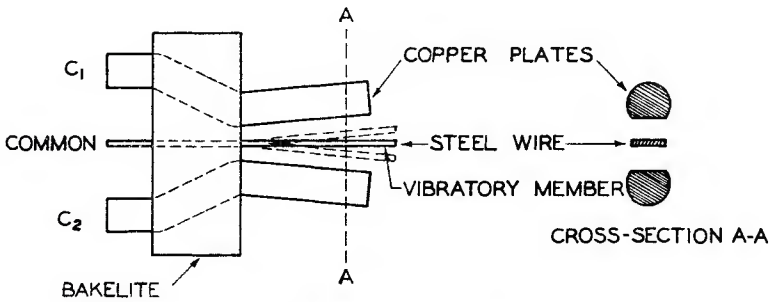


Fig. 6—One form of push-pull FM transducer.

amplitude as in the case when the oscillator was modulated. In other words, both circuits were designed to produce equal output voltage variations.

In each case, however, the exact oscillator center frequency position on the discriminator curve was noted, and the fixed substituted capacity adjusted so that the position was the same. This is important, since if the position were to change, a different amount of diode voltage would be developed for a given modulation, depending on the slope of the discriminator curve in the section used for discrimination. This position was noted by observing a microammeter connected between ground and the diode load resistor. The reading then was highest at the peak and proportional to the position of the oscillator center frequency on the discriminator curve. By thus observing the reading, the position and the side of the discriminator curve can be duplicated in each case.

Since the output voltage of the 2 sides of a push-pull system are additive, only half the voltage is required from each side of the system and, therefore, only half the frequency change to produce it. It is, therefore, evident that the vibratory member is capable of producing a large voltage amplitude with a relatively small capacity, resulting in a very small unit.

2—Operation

The push-pull FM circuit, built and tested, included a push-pull condenser as its controlling element. This condenser consisted of a

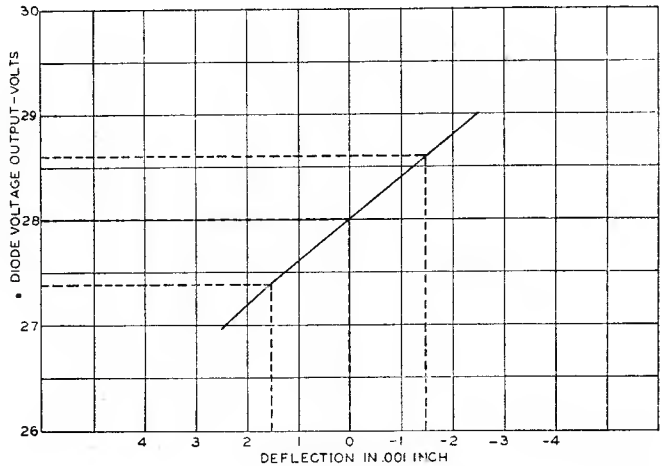


Fig. 7—Linearity of the push-pull FM system (distortion).

steel wire ⅛ in. long, 20 mils wide, and 8 mils thick, on each side of which is a plate of similar dimensions and spaced about 5 mils from the center movable plate as illustrated in Figure 6. The movable steel wire, which is the common grounded element of the push-pull condenser, is clamped at one end. Its free end can be attached to any mechanical system whose vibrations are to be measured, as, for instance, the stylus tip engaged in the groove of a phonograph record. The rigidity of this wire is low, being in the order of 1.4×10^6 dynes per centimeter. The 2 outside plates as previously described are connected; one across the oscillator coil, and the other across the discriminator coil, with the center movable plate grounded. Using the circuit as described, the diode voltage was measured by a sensitive voltmeter and a curve plotted of the diode voltage versus the deflection of the center condenser plate.

This curve is shown in Figure 7. It should be noted that throughout most of its length this curve is practically linear. The total harmonic content represented by the curvature between the two dotted lines, A and B, amounts to less than one per cent. The curvature above that is caused by the nonlinearity of the discriminator resonance slope. When the Q of the discriminator coil is in the order of 200 (which is easy to attain at frequencies of 40 megacycles), and the mechanical vibration applied to the tip of the wire is ± 1 mil, the output from the diode

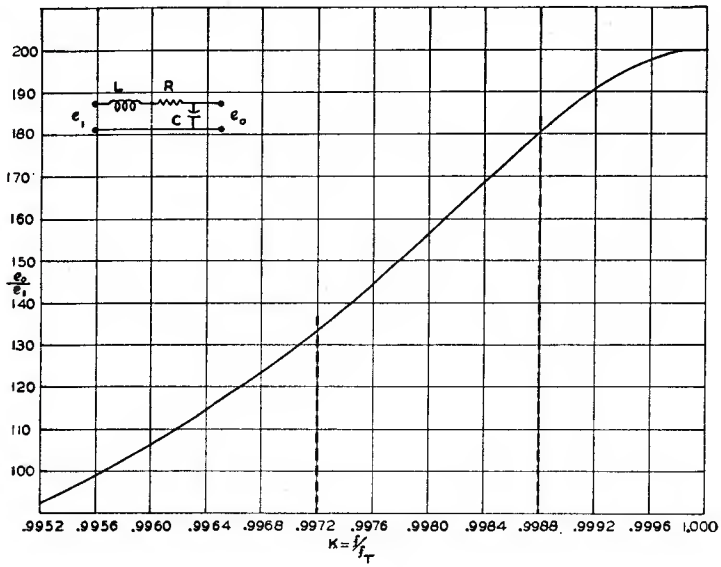


Fig. 8—One side of the discriminator resonance curve.

rectifier, as measured across the diode load resistance R_1, will be in the order of —30 db (reference 0.006 w).

The only nonlinear factor in this circuit is the discriminator resonance peak, one side of which is used as the slope against which the frequency modulation is changed to amplitude variations of the carrier. The calculated linearity of that slope is represented in Figure 8, and is based on the formula[3]

$$\frac{e_0}{e_1} = [(1 - K^2)^2 + 2.5 \times 10^{-5} K^2]^{-1/2}$$

as taken from an equivalent circuit shown in Figure 8 where L and C are the inductance and capacity of the resonant circuit and R is the

loss and therefore the limiting Q factor. K stands for the ratio of the oscillator frequency to the resonant frequency of the discriminator ($K = f/f_r$). When the distortion is thus computed, the following table represents the per cent of second harmonic distortion as compared to the frequency swing, relative gain, and the mean frequency of the point on the resonance curve against which the carrier is sweeping.

K	Second Harmonic in Per Cent of Fundamental, ±45 KC	Relative Gain
0.980	2.94	25.3
0.982	3.26	28.0
0.984	3.67	31.5
0.986	4.20	36.0
0.988	4.90	41.9
0.990	5.88	48.8
0.992	6.9	60.0
0.994	8.0	77.5
0.9968	6.5	123.0
0.9976	3.2	144.5
0.9980	Less than 0.5	156.0
0.9981	Less than 0.5	160.0
0.9988	12.5	180.0

The curve is practically straight from $K = 0.9972$ to 0.9988. Thus a discriminator will cause negligible distortion if the total range of modulation is restricted to 80 kc for a 40-megacycle carrier. This is also the portion of the characteristic which provides maximum output. The range actually used covered 30 kc in each push-pull component, thereby effectively covering 60 kc on the discriminator curve.

The push-pull FM circuit presents a new tool which can be applied to any mechanical translating or measuring device. As an example, in translating the mechanical vibration of a reed it is possible to place the two outside plates of the push-pull condenser on each side of the reed, having the reed act as the center grounded plate. In that case large amplitudes can be translated with minimum distortion without adding anything to the reed itself.

3—Push-Pull FM Cutter Calibrator

During recent years as the art of disk recording has become refined, a number of improvements were made in the recording head. Several ways were devised to measure the ability of the cutting head to repro-

duce faithfully the sounds engraved on a disk.[4] The most important of these measurements is the frequency response of the head after compensation. This type of measurement is made in several ways. One of the best-known and widely used direct methods is by the light

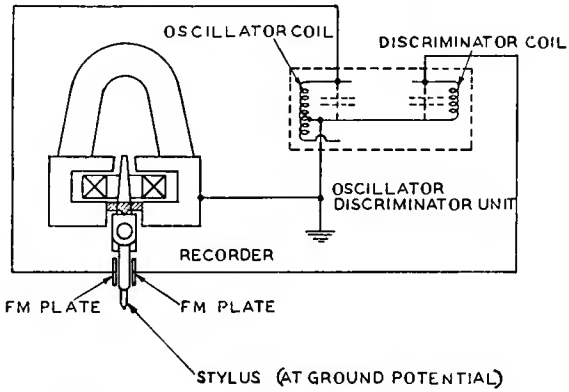


Fig. 9—Arrangement of the FM calibrator adapted to a recorder head.

pattern. As is generally known, this method consists of recording several frequencies on a disk and comparing their amplitudes with the aid of reflection of light. This method is fairly accurate when certain precautions are taken and is really a true indication of the performance

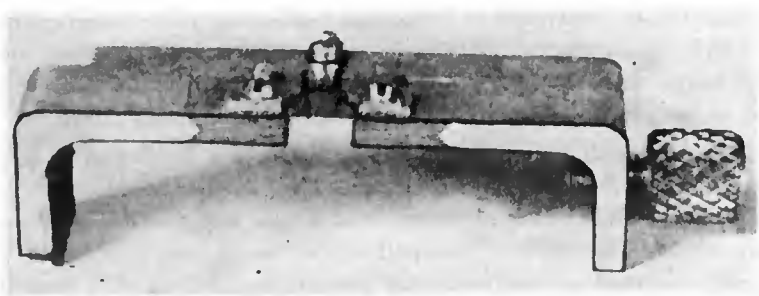


Fig. 10—Push-pull FM plate assembly.

of the head, since the head is actually engaged in cutting a disk that represents the normal load on the stylus.

Another method which has been widely used consists in measuring the amplitude of the stylus vibrating freely in air by means of a microscope which has been precalibrated to measure small distances. In this

method it is very difficult, however, to obtain accurate results, since in some cases the amplitude of the vibrations is extremely small. By substituting a photoelectric cell for the human eye it is possible to improve the accuracy to a large extent. Such optical cutter calibrators have been in use for years. Since the measurements are taken in air, it is assumed that the difference in the measurements is small when compared to such made while the head is actually cutting a record.

The problem of being able to calibrate the recorder under actual cutting conditions still remained, however, and it was not until the push-pull FM circuit was developed that a solution to the problem was

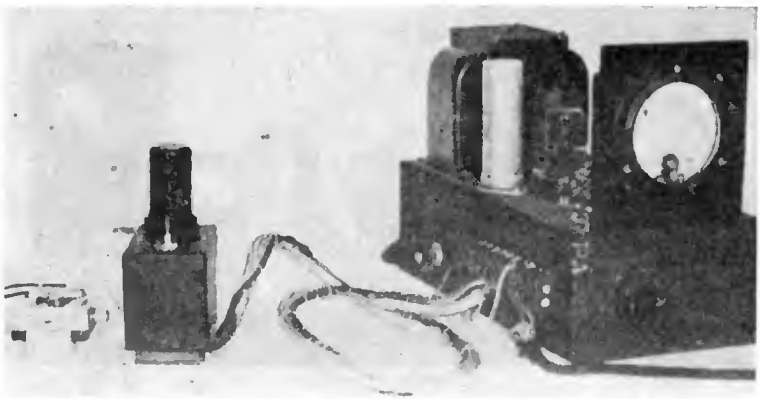


Fig. 11—Component parts of a complete FM calibrator for recording heads.

realized. A device was needed which could be attached to the recorder without requiring much space or adding any mass to the moving system, as well as one which would not couple electrically to the driving coil of the recording head, and of course one which would not interfere with the cutting action of the stylus.

The push-pull FM system meets these requirements very nicely. However, it is practically essential to use the push-pull circuit, because the distortion would be too high if only one side of the stylus were used as one of the plates of a condenser, owing to the low ratio of the capacitor plate spacing and the large variation of the spacing. As illustrated in Figure 9, two small plates, one on each side of the stylus shank, insulated from each other and from the recorder, are spaced a few thousandths of an inch from the stylus. Since neither mass nor rigidity is added to the moving system, no change in its action can

occur. Two flexible leads from these plates and another lead from the cutter mechanism are connected to the oscillator discriminator unit mounted on the carriage and located as close to the cutter head as feasible. The oscillator-discriminator unit is of the same type as described in Section 1. The variation of the capacitance, between the two plates and the stylus, because of its motion, changes the oscillator frequency and varies the tuning of the discriminator in accordance with the method of operation described in Section 2. The audio output of the oscillator-discriminator unit, since it also contains a diode, is transferred to another unit containing an amplifier and a power supply. The output from this amplifier is then either measured with the aid of

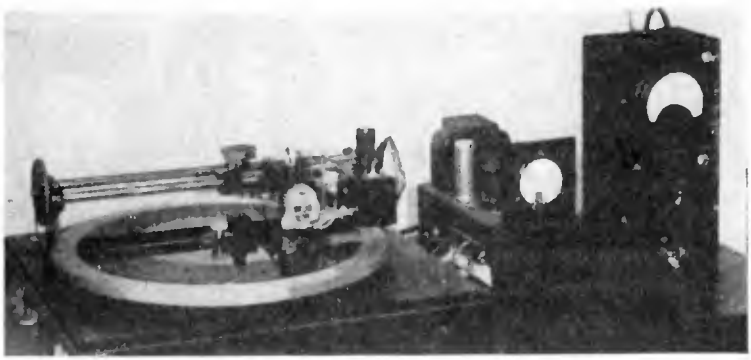


Fig. 12—The FM calibrator in operation.

a tube voltmeter, or is further amplified for listening or other purposes by using a suitable transformer located between the amplifier tube, which has a high-impedance output, and a conventional amplifier of 500 ohms input impedance.

The push-pull FM plates are mounted on the bottom side of the cutter in such a way that in combination with the cutting stylus they form a small push-pull condenser. As shown in Figure 10, this assembled bakelite plate on which the two FM plates are mounted fits on the under side of the cutter and replaces the cover plate. This assembly is again shown separately in Figure 11. The FM plates are, of course, adjustable so as to provide any width of gap to accommodate any cutting styli or for unusual circumstances where a wide gap is necessary. Normally, however, the FM plates are spaced 0.01 in. away from the stylus for frequency response measurements and 0.015 in. for distortion measurements at the lower frequencies. This spacing is ample for any modulation and yet provides more than ample output.

For calibration during the cutting of a record, the cutter is mounted in the usual way. The oscillator-discriminator unit is mounted immediately beside the cutter (see Figure 12). The FM plates are connected to the oscillator and discriminator circuits by means of two 0.004-in. diameter steel wires covered with vinylite insulation. The ground wire between the two units is of the same material. The complete schematic diagram of the calibrator is shown in Figure 13.

The output of the oscillator-discriminator unit is fed to an amplifier, which is on the same chassis with the power supply. It consists of a conventional circuit using the 6SJ7 triode connected. From there the

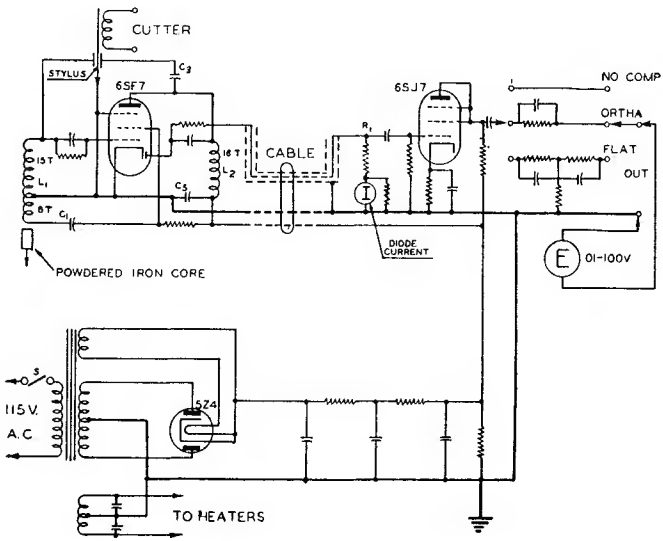


Fig. 13—Schematic of the push-pull FM calibrator.

signal passes through a 3-position switch which selects the "flat compensator," the "orthacoustic," or the "no compensation" circuits. These positions can be used for different purposes as follows: "No compensation" is used for measurements with a vacuum tube voltmeter, for frequency response tests, and input versus output curves. "Flat" response can be used for monitoring on disks requiring that type of response; however, the important function of this position is for intermodulation tests. "Orthacoustic" can also be used for monitoring while the record is being cut, and is designed for that purpose.

To obtain consistent readings, it is important to have the oscillator tuned to the same frequency relative to the optimum point of operation

of the discriminator circuit. To accomplish this, a meter is provided in the diode circuit. The oscillator circuit is tuned for maximum diode current and is then backed down, in the same direction each time, to 70 per cent of the peak current.

The use of this instrument and results of measurements are completely covered in the paper by H. E. Roys entitled "Experience with an FM Calibrator for Disk Recording Heads."[5]

Acknowledgment

The author wishes to acknowledge the helpful suggestions and advice of E. W. Kellogg, H. E. Roys, C. M. Sinnett, and A. C. Blaney in this work.

References

[1] G. L. Beers and C. M. Sinnett, "Some Recent Developments in Record Reproducing Systems," *Jour. Soc. Mot. Pic. Eng.*, XL, 4 (April, 1943), p. 222.

[2] F. E. Terman, "Radio Engineering," McGraw-Hill (New York), 1937, p. 360; J. B. Dow, "A Recent Development in Vacuum Tube Oscillator Circuits," *Proc. I.R.E.*, 19 (December, 1931), p. 2095.

[3] V. D. Landon and K. McIlwain, "Pender's and McIlwain's Electrical Engineers' Handbook (Electric Communication and Electronics)," John Wiley and Sons (New York), 3d Ed., p. 7.

[4] D. G. Fink, "Electronic's Engineering Manual," McGraw-Hill (New York), p. 240.

[5] H. E. Roys, "Experience with an FM Calibrator for Disk Recording Heads," *Jour. Soc. Mot. Pic. Eng.*, 44, 6 (June, 1945), p. 461.

FREQUENCY MODULATION AND CONTROL BY ELECTRON BEAMS*†

By

Lloyd P. Smith‡ and Carl I. Shulman

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*General formulas for the effect of electron beams on resonant systems in terms of frequency shift and change in Q are derived from the point of view of lumped circuits and from a general electromagnetic field standpoint. A way of introducing the electrons has been found which materially enhances their effectiveness in producing a shift of frequency in a resonant system. Measurements of the frequency shift produced by such an electron beam in a typical geometry were made which check well with values calculated from the general theory. Possible amplitude and phase distortions are calculated when such a beam is used to frequency-modulate a system, and these are found to be negligibly small, even for very high modulating frequencies.*

It is shown that this method of frequency control is ideally suited for frequency modulation or automatic frequency stabilization of continuous-wave magnetron oscillators using negative-grid-controlled electron beams for controlling the frequency of oscillations.

I. Introduction

MANY occasions arise when it is desirable to control the frequency of a high-frequency oscillating system so rapidly that only electronic means can be used. The theory of reactive electron beams and the computation of the change in frequency and loading they produce in an oscillating system is presented here from two points of view; the first of which is a treatment essentially based on lumped-constant concepts, and the second is a general field-theory approach especially applicable to cavity systems or systems with distributed constants. The change of frequency and loading can be finally expressed in terms of rather simple formulas.

A part of the problem of obtaining the largest frequency shift possible for a given number of electrons entering the oscillating electromagnetic field region per second is that of introducing the electrons into that combination of oscillating electromagnetic field and steady electric and magnetic fields for which the largest reactive current is produced. For many applications it is necessary to obtain this maximum reactive current consistent with the auxiliary condition that

* Decimal Classification: R138 × R355.912.1.
† Reprinted from *Proc. I.R.E.*, July, 1947.
‡ Now Research Consultant, RCA Laboratories Division.

there be no resistive component. A complete answer to this problem is not attempted here, but a particular combination of fields has been found which yields a much larger reactive current than many other usual field combinations. It also has the advantage that the loading can be made negligibly small.

II. Lumped-Circuit Treatment

When free electrons are introduced into a rapidly oscillating electromagnetic field, there is in general an exchange of energy between the electrons and the field. The electrons oscillate and induce oscillating image currents on the surfaces of the field boundaries. These surface currents are proportional to the field amplitude and in general are not in phase with the oscillating field. Since these currents are proportional to the field, a linear admittance function relating the currents to the field can be written. This electronic admittance can be handled as a conventional circuit element in conjunction with a proper equivalent circuit to represent the over-all system. The real part of the electronic admittance represents an energy transfer between the oscillating electrons and the electromagnetic field, and the imaginary part represents a change in the resonant frequency of the system.

The first section of the paper treats the problems of estimating the change in admittance across a pair of parallel-plane electrodes due to the introduction of an electron beam moving between the parallel planes and in the direction of a constant magnetic field. From this change in electronic admittance one can calculate the change in the resonant frequency of the oscillating system coupled to the parallel-plane electrodes.

Calculation of Electronic Admittance

L. Malter[1] has studied certain phases of the problem, but his results are not readily applicable to the problem considered here.

Consider a pair of parallel plates a distance d apart, infinitely long in the y direction, and of length L in the z direction, as indicated in Figure 1. In the z direction is a constant magnetic field H. Let a rectangular electron beam enter the region between the plates at $z = x = 0$, with velocity v_0 in the z direction. Between the plates is an oscillating electric field E_x in the x direction which is constant over the length and width of the plates. (Assume no fringe field. This

[1] L. Malter, "Deflection and Impedance of Electron Beams at High Frequencies in the Presence of a Magnetic Field," *RCA REVIEW*, vol. 5, pp. 439-454; April, 1941.

point is considered in the Appendix.) The equations of motion for the electrons in the space between the plates are

$$m\ddot{x} = -E_x \mid e \mid - H \mid e \mid \dot{y} \tag{1}$$

$$m\ddot{y} = +H \mid e \mid \dot{x} \tag{2}$$

$$m\ddot{z} = 0 \tag{3}$$

where $\mid e \mid$ is the absolute value of the electronic charge.

For a field $E_x = E_0 e^{i\omega t}$, the solution of these equations for the x component of velocity v_x, subject to the initial conditions at $t = t_0$

$$x = z = 0, \qquad \dot{x} = v_x = \dot{y} = 0, \qquad \dot{z} = v_0, \tag{4}$$

is

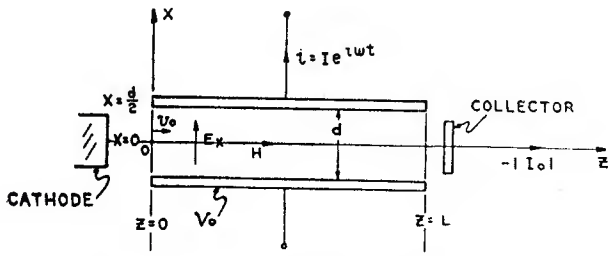


Fig. 1—Schematic diagram of beam and electrode arrangement.

$$v_x = -E_0 \frac{\mid e \mid}{m} \frac{i\omega}{\omega^2 - \omega_c^2} \left\{ \left(\frac{\omega_c + \omega}{2\omega_c} \right) e^{i(\omega_c - \omega)(t - t_0)} + \left(\frac{\omega_c - \omega}{2\omega_c} \right) e^{-i(\omega + \omega_c)(t - t_0)} - 1 \right\} e^{i\omega t} \tag{5}$$

where t_0 is the time at which an electron enters the electric field and $\omega_c = He/m$. If we restrict our interest to the region where $\omega \simeq \omega_c$ or $(|\omega_c - \omega| / \omega) << 1$, the second term in (5), in addition to being rapidly oscillating with respect to the first term, becomes very small. In almost all cases of interest the second term can be neglected, in which case (5) becomes

$$v_x = \frac{-E_0}{2} \frac{\mid e \mid}{m} \frac{i}{\omega - \omega_c} [e^{i(\omega_c - \omega)(t - t_0)} - 1] e^{i\omega t}. \tag{6}$$

The current to the parallel-plane electrodes due to an oscillating charge dq between them is

$$dIe^{i\omega t} = \frac{dq}{d} v_x = \frac{-|I_0|}{v_0} \frac{v_x}{d} dz \tag{7}$$

which on substitution of v_x becomes

$$dI = E_0 \frac{|e|\,|I_0|}{2m} \frac{i}{v_0 d} \frac{}{\omega - \omega_c} [e^{i(\omega_c - \omega)z/v_0} - 1] dz \tag{8}$$

for

$$t = t_0 + \frac{z}{v_0}.$$

The total current at any instant is the sum of the image currents for each differential element of charge from $z = 0$ to $z = L$, so that

$$I = \frac{E_0}{2d} \frac{|e|}{m} |I_0| \frac{i}{v_0} \frac{}{\omega - \omega_c} \int_v^L (e^{i(\omega_c - \omega)z/v_0} - 1)\, dz.$$

On evaluating the integral, one obtains

$$I = E_0 \frac{|e|\,|I_0|}{m} \frac{}{2d} \tau^2 \left\{ \frac{1 - \cos\theta}{\theta^2} + i \frac{\theta - \sin\theta}{\theta^2} \right\} \tag{9}$$

where

$$\tau = \frac{L}{v_0} \text{ and } \theta = (\omega_c - \omega)\, \tau.$$

Letting

$$E_0 d = V_{ac} = \frac{I}{Y_e},$$

then

$$Y_e = \frac{L^2}{4d^2} \frac{I_0}{V_0} \left\{ \frac{1 - \cos\theta}{\theta^2} + i \frac{\theta - \sin\theta}{\theta^2} \right\} = G_e + iB_e \tag{10}$$

where Y_e is an electron admittance and

$$V_0 = \frac{1}{2} \frac{m}{|e|} v_0^2.$$

Y_e versus θ is shown in Figure 2.

It is apparent from (10) that the electronic admittance can be made purely reactive when θ is 2π or -2π. This can be clearly understood physically by considering an electron's contribution to the total current at various points of its trajectory between the parallel plates. Mathematically, of course, this contribution is proportional to the current

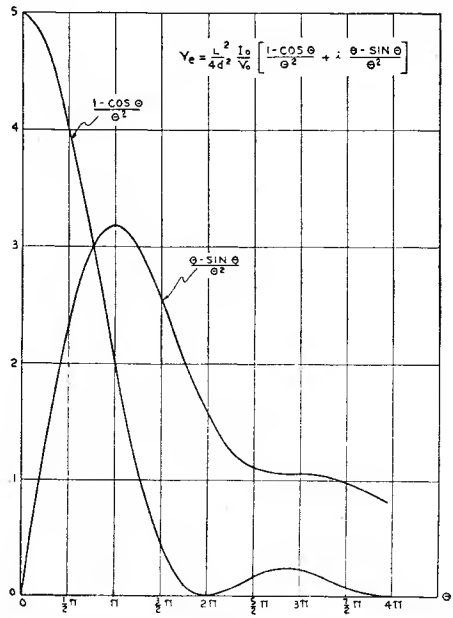


Fig. 2—Electronic admittance as a function of θ.

differential given by (8). The physical situation is as follows: When an electron has a velocity in a direction at right angles to a uniform magnetic field H, it will describe a circular path. The time T required to traverse a complete circle is $T = 2\pi m/H|e|$ where m is the mass of the electron, $|e|$ is its charge, and H is the magnetic field intensity. This period of rotation is independent of the electron's velocity or the radius of its circular path. The angular frequency of rotation is then

$$\omega_c = \frac{2\pi}{T} = \frac{H|e|}{m}.$$

Consider a pair of parallel plates across which an alternating potential difference is applied, resulting in an electric field of amplitude E and angular frequency ω between the capacitor plates shown in Figure 3(a). The uniform magnetic field is in the direction shown. An electron is projected into the field between the plates along the line ab parallel to the magnetic field. Its velocity in this direction will remain constant so that at time $t = 0$ the electron will be at position 1, and at time $\triangle t$ seconds later it will be at position 2, etc. At position 1 the electron has no velocity in the direction of E, but on encountering the electric field E between positions 1 and 2, it will be accelerated in this instantaneous direction. If the frequency of the oscillating electric field ω were equal to the frequency of ω_c, the electron would remain in phase with the electric field and would be continuously accelerated so that it would receive more and more energy from the oscillating field and its motion would be a spiral of ever-increasing radius. This condition would produce an electron current in phase with E at all positions 1, 2, · · · and would constitute a loading or resistive component of current. If the capacitor plates were to form a part of a resonant circuit, this "in-phase component" would have no effect on the frequency but would reduce the Q of the circuit.

The transit time of the electron through the capacitor plates and the relative values of ω and ω_c can be chosen so that the net electron current is not in the direction of E but in quadrature with E, so that it constitutes a purely reactive current. To see how this comes about, let ω_c be somewhat less than ω. In this circumstance the electron will not complete one revolution during one complete cycle of the oscillating field. Consequently it will lag behind E by a certain phase angle $\triangle\phi$, so that as times goes on the electron lags more and more behind E. It will, however, continue to be accelerated until it lags behind E by 90 degrees. The electron then still increases its lagging phase angle by $\triangle\phi$ for each revolution, but the electron now is retarded by the oscillating field and loses energy. It is evident that, for a given value of $\omega - \omega_c$, the time of flight of the electron between the capacitor plates can be adjusted so that the energy acquired in the time interval during which acceleration took place is given back to the field during the interval of retardation, and the net energy transfer from field to electron is zero. This situation is illustrated in Figures 3(a), (b) and (c). When an electron enters the plates at position 1 in Figure 3(a), it has no displacement from the median plane between the plates. The displacement for various positions is shown in Figure 3(b). Also the transverse current at position 1 is zero, as is shown in Figure 3(c). At position 2, Figure 3(a), the electron has acquired some

circular motion and its amplitude of oscillation about the median plane is shown by the vector r_2 in Figure 3(b). Since $\omega_e < \omega$, the electron lags behind E by the phase angle ϕ when it reaches position 2. The transverse electron current in magnitude and phase relative to E is shown by the vector i_2 in Figure 3(c). When the electron has reached position 3, its amplitude of oscillation is given by r_3 and the transverse current by i_3. Because of the electron's increased velocity, it is evident that at position 3 the magnitude of the current i_3 is greater, as also is the lagging phase angle. At position 5 the electron has acquired its maximum displacement and the current due to it is 90 degrees out of phase with E and is purely reactive. From this point on, the electron is retarded and the current produced by it has a

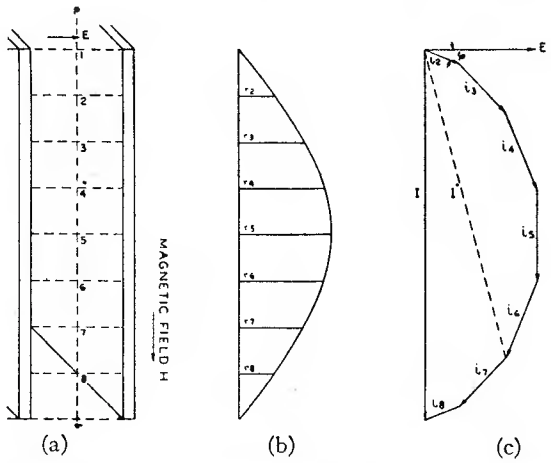


Fig. 3—Vector representation of electron current between parallel plates.

negative resistive component (a component opposite to E). When electrons enter the capacitor plates at all times, as in a continuous beam, there will be electrons at all positions from 1 to 9 simultaneously, and the total effective current will be the vector sum of the current at each position. This vector sum is shown by I in Figure 3(c). Thus the total effective current lags behind the field E by 90 degrees. Thus E and I are related in phase exactly the same as the potential difference across an inductance and the current through it. If the capacitor plates as shown were in parallel with an inductance to form a resonant circuit with resonant frequency ω, the passage of the beam of electrons through the capacitor plates would increase the resonant frequency.

On the other hand, if $\omega_e > \omega$, then the currents i_2, i_3, etc., would lead

E and the total current would lead E by 90 degrees, giving rise to a decrease in the resonant frequency of the parallel circuit.

If the electrons had emerged from the plates at position 6, then the total current would be I' as shown by the dashed vector in Figure 3(c). In this case, the total current will have a resistive and reactive component as indicated by (9).

Calculation of Frequency Change

If a beam of electrons is to be used in the manner described above to control the resonant frequency of an oscillating system, the frequency of the over-all system can be estimated by applying the Y_e calculated above to the equivalent circuit for the system, and calculating the frequency from conventional circuit theory. Let the L_0C circuit shown in Figure 4 represent the equivalent circuit of the system before the beam is introduced. C_0 is defined as the total equivalent capacitance of the system, or more exactly,

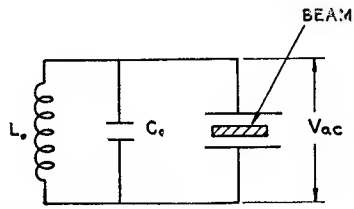


Fig. 4—Equivalent lumped circuit of resonant system.

$$C_0 V_{ac}^2 = \mathcal{E} \int_v E^2 dv$$

where $\mathcal{E}$ is the dielectric constant of the space, V_{ac} is the radio-frequency voltage across the parallel planes between which the beam moves, and the integral is taken over the entire volume of the oscillating system. C_0 is then determined by the stored energy of the system and the point at which the beam is introduced, for V_{ac} depends, in general, on the geometry of the system.

To calculate the resonant frequency, write

$$B_e + B_c = 0 \tag{11}$$

where

B_e = electronic susceptance

and $$B_c = \text{circuit susceptance} = \frac{C_0}{\omega}(\omega^2 - \omega_0^2) \tag{12}$$

where $\omega_0^2 = 1/L_0C_0 =$ resonant frequency of the system before the electron beam is introduced. Let $\omega = \omega_0 + \Delta\omega$ where $\Delta\omega$ is the small change in frequency due to the introduction of the beam. If

$$\Delta\omega << \omega_0 \text{ so that } \frac{\omega_0 + \omega}{\omega} \simeq 2, \text{ and } B_c \simeq 2C_0\Delta\omega,$$

(11) becomes

$$2C_0\Delta\omega + \frac{L^2}{4d^2}\frac{|I_0|}{V_0}\frac{(\omega_c - \omega_0 - \Delta\omega)\,\tau - \sin(\omega_c - \omega_0 - \Delta\omega)\,\tau}{(\omega_c - \omega_0 - \Delta\omega)^2\,\tau^2} = 0. \tag{13}$$

For most cases of interest we may take $\Delta\omega << (\omega_c - \omega_0)$. Then $(\omega_c - \omega)\,\tau = (\omega_c - \omega_0)\,\tau = \theta$, so that

$$\Delta\omega(\theta) = \frac{L^2}{8d^2}\frac{|I_0|}{V_c}\frac{1}{C_0}\frac{\sin\theta - \theta}{\theta^2}. \tag{14}$$

The electronic conductance introduced by the beam is

$$G_e(\theta) = \frac{L^2}{4d^2}\frac{|I_0|}{V_0}\frac{1 - \cos\theta}{\theta^2}. \tag{15}$$

Figure 5(a) shows $\Delta\omega$ and G_e versus ω_c/ω_0. Note that the conductance is a maximum at $\omega_c = \omega$. This means that the electrons rotate in phase with the electric field, absorbing energy continuously. There are no reactive currents, so that $\Delta\omega = 0$. At $(\omega_c - \omega_0)\,\tau = 2\pi$, $G_e = 0$. The electrons gain and lose energy during transit and come out with no net exchange, as inferred from Figures 3(a), (b), and (c). However, there are reactive currents and the frequency does change. For a given value of τ, $\Delta\omega(\theta)$ is a maximum at $(\omega_c - \omega_0)\,\tau = \pm\pi$.

$$\Delta\omega(\pm\pi) = \pm\frac{L^2}{8d^2}\frac{|I_0|}{V_0}\frac{1}{\pi C_0} \tag{14a}$$

$$G_e(\pm\pi) = \frac{L^2}{2d^2}\frac{|I_0|}{V_0}\frac{1}{\pi^2}, \tag{15a}$$

while

$$\triangle\omega\ (\pm 2\pi) = \pm \frac{L^2}{8d^2} \frac{|I_0|}{V_0} \frac{1}{2\pi C_0} = \frac{1}{2} \triangle\omega\ (\pm \pi) \tag{14b}$$

$$G_e\ (\pm 2\pi) = 0. \tag{15b}$$

It is clear from the above that it is possible to frequency-modulate an oscillating system by introducing into the resonant circuit an electron beam moving in the direction of a constant magnetic field. Furthermore, this is possible without absorption of power by the electron beam.

If, instead of being collected after traversing the length of the parallel planes, the electrons are reflected and caused to make a return

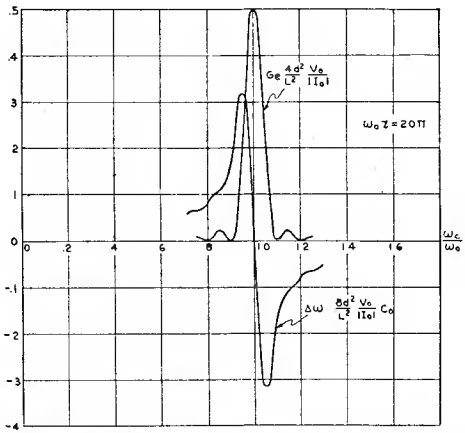


Fig. 5(a)—Theoretical frequency shift and electronic loading as a function of magnetic field strength.

trip through the radio-frequency field space, the effective transit time may be increased. If the turn-around time is zero, (14) and (15) may be used to calculate the resultant electronic admittance where τ is the total transit time. For finite turn-around time, the electronic admittance Y_e is a fairly complex function of the transit angle in the end space, except in the case for $\theta = 2\pi$, where Y_e is independent of the turn-around time. This follows because at $\theta = 2\pi$ the electrons emerge from the active space with zero transverse velocity and, of course, re-enter with zero transverse velocity, in which case Y_e is independent of the entrance time.

Some Experimental Results

An experimental check of (14) was carried out at 4000 megacycles

using a twelve-vane magnetron oscillator shown schematically in Figures 6 and 7, which show a multivane magnetron with two rectangular beams introduced between the vane structures. On one side of the resonant cavity block is the beam gun structure, which includes a cathode, control grid, and acceleration aperture or grid; on the other side of the block is a collector (or reflector for multiple transit). The beam is focused by the magnetic field. Measured values of the change in frequency per milliampere of beam current are shown in Figure 5(b) plotted as a function of ω_c/ω_0 for two different transit times. These curves have the same form as the theoretical curves shown in Figure 5(a). The L/d ratio for the cavity was 3, and the effective capacitance C_0 was estimated to be about 2×10^{-12} farads. At a beam

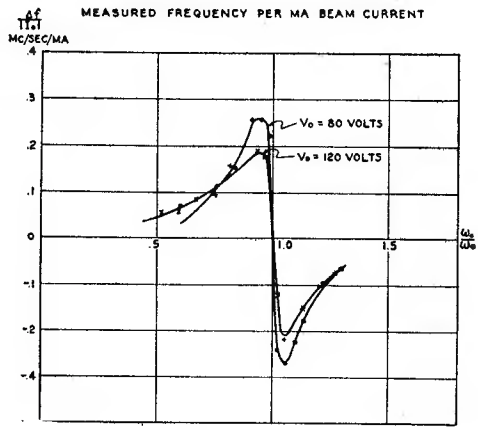


Fig. 5(b)—Measured frequency shift as a function of magnetic field strength.

voltage of 80 volts, (14) gives for this cavity a maximum frequency change of 0.36 megacycle per milliampere of beam current. The measured value at 80 volts was about 0.25 megacycle per milliampere. Since the estimate of C_0 was rather rough, the check is considered good. Magnetrons of the type shown in Figures 6 and 7 were built and successfully modulated at these laboratories under U. S. Navy Contract No. NXsa-35042 during the past two years.

Design Considerations

It would appear from the foregoing analysis that the effect of the electron beam is independent of the electric field intensity. This is indeed true if none of the electrons strike the electrodes. It is clear from (10) that $\triangle\omega$ increases with τ^2/d^2. Integration of (6) shows that the maximum amplitude of oscillation about the median plane increases with τ. Hence, $\triangle\omega$ increases with $x_{max}{}^2/d^2$ and is obviously a maximum

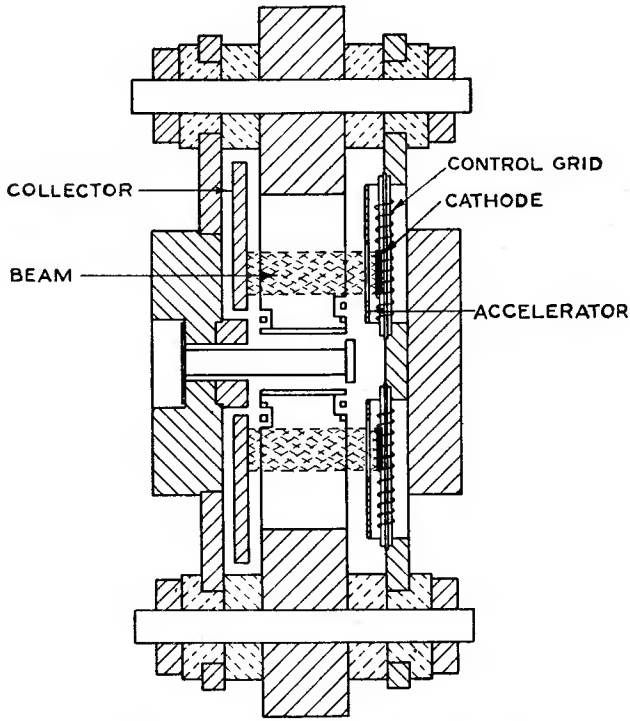


Fig. 6—Schematic view of multivane magnetron with auxiliary beams.

at grazing incidence. In order to obtain the maximum frequency shift consistent with electron orbits just grazing the plates, it is necessary to look at the displacement of the spiraling electron which does depend on the electric field intensity.

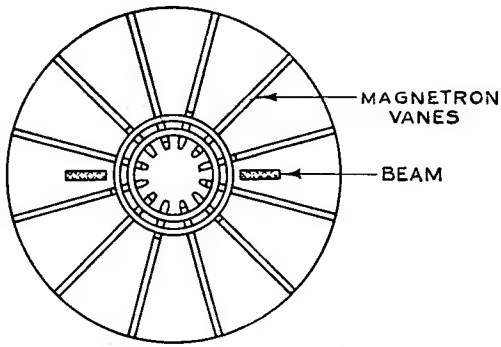


Fig. 7—Cross-sectional view of multivane magnetron with auxiliary beams.

Integration of (6) gives

$$x = -E_0 \frac{|e|}{m} \frac{e^{i\omega t}}{2(\omega - \omega_c)\omega} [e^{i(\omega_c - \omega)\tau(z/L)} - 1] \tag{16}$$

$$|x|_{\max} = \frac{E_0 |e|}{m} \frac{\tau}{\omega_c \theta}; \qquad (\omega \neq \omega_c). \tag{17}$$

As long as $|x|_{\max} \leqq (d/2)$, $\Delta\omega$ and G_e are independent of the field intensity.

It is of interest to estimate, in a rough way, the largest possible frequency change consistent with space-charge limitations and grazing incidence that can be effected with the introduction of a resonant beam of the type herein described. Consider a beam of thickness t and width W moving between a pair of parallel-plane electrodes also of width W, of length L, and a distance d apart. We shall assume that all the stored energy in the system is contained between the parallel-plane electrodes and select a case of particular interest, namely $(\omega_c - \omega_0)\tau = 2\pi$ (no loading). From (14b),

$$\Delta\omega = \frac{L^2}{8d^2} \frac{|I_0|}{V_0} \frac{1}{2\pi C_0}. \tag{18}$$

The maximum current that can be established between parallel planes is[2]

$$|I_0|_{\max} = \frac{4}{9\pi} \sqrt{\frac{2e}{m}} \frac{W}{d} V_0^{3/2} F \tag{19}$$

where F is a function of t/d, the ratio of beam thickness to plate separation, varying between 2 at $t/d = 1$ and 0.87 at $t/d = 0$, and W is the beam width. Since all the stored energy is assumed to be between the plane electrodes, the effective capacitance C_0 is

$$C_0 = \frac{1}{4\pi} \frac{WL}{d}. \tag{20}$$

The $\Delta\omega$, where I_0 has been maximized, would then be (combining (18), (19), and (20))

[2] A. V. Haeff, "Space-Charge Effects in Electron Beams," *Proc. I.R.E.*, vol. 27, pp. 586-602; September, 1939.

$$\Delta\omega = \frac{F}{9\pi d^2}\frac{L^2}{\tau}. \tag{21}$$

It is clear from (21) that $\Delta\omega$ may be increased further by reducing d. However, d may not be reduced beyond the point where electrons are caught by the electrodes. The minimum d is determined by the requirement that the maximum orbit diameter be just equal to $d/2$ for a beam of negligible thickness, or

$$|x|_{\max} = E_0 \frac{|e|}{m}\frac{\tau}{\omega_c 2\pi} = \frac{d}{2}.$$

Recalling that $V_{ac} = E_0 d$, we have

$$\frac{d^2}{2} = \frac{V_{ac}}{2\pi\omega_c}\frac{|e|}{m}\tau. \tag{22}$$

Combining (21) and (22),

$$\left(\frac{\Delta\omega}{\omega_c}\right)_{\max} = \frac{2F}{9}\frac{V_0}{V_{ac}}.$$

Since $\omega_c \simeq \omega_0$, write

$$\left(\frac{\Delta\omega}{\omega_0}\right)_{\max} = \frac{2F}{9}\frac{V_0}{V_{ac}}. \tag{23}$$

Equation (23) gives the maximum fractional frequency change for a given beam voltage and radio-frequency voltage across the electrodes.

If the stored energy is not entirely contained between the parallel planes, then

$$\left(\frac{\Delta\omega}{\omega_0}\right)_{\max} = \frac{2F}{9}\frac{V_0}{V_{ac}}\frac{C}{C_0} \tag{24}$$

where C_0 is the total effective capacitance of the system, and C is the capacitance between the parallel planes.

It should be pointed out that (24) may be expected to furnish somewhat too low a value for $\Delta\omega$ because all electrons were presumed to have space-charge-free transit times.

The limitations described in this section apply also to multiple-transit conditions, for the total current that can be maintained in the

vane space is the sum of the magnitudes of the current away from and toward the cathode.

Effect of Modulation Rate on Electronic Admittance

If a resonant beam of the type herein described is to be used to modulate or control oscillating systems, it is important to know how the electronic admittance Y_e depends on the modulation rate. Assume, for instance, that the beam current is not constant but is of the form

$$| I_0 | (1 + p \cos \omega_m t_0) \tag{25}$$

where ω_m is the modulation frequency and p is the modulation coefficient. Then (8) becomes

$$dI_m = iE_0 \frac{| e |}{m} \frac{| I_0 |}{2(\omega - \omega_c) v_0 d} (1 + p \cos \omega_m t_0) \left[e^{i(\omega_c - \omega) z/v_0} - 1 \right] dz. \tag{26}$$

Let $I_m = I_1 + I_2$ where I_1 is the solution with no modulation given by (9), and I_2 is the contribution from the modulation terms.

$$I_2 = i \frac{E_0}{d} \frac{| e |}{m} p \frac{| I_0 |}{2 (\omega - \omega_c) v_0} \int_0^L \cos \omega_m t_0 \left[e^{i(\omega_c - \omega) z/v_0} - 1 \right] dz. \tag{27}$$

Let $\omega_m \tau = \phi$ and recall that $(\omega_c - \omega) \tau = \theta$ where $\tau = L/v_0$. Integration of (27) gives

$$I_2 = \frac{E_0}{d} \frac{| e |}{m} \frac{|I_0|}{4} \frac{\tau^2}{\theta} p \left\{ e^{i\omega_m t} \left[\frac{1 - e^{i(\theta - \phi)}}{\theta - \phi} + \frac{1 - e^{-i\phi}}{\phi} \right] + e^{i\omega_m t} \left[\frac{1 - e^{i(\theta + \phi)}}{\theta + \phi} - \frac{1 - e^{i\phi}}{\phi} \right] \right\}. \tag{28}$$

Combining I_1 and I_2, one can write the over-all modulated admittance Y_{em} in expanded trigonometric terms as follows:

$$Y_{em} = \frac{L^2}{4d^2} \frac{| I_0 |}{V_0} \left\{ \frac{1 - \cos \theta}{\theta^2} \left[1 + pS_1 (\theta, \phi) \cos (\omega_m t + \beta (\theta, \phi)) \right] + i \frac{\theta - \sin \theta}{\theta^2} \left[1 + pS_2 (\theta, \phi) \cos (\omega_m t + \alpha (\theta, \phi)) \right] \right\} \tag{29}$$

where

$$S_1^2(\theta,\phi)=\frac{\theta^2}{4(1-\cos\theta)^2}\left\{\left[\frac{1-\cos(\theta-\phi)}{\theta-\phi}+\frac{1-\cos(\theta+\phi)}{\theta+\phi}\right]^2+\left[\frac{\sin(\theta-\phi)}{\theta-\phi}-\frac{\sin(\theta+\phi)}{\theta+\phi}\right]^2\right\} \tag{30}$$

$$S_2^2(\theta,\phi)=\frac{\theta^2}{4(\theta-\sin\theta)^2}\left\{\left[2\frac{\sin\phi}{\phi}-\frac{\sin(\theta-\phi)}{\theta-\phi}-\frac{\sin(\theta+\phi)}{\theta+\phi}\right]^2+\left[\frac{1-\cos(\theta-\phi)}{\theta-\phi}-\frac{1-\cos(\theta+\phi)}{\theta+\phi}+2\frac{1-\cos\phi}{\phi}\right]^2\right\} \tag{31}$$

$$\tan\beta(\theta,\phi)=\frac{\dfrac{\sin(\theta+\phi)}{\theta+\phi}-\dfrac{\sin(\theta+\phi)}{\theta-\phi}}{\dfrac{1-\cos(\theta-\phi)}{\theta-\phi}+\dfrac{1-\cos(\theta+\phi)}{\theta+\phi}} \tag{32}$$

$$\tan\alpha(\theta,\phi)=\frac{\dfrac{1-\cos(\theta+\phi)}{\theta+\phi}-\dfrac{1-\cos(\theta-\phi)}{\theta-\phi}-2\dfrac{1-\cos\phi}{\phi}}{2\dfrac{\sin\phi}{\phi}-\dfrac{\sin(\theta-\phi)}{\theta-\phi}-\dfrac{\sin(\theta+\phi)}{\theta+\phi}}. \tag{33}$$

For low modulating frequencies where $\theta>>\phi$, ϕ can be neglected with respect to θ. Then

$$S_1(\theta,\phi)=1 \qquad \theta>>\phi \tag{34}$$

$$\beta(\theta,\phi)=0 \qquad \theta>>\phi \tag{35}$$

$$S_2(\theta,\phi)=\frac{\theta}{\theta-\sin\theta}\sqrt{\left(\frac{\sin\phi}{\phi}-\frac{\sin\theta}{\theta}\right)^2+\left(\frac{1-\cos\phi}{\phi}\right)^2} \tag{36}$$

$$\tan\alpha(\theta,\phi)=\frac{\dfrac{1-\cos\phi}{\phi}}{\dfrac{\sin\phi}{\phi}-\dfrac{\sin\theta}{\theta}}. \tag{37}$$

It is of interest to look at the case where $\theta = 2\pi >> \phi$:

$$S_2(2\pi, \phi) = \frac{\sqrt{2}}{\phi} \sqrt{1 - \cos \phi} \qquad \theta >> \phi \tag{38}$$

and

$$\alpha(2\pi, \phi) = \frac{\phi}{2} \qquad \theta >> \phi \tag{39}$$

$$S_1(2\pi, \phi) = 1 \qquad \theta >> \phi. \tag{40}$$

So that, when $\theta = 2\pi >> \phi$,

$$Y_{em}(2\pi, \phi) = i \frac{L^2}{4d^2} \frac{|I_0|}{V_0} \frac{1}{2\pi} \left[1 + p \frac{\sqrt{2}}{\phi} \sqrt{1 - \cos \phi} \cos \omega_m \left(t + \frac{\tau}{2} \right) \right] \tag{41}$$

and

$$\triangle\omega(2\pi, \phi) = \frac{-L^2}{16\pi d^2} \frac{|I_0|}{V_0} \frac{1}{C_0} \left[1 + p \frac{\sqrt{2}}{\phi} \sqrt{1 - \cos \phi} \cos \omega_m \left(t + \frac{\tau}{2} \right) \right]. \tag{42}$$

Note that, for $\theta >> \phi$, distortion terms appear only in the reactive component of the modulated admittance, and appear in the form of both amplitude and phase distortion. For $\theta = 2\pi >> \phi$, the loading term becomes zero and the phase distortion in the reactive component disappears. Equation (42) is of great practical interest in the design of electronic reactance control devices.

When θ and ϕ are of the same order of magnitude, Y_{em} is given by (29) and is rather involved. Two cases of interest are $\theta = \pi$ and 2π.

$\theta = \pi$

$$S_1(\pi, \phi) = \frac{\pi^2}{2(\pi^2 + \phi^2)} \sqrt{2(1 + \cos \phi)} \tag{43}$$

$$\beta(\pi, \phi) = -\frac{\phi}{2} \tag{44}$$

$$S_2(\pi, \phi) = \frac{\sqrt{2}}{(\pi^2 - \phi^2)\phi} \sqrt{\pi^2(\pi^2 - 2\phi^2)(1 - \cos \phi) + 2\phi^4} \tag{45}$$

$$\alpha(\pi, \phi) = \tan^{-1} \frac{(\pi^2 - 2\phi^2) \cos \phi - \pi^2}{(\pi^2 - 2\phi^2) \sin \phi} \tag{46}$$

so that $Y_{em}(\pi, \phi)$

$$= \frac{L^2}{4d^2} \frac{|I_0|}{V_0} \left\{ \frac{2}{\pi^2} \left[1 + p \frac{\pi^2}{2(\pi^2 - \phi^2)} \sqrt{2(1 + \cos \phi)} \cos \omega_m \left(t - \frac{\tau}{2} \right) \right] \right.$$

$$\left. + i \frac{1}{\pi} \left[1 + \frac{p \sqrt{2} \sqrt{\pi^2(\pi^2 - 2\phi^2)(1 - \cos \phi) + 2\phi^4} \cos (\omega_m t + \alpha)}{(\pi^2 - \phi^2)\phi} \right] \right\}. \tag{47}$$

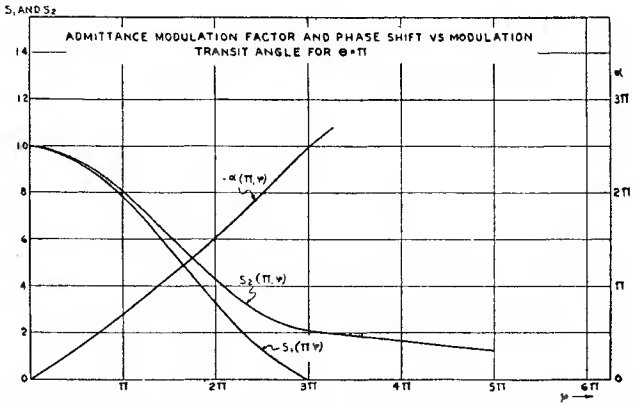


Fig. 8—Admittance modulation factor and phase shift versus modulation transit angle for $\theta = \pi$.

Figure 8 shows S_1, S_2, and α versus ϕ for $\theta = \pi$. It is interesting to note that there is only amplitude distortion in the loading term, while there is both amplitude and phase distortion in the reactance term.

$\theta = 2\pi$: $(\cos \theta = 1)$ Although in this case $S_1(2\pi, \phi)$ becomes infinite,

$$\frac{1 - \cos \theta}{\theta^2} S_1(2\pi, \phi) \text{ remains finite.}$$

$$\frac{(1 - \cos \theta)}{\theta^2} S_1(2\pi, \phi) = \frac{1}{4\pi^2 - \phi^2} \sqrt{2(1 - \cos \phi)} \tag{48}$$

$$\beta(2\pi, \phi) = \frac{\pi}{2} - \frac{\phi}{2} \tag{49}$$

$$S_2(2\pi, \phi) = \frac{4\pi^2}{4\pi^2 - \phi^2} \frac{\sqrt{2(1 - \cos \phi)}}{\phi} \tag{50}$$

$$\alpha(2\pi, \phi) = -\frac{\phi}{2} \tag{51}$$

so that $Y_{cm}(2\pi, \phi)$

$$= \frac{L^2}{4d^2} \frac{|I_0|}{V_0} \left\{ p \frac{1}{4\pi^2 - \phi^2} \sqrt{2(1 - \cos \phi)} \cos \left[\omega_m \left(t - \frac{\tau}{2} \right) + \frac{\pi}{2} \right] \right.$$

$$\left. + i \frac{1}{2\pi} \left[1 + p \frac{4\pi^2}{4\pi^2 - \phi^2} \frac{\sqrt{2(1 - \cos \phi)}}{\phi} \cos \omega_m \left(t - \frac{\tau}{2} \right) \right] \right\}. \tag{52}$$

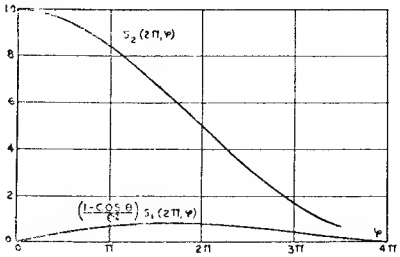


Fig. 9—Admittance modulation factor versus modulation transit angle for $\theta = \pi$.

No phase distortion appears in either the loading or the reactive term. Figure 9 shows $(1 - \cos \theta/\theta^2)$ S_1 and S_2 versus ϕ for $\theta = 2\pi$. There is no average loading, but there is instantaneous loading with a maximum near $\phi = 2\pi$.

$$\Delta\omega_m (2\pi, \phi) = \frac{L^2}{16\pi d^2} \frac{|I_0|}{V_0} \frac{1}{C_0} pG_2 (2\pi, \phi) \cos \omega_m \left(t - \frac{\tau}{2} \right). \tag{53}$$

ϕ may be as large as $3\pi/2$ before S_2 $(\theta = 2\pi)$ is reduced to half its value at $\phi = 0$. For example, if

$$\omega_0 = 2\pi \times 4000 \times 10^6 \text{ and } \frac{\omega_c}{\omega_0} = 1.1$$

$$\phi_m = \frac{3\pi}{2} \text{ corresponds to } \omega_m$$

$$= 2\pi \times 300 \times 10^6 \text{ radians per second.}$$

III. General Field-Theory Treatment

When electrons are injected into cavities or systems with distributed constants, it is convenient to be able to calculate the resulting changes in frequency in terms of the electromagnetic field quantities and the forced electron convection current, rather than in terms of currents induced in a circuit containing lumped constants as was done in Section I. This can be done quite simply with adequate accuracy by means of a perturbation method.

The electromagnetic field in a region of space entirely enclosed by a perfectly conducting surface will be given by a solution of Maxwell's equations for which the tangential component of electric field at the surface vanishes. When the field quantities are expressed in the meter-kilogram-second[3] system of units, these equations are:

$$\text{Curl } H = J + \epsilon \dot{E} \tag{54a}$$

$$\text{Div } H = 0 \tag{54b}$$

$$\text{Curl } E = -\mu \dot{H} \tag{54c}$$

$$\text{Div } E = \rho/\epsilon. \tag{54d}$$

In obtaining solutions of (54), it is convenient to reduce the number of equations by expressing E and H in terms of the scalar potential ϕ and the vector potential A by means of the relations

$$E = -\dot{A} - \text{grad } \phi \tag{55a}$$

$$\mu H = \text{Curl } A \tag{55b}$$

where

$$\dot{A} = \frac{\partial A}{\partial t}.$$

When so expressed, E and H automatically satisfy (54b) and (54c) and we are left with the problem of solving the equations

$$\nabla \cdot \nabla A - \epsilon\mu \ddot{A} = -\mu J + \mu\epsilon \text{ grad } \dot{\phi} + \text{grad div } A$$

$$\nabla^2 \phi + \text{div } A = -\rho/\epsilon.$$

For the interpretation of $\nabla \cdot \nabla A$, see footnote reference 3, page 49. Since A is not uniquely defined by (55b), we are free to impose the

[3] J. D. Stratton, ELECTROMAGNETIC THEORY, p. 16, McGraw-Hill Book Co., New York, N. Y., 1941.

additional condition that $\operatorname{div} A = 0$, whereafter the above equations reduce to

$$\nabla \cdot \nabla A - \epsilon\mu\ddot{A} = -\mu J + \mu\epsilon \operatorname{grad} \dot{\phi} \tag{56a}$$

$$\nabla^2\phi = -\rho/\epsilon \tag{56b}$$

When there is no charge or current in the enclosed space, ϕ may be taken to be identically zero, and only one equation remains, namely,

$$\nabla \cdot \nabla A - \epsilon\mu\ddot{A} = 0. \tag{57}$$

This means that both the electric and magnetic fields are derivable from the vector potential.

The general procedure in computing the change in resonant frequencies of a cavity produced by the injection of a stream of electrons is as follows. Before the electrons are injected, the field in the cavity must satisfy (57), since there is no charge or current in the interior. Solutions of this equation, which are harmonic in the time and which satisfy the boundary conditions already mentioned, exist only for discrete frequencies. Such particular solutions can be written in the form

$$A(r, t) = A_{0n}(r)\, e^{i\omega_{0n}t} \tag{58}$$

where

$$\nabla \cdot \nabla A_{0n} + \mu\epsilon\omega_{0n}^2 A_{0n} = 0. \tag{59}$$

The solutions of the last equation give the special distribution of field for the characteristic modes of oscillation of the cavity and for every geometry of cavity it is, in principle, possible to compute the corresponding resonant frequency ω_{0n}. This will be done later for special cases.

When electrons are injected into the cavity, the electromagnetic field changes, and the resonant frequency may also change. The new field will be given by solutions of (56a) and (56b). The field in the cavity will act on the electrons, and so the current will not remain uniform but will vary with time and position in the cavity. This modified current in turn modifies the field, and so on until a steady state is reached. If the injected current is not too large it is convenient to adopt a method of successive approximations for arriving at the steady-state condition. To this end we shall assume that before electrons are introduced the field in the cavity is given in terms of the

vector potential of the nth mode, i.e., A_{0n}, corresponding to the frequency ω_{0n}. If a uniform beam of electrons is introduced into the cavity, then we may calculate a first approximation to the current inside the cavity by calculating the electron trajectories that would be produced by the unperturbed or zeroth-order field existing before the electrons were introduced. Let the first-order approximation to the current density inside the cavity be denoted by $\delta J_1(r, t)$. δ is a parameter introduced to indicate the order of smallness of the current. The corresponding first approximation to the charge distribution in the cavity can be denoted by $\delta\rho_1(r, t)$. The first-order current and charge can be substituted in (56a) and (56b) and these equations solved to obtain a first approximation for the vector and scalar potentials which may be denoted by A_{1n} and ϕ_{1n}. In principle this procedure can be repeated and would be expected to yield a convergent series (when J_1 is not too large) which would give a solution of the problem with any desired accuracy.

Carrying out this procedure, the first-order approximation for the vector and scalar potentials will be found by solving the following equations:

$$\nabla \cdot \nabla A_{1n}(r, t) - \epsilon\mu \ddot{A}_{1n}(r, t) = -\mu\delta J_{1n}(r, t) + \mu\epsilon \operatorname{grad} \phi_{1n}(r, t) \quad (60)$$

$$\nabla^2 \phi_{1n}(r, t) = \frac{\delta\rho_{1n}(r, t)}{\epsilon}. \quad (61)$$

It is to be understood that J_1 and ρ_1 are to be determined from the unperturbed field distribution $A_{0n}(r)\, e^{i\omega t}$ with ω as yet unknown. When this is done, the first-order current density and charge can be expressed in terms of their harmonic components as follows:

$$\begin{aligned} J_{1n} &= J_{1n}{}^0 + J_{1n}{}^{(1)}(r)\, e^{i\omega t} + J_{1n}{}^{(2)}(r)\, e^{2i\omega t} + \cdots \\ \rho_{1n} &= \rho_{1n}{}^0 + \rho_{1n}{}^{(1)}(r)\, e^{i\omega t} + \rho_{2n}{}^{(2)}(r)\, e^{2i\omega t} + \cdots \end{aligned} \quad (62)$$

where $J_{1n}{}^0$ and $\rho_{1n}{}^0$ are the steady current and charge densities. Solutions of (60) and (61) are facilitated by writing the potentials in terms of harmonic components, so that

$$\begin{aligned} A_{1n} &= A_{1n}{}^0 + A_{1n}{}^{(1)} e^{i\omega t} + A_{1n}{}^{(2)} e^{2i\omega t} + \cdots \\ \phi_{1n} &= \phi_{1n}{}^0 + \phi_{1n}{}^{(1)} e^{i\omega t} + \phi_{1n}{}^{(2)} e^{2i\omega t} + \cdots . \end{aligned} \quad (63)$$

Equations (62) and (63) can now be substituted in (60) and (61) and coefficients of the linearly independent time functions $e^{i\omega t}$, $e^{2i\omega t}$, etc.,

equated. This gives the following infinite set of equations:

$$\nabla \cdot \nabla A_{1n}{}^{0} = -\mu \delta J_{1n}{}^{0} \tag{64a}$$

$$\nabla \cdot \nabla A_{1n}{}^{(1)} + \epsilon\mu\omega^2 A_{1n}{}^{(1)} = -\mu \delta J_{1n}{}^{(1)} + i\mu\epsilon\omega \operatorname{grad} \phi_{1n}{}^{(1)} \tag{64b}$$

$$\nabla^2 \phi_{1n}{}^{0} = \frac{\delta \rho_{1n}{}^{0}}{\epsilon} \tag{65a}$$

$$\nabla^2 \phi_{1n}{}^{(1)} = \frac{\delta \rho_{1n}{}^{(1)}}{\epsilon}. \tag{65b}$$

Equations (64a) and (65a) have no bearing on the problem of determining the resonant frequencies, since the quantities involved are independent of the time.

Since $\delta J_{1n}{}^{(1)}$ is small, we may make use of a standard perturbation method for determining A_{1n} and ω. To do this we write

$$A_{1n}{}^{(1)} = A_{0n} + \delta B_{1n}{}^{(1)} + \delta^2 C_{1n}{}^{(1)} \cdots \tag{66}$$

$$\omega = \omega_{0n} + \delta\omega_{0n}{}' + \delta^2{}_{0n}{}'' \cdots \tag{67}$$

$$\phi_{1n}{}^{(1)} = \delta\psi_{1n}{}' + \delta^2\psi_{1n}{}'' \cdots . \tag{68}$$

$J_{1n}{}^{(1)}$ will in general depend on the frequency ω assumed by the system. When it does, the perturbation method can be kept consistent by expanding $J_{1n}{}^{(1)}$ in a power series in $\omega - \omega_0$. Then

$$J_{1n}{}^{(1)}(\omega, r) = J_{1n}{}^{(1)}(\omega_{0n}, r) + \left\{ \frac{d}{d\omega} J_{1n}{}^{(1)}(\omega, r) \right\}_{\omega 0n} (\omega - \omega_{0n}) + \cdots$$

$$= J_{1n}{}^{(1)}(\omega_{0n}, r) + \left\{ \frac{d}{d\omega} J_{1n}{}^{(1)}(\omega, r) \right\}_{\omega 0n} \delta\omega_{0n}{}' + \cdots .$$

When these equations are used in (64b) and coefficients of like powers of δ are equated, we have

$$\nabla \cdot \nabla A_{0n} + \epsilon\mu\omega_{0n}{}^2 A_{0n} = 0 \tag{69}$$

$$\nabla \cdot \nabla B_{1n}{}^{(1)} + \epsilon\mu \left[2\omega_{0n}\omega_{0n}{}' A_{0n} + \omega_{0n}{}^2 B_{1n}{}^{(1)} \right] = -\mu J_{1n}{}^{(1)}(\omega_{0n}) + i\mu\epsilon\omega_{0n} \operatorname{grad} \psi_{1n}{}'. \tag{70}$$

Equation (69) is already satisfied by A_{0n}. Since it is known that the vector potentials form an orthogonal set of functions, i.e.,

$$\int A_{0m} \cdot A_{0n}^* d\tau = \begin{cases} 0, & m \neq n \\ \text{constant}, & m = n \end{cases} \tag{71}$$

where A_{0n}^* is the complex conjugate of A_{0n} and the integral is taken over the volume of the cavity, we may expand $B_{1n}^{(1)}$ in terms of the functions A_{0n}. Thus

$$B_{1n}^{(1)} = \sum_{k=1}^{\infty} C_{1nk}^{(1)} A_{0k} \tag{72}$$

Substituting this into (70) and rearranging and making use of (69), we obtain

$$\sum_k \epsilon\mu \left[\omega_{0n}^2 - \omega_{0k}^2\right] C_{1nk}^{(1)} A_{0k} = -2\epsilon\mu\omega_{0n}\omega_{0n}' A_{0n} - \mu J_{1n}^{(1)}(\omega_{0n}) + i\mu\epsilon\omega_{0n} \operatorname{grad} \psi_{1n}'. \tag{73}$$

If we multiply this equation by A_{0n}^* scalarly and integrate the result over the volume of the cavity, we have the result

$$2\epsilon\mu\omega_{0n}\omega_{0n}' \int A_{0n}' \cdot A_{0n}^* d\tau + \mu \int \{J_{1n}^{(1)}(\omega_{0n}) + i\mu\epsilon\omega_{0n} \operatorname{grad} \psi_{1n}'\} \cdot A_{0n}^* d\tau = 0.$$

The orthogonality property of the A_{0n}'s has been used. If the potential $\phi_{1n}^{(1)}$ and consequently ψ_{1n}' is made zero on the boundary of the cavity, then it can be shown that

$$\int \operatorname{grad} \psi_{1n}' \cdot A_{0n}^* d\tau = 0$$

so that

$$\omega_{0n}' = -\frac{\int J_{1n}^{(1)}(\omega_{0n}) \cdot A_{0n}^* d\tau}{2\epsilon\omega_{0n} \int A_{0n} \cdot A_{0n}^* d\tau}. \tag{74}$$

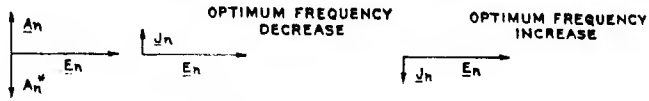
Since $J_{1n}^{(1)}$ is calculated from the unperturbed field distribution, the above formula shows that the first-order frequency change is

determined from the unperturbed vector potential, so that it is not necessary to calculate the field as modified by the electrons in order to determine the change in frequency caused by the electron stream, provided the change is small.

The quantity ω_{0n}' computed from (74) will, in general, be complex, so that the frequency change is actually the real part of (74), the imaginary terms giving rise to a net energy transfer from field to electrons. If the electrons gain energy and the wall losses in the cavity are small, this loss of energy to the electrons determines the Q of the cavity, so that (74) furnishes the basis for the following two formulas:

$$\triangle\omega = -Rp \frac{\int J_{1n}^{(1)}(\omega_{0n}) \cdot A_{0n}^* d\tau}{2\epsilon\omega_{0n} \int |A_{0n}|^2 d\tau} \tag{75}$$

$$\frac{\omega_{0n}}{2Q_{0n}} = -Im \frac{\int J_{1n}^{(1)}(\omega_{0n}) \cdot A_{0n}^* d\tau}{2\epsilon\omega_{0n} \int |A_{0n}|^2 d\tau} \tag{76}$$



(a) (b) (c)

Fig. 10—Phase diagrams for A, E, and J.

where $\triangle\omega$ is the change in frequency, Q_{0n} is the Q of the cavity with injected electrons, and the cavity is oscillating in the nth mode.

Before calculating the electron currents in some special cases, it will be of interest to discuss the implications of (75) and (76) with regard to the nature of the currents necessary to produce optimum frequency shifts, etc. If the unperturbed electric field in the cavity is represented by

$$E_n(r)\ e^{i\omega_{0n}t},$$

then from the relation (55) the vector potential may be written as

$$A_{0n}(r)\, e^{i\omega_{0n}t} = \frac{i}{\omega_{0n}} E_n(r)\ e^{i\omega_{0n}t} = \frac{E_n(r)}{\omega_{0n}} e^{i\pi/2}\, e^{\omega_{0n}t}. \tag{77}$$

Consequently the vector potential leads the electric field intensity in time phase by 90 degrees. The vector diagram in Figure 10(a) shows the time-phase relationship of E_{0n}, A_{0n}, and A_{0n}^* taking the phase angle of E_n as zero.

Since the denominators of (75) and (76) are real, the real and imaginary parts are determined by the scalar product

$$J_{1n}^{(1)}(r) \ : A_n^*(r)$$

in the integrals in the numerators. This scalar product, expressed in terms of $E_n(r)$, which is real, is

$$J_{0n}^{(1)}(\omega_{0n}, r) \cdot A_n^*(r) = J_{0n}^{(1)}(\omega_{0n}, r) \cdot \frac{E_n(r)}{\omega_{0n}} e^{-i\pi/2}. \qquad (78)$$

Consequently, for the optimum real part of this expression and therefore the optimum frequency shift, the electron current should lead or lag the electric field intensity by 90 degrees. For maximum decrease in frequency (75) shows that the electron current must lead E_n by 90 degrees, while for maximum increase in frequency the electron current should lag E_n by 90 degrees. This is analogous to the situation in circuit theory. The vector diagrams for these two cases are shown in Figures 10(b) and 10(c), respectively.

Formulas (76) and (78) show that optimum loading takes place when the current is in phase with E_n. This is the case when energy is transferred from the electro-magnetic field to the electrons. If the current were 180 degrees out of phase with E_n, energy would be transferred from the electrons to the field. In this case one would have a generator and a negative Q.

If the time phase of the current is a function of position in the cavity, as it usually will be, the net frequency shift is, of course, given by the integral.

In addition to the optimum time-phase relationships, (75) and (76) show that the greatest frequency change or loading for a given current occurs when that current is parallel to the electric field and located in that region of the cavity where the electric field is the largest. This is the region of largest interaction of field and electrons.

Calculation of Current

The frequency shift and loading will be computed for a rectangular cavity for a special way of introducing the electrons. We shall assume that before the electrons are introduced the mode of oscillation is such that the electric field is parallel to the x direction only, as shown in Figure 11. The electric field for this mode is given by

$$E_x(r, t) = E_0 \sin \frac{\pi y}{Y} \sin \frac{\pi z}{Z} e^{i\omega t} \tag{79}$$

and the frequency by

$$\omega_{0n} = \omega_0 = \pi C \sqrt{\frac{1}{Y^2} + \frac{1}{Z^2}}. \tag{80}$$

The x component of the vector potential corresponding to this mode is

$$A_x(r, t) = \frac{i}{\omega_0} E_0 \sin \frac{\pi y}{Y} \sin \frac{\pi z}{Z} e^{i\omega t}. \tag{81}$$

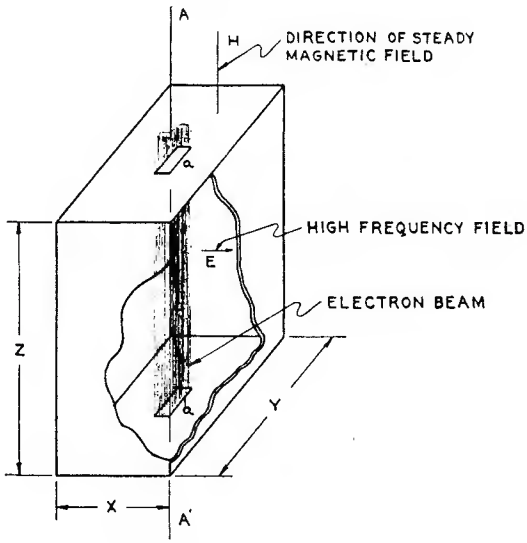


Fig. 11—Schematic diagram showing beam in rectangular resonant cavity.

We shall assume that the electron stream traverses the cavity in a direction parallel to the z axis, in which direction a uniform magnetic field H is assumed to be present. The determination of the first-order current $J_{1n}(\omega_0)$ for this case (see (62)) requires the calculation of the electron trajectories through the cavity under the influence of the unperturbed field given in (79). For definiteness we shall take the electron beam to be of rectangular cross section with center at $x = X/2$ and $y = Y/2$, all electrons entering the cavity at $z = 0$ with a velocity v_0.

Taking the initial conditions at $t = t_0$, the time of entrance, to be

$$x = \frac{X}{2}; \qquad y = \frac{Y}{2}; \qquad z = 0$$

$$\dot{x} = \dot{y} = 0 \quad \text{and} \quad \dot{z} = v_0,$$

the solutions of the equations of motion, (1), (2), (3), for the x component of velocity, $\dot{x}$, with the above applied field, is

$$\dot{x} = -\frac{|e|E_0}{m} F(z) e^{i\omega_0 t_0} \tag{82}$$

where

$$F(z) = \int_0^{z/v_0} \cos \omega_c(\eta - z/v_0) \sin \frac{\pi v_0}{Z} \eta e^{i\omega\eta} d\eta. \tag{83}$$

The charge entering the cavity per unit area of the beam between t_0 and $t_0 + dt_0$ will be

$$dq = -|J_0| dt_0$$

where J_0 is the current density in the beam before entering the cavity. The current density J in the x direction in the beam between z and $z + dz$ at time t is

$$Jdz = -|J_0| \frac{dz}{v_0} \dot{x} = \frac{|e||J_0|}{m} \frac{1}{v_0} E_0 F(z) e^{i\omega t_0} dz$$

$$J = \frac{|e||J_0|}{m} \frac{1}{v_0} E_0 F(z) e^{i\omega_0(t - z/v_0)}. \tag{84}$$

The amplitude of the current density entering (75) and (76), namely, $J_{1n}^{(1)}$, is just the time-independent part of the above expression; hence,

$$J_{1n}^{(1)}(\omega_0) = \frac{|e||J_0|}{m} \frac{1}{v_0} E_0 F(z) e^{-i\omega_0 z/v_0}. \tag{85}$$

This completes the calculation of the first-order approximation for the current.

Calculation of the Frequency Shift

The integral in the denominators of (75) and (76) is easy to compute. The result is

$$\int_0^Z \int_0^Y \int_0^X |A_{0n}|^2 dxdydz = \frac{E_0^2}{\omega_0^2} \frac{XYZ}{4}. \tag{86}$$

In the numerator of the same formulas, we require the value of

$$\int_0^Z \int_0^Y \int_0^X J_{1nx}^{(1)}(r) A_x^*(r)\, dxdydz$$

$$= \frac{-i|e|}{m} \frac{E_0^2}{vv_0} \int_0^Z \int_0^Y \int_0^X |J_0| F(z) e^{i\omega_0 z/v_0}$$

$$\cdot \sin \frac{\pi y}{Y} \sin \frac{\pi z}{Z}\, dxdydz.$$

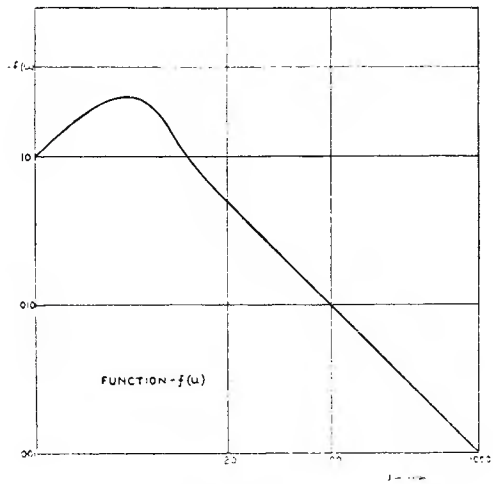


Fig. 12—$f(u)$ versus u.

The results (85) and (81) have been used. On the right side $|J_0|$ is different from zero only within the beam, so that, neglecting the variation in $\sin \pi y/Y$ over the beam width as before, the integrals over x and y can be evaluated at once since $\int\int |J_0|\, dxdy = |I_0|$ where $|I_0|$ is the total beam current. Then

$$\int\int\cdot\int J_{1nx}^{(1)}(r) A_x^*(r)\, dxdydz = |I_0| \frac{|e|}{m} \frac{E_0^2}{v_0\omega_0} K \tag{87}$$

where

$$K = -i \int_0^Z F(z)\, e^{-i\omega_0 z/v_0} \sin \frac{\pi z}{Z}\, dz. \tag{88}$$

K can be evaluated and for the real part we find

$$Rp\, K = \frac{Z}{4\pi} \left[\pi \left\{ \frac{(\omega_0 + \omega_c)}{\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 + \omega_c)^2} + \frac{\omega_0 - \omega_c}{\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 - \omega_c)^2} \right\} \right.$$
$$\left. -2\left(\frac{\pi v_0}{Z}\right)^3 \left\{ \frac{\sin (\omega_0 + \omega_c) z/v_0}{\left[\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 + \omega_c)^2\right]^2} + \frac{\sin (\omega_0 - \omega_c) z/v_0}{\left[\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 - \omega_c)^2\right]^2} \right\} \right] \tag{89}$$

The final formula for the frequency shift can be materially simplified by adopting new variables, namely,

$$u = (1 + k)\alpha \quad \text{and} \quad w = (1 - k)\alpha \tag{90}$$

where

$$k = \frac{\omega_c}{\omega_0} \quad \text{and} \quad \alpha = \frac{Z\omega_0}{v_0} = \text{transit angle.} \tag{91}$$

When the above formulas are substituted in (75), we have the expression for the frequency shift. This is

$$\Delta\omega = -\frac{|I_0|\,|e|\,\alpha^2}{2\epsilon m X Y Z \omega_0^2} [f(u) + f(w)] \tag{92}$$

where

$$f(u) \frac{u(\pi^2 - u^2) - 2\pi^2 \sin u}{(\pi^2 - u^2)^2}. \tag{93}$$

A graphical representation of $f(u)$ is shown in Figure 12.

The main characteristics of this method of changing the frequency are that the frequency is proportional to the total beam current and inversely proportional to the energy stored in the electromagnetic field within the cavity. A further examination of (92) and (93) shows that when the transit angle α is not too small, the maximum frequency shift $\triangle\omega_{max}$ occurs when

$$(1 - k)\alpha \approx 4 \tag{94}$$

and that $\triangle\omega_{max}$ is proportional to α^2. The above condition for maximum

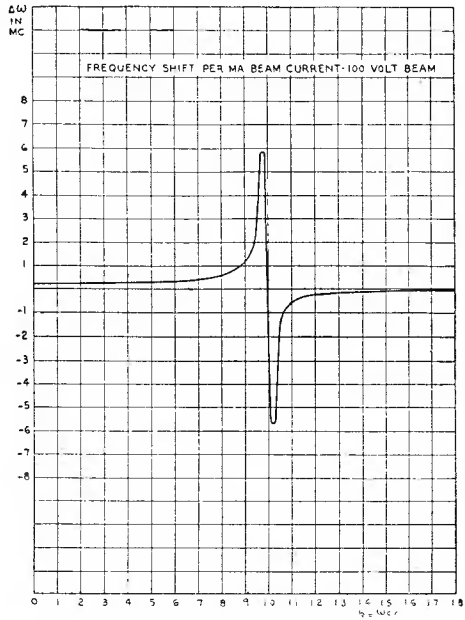


Fig. 13—Frequency shift versus magnetic field for the rectangular cavity structure.

frequency shift is not quite the same as that found for electrons moving between parallel plates.

In order to indicate the magnitude of the frequency change and its behavior as a function of the magnetic field, a curve is shown in Figure 13 which gives $\triangle\omega$ as a function of k for a rectangular cavity where $Y = Z = 5.3$ centimeters, $X = 0.2$ centimeter, and $\omega_0 = 4000$ megacycles, and the transit angle corresponds to that of electrons having been accelerated through 100 volts. One may readily verify that the maximum of $\triangle\omega$ occurs when the condition (94) is satisfied. The curve shows the greatly increased frequency shift that can be obtained by adjusting the magnetic field so that the electron is nearly

resonant with the oscillating electric field over that which would result for zero magnetic field.

When $k < 1$, the electron current is inductive, i.e., lags behind the electric field producing it, and gives rise to an increase in resonant frequency of the system. When $k > 1$ the electron current is capacitive, i.e., leads the electric field, and produces a decrease in resonant frequency.

It should be noted that the above behavior applies to cases where the beam is not limited by space-charge effects and the orbits of the electrons are not large enough to strike the walls of the cavity. An estimate of the effects of these factors can be made for any special case.

Electron Loading

In order to calculate the electronic loading, we note from (76) that the imaginary part of K given by (88) is required. Thus

$$\frac{\omega_0}{2Q_0} = \frac{2\,|I_0|\,|e|}{\epsilon m v_0 XYZ}\, Im\, K. \tag{95}$$

The value of the ImK is

$$ImK = v_0 \left[\frac{\left(\frac{\pi v_0}{Z}\right)^2 \cos^2(\omega_0 - \omega_c)\frac{Z}{2v_0}}{\left\{\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 - \omega_c)^2\right\}^2} + \frac{\left(\frac{\pi v_0}{Z}\right)^2 \cos^2(\omega_0 + \omega_c)\frac{Z}{2v_0}}{\left\{\left(\frac{\pi v_0}{Z}\right)^2 - (\omega_0 + \omega_c)^2\right\}^2} \right]. \tag{96}$$

Using the variable u and w given in (90) and (91), the above expressions can be combined and rearranged so that

$$\frac{\omega_0}{2Q_0} = \frac{I_0\,|\,e\,|\,\alpha^2}{2\epsilon m XYZ\omega_0^2}\,[T(w) + T(u)] \tag{97}$$

where

$$T(w) = \frac{4\pi^2 \cos^2 w}{[\pi^2 - w^2]^2}. \tag{98}$$

A graph of $T(w)$ is shown in Figure 14.

Just as in the case of the frequency shift, the loading is propor-

tional to the total beam current and inversely proportional to the cavity volume and the square of the resonant frequency. The maximum loading occurs when $w = (1 - k)\ \alpha = 0$. For a transit angle different from zero this occurs at $k = 1$, i.e., $\omega_c = \omega_0$, which means that the electron circular motion is exactly in resonance with the electric field. As seen from Figure 13, $\triangle\omega$ is zero at this point, so that the electron current is entirely a resistive component.

When the transit angle is not too small the loading is essentially zero at

$$w = (1 - k)\ \alpha = 3\pi \tag{99}$$

since $T(w)$ is zero at this point and $T(u)$ is very small. In this case we find that minimum loading occurs at $(1 - k)\ \alpha = 3\pi$, instead of 2π

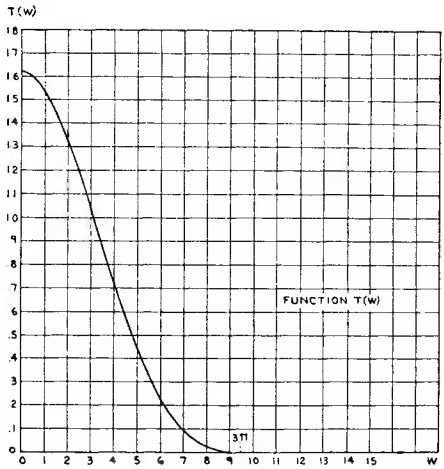


Fig. 14—$T(w)$ versus w.

as it was for electrons moving in a uniform oscillating field between parallel plates.

If we compare Figures 12 and 13, we find that for a given transit time (not too short) when $(1 - k)\ \alpha$ is adjusted for no loading, the frequency shift is about four-tenths of the maximum frequency shift, i.e.,

$$\triangle\omega_{\text{no loading}} \approx 0.4\triangle\omega_{\text{max}}. \tag{100}$$

Again this is a somewhat different ratio from the case of electrons moving between parallel planes.

APPENDIX

Fringe-Field Effect

In general, the problem of estimating the effect of the fringe field depends on the geometry of the specific system. If, however, the transit time in the fringe field is small, the effects can be approximated as follows.[1]

Consider a charge q emerging from between a pair of parallel-plane electrodes, and let the charge be displaced a distance x from the median plane. The induced charge on one plane is

$$-\frac{q}{d}\left(x+\frac{d}{2}\right)$$

and that on the other is

$$-\frac{q}{d}\left(-x+\frac{d}{2}\right).$$

The difference between the charge on the two planes is $-q(2x/d)$. After the electron has traversed the fringe field in the exit space, this difference of charge must have disappeared, so that during the transit time in the fringe field a charge $-q(x/d)$ must have moved from one electrode to the other through the external circuit.

If a beam of current $-|I_0|$ is moving between the plates at a velocity v_0, the charge at the exit point is

$$q=\frac{-|I_0|}{v_0}dz$$

and the charge

$$+\frac{|I_0|}{dv_0}xdz=\int i_f dt$$

is the charge moved from one electrode to the other during transit through the fringe field. i_f is the current in the external circuit due to this rearrangement of charge.

$$i_f=+\frac{|I_0|}{dv_0}x\frac{dz}{dt}=+\frac{|I_0|}{d}x. \tag{101}$$

The fringe effects on *entering* the space between the parallel planes are negligible, for the electrons are assumed to enter at $x = 0$. From integration of (6)

$$x = + E_0 \frac{|e|}{m} \frac{\tau}{2\omega_c \theta} (e^{i\theta} - 1)\ e^{i\omega t}. \tag{102}$$

If

$$i_f = I_f e^{i\omega t}$$

$$I_f = \frac{|I_0|}{2d} E_0 \frac{|e|}{m} \frac{\tau}{\omega_c \theta} (e^{i\theta} - 1).$$

The total current can be regarded as the sum of two parts, so that $I = I_1 + I_f$ where I_1 is given by (9).

$$I = \frac{|I_0|}{2d} E_0 \frac{|e|}{m} \frac{\tau^2}{\theta^2} \left[\frac{\omega_c - \omega}{\omega_c} (e^{i\theta} - 1) + (1 + i\theta - e^{i\theta}) \right]$$

Y_{cf} = over-all admittance including fringe-field effects.

$$Y_{cf} = \frac{L^2}{4d^2} \frac{|I_0|}{V_0} \left\{ \frac{1 - \cos\theta}{\theta^2} \left(1 - \frac{\omega_c - \omega}{\omega_c} \right) + i \frac{\theta - (\sin\theta) \left(1 - \dfrac{\omega_c - \omega}{\omega_c} \right)}{\theta^2} \right\}. \tag{103}$$

It is clear from the above that, for small transit times through the fringe field, the contribution to the image currents from the fringe electrons is of the same order of magnitude of terms already neglected in this theory and should be neglected here.

CARRIER AND SIDE-FREQUENCY RELATIONS WITH MULTI-TONE FREQUENCY OR PHASE MODULATION*†

By

MURRAY G. CROSBY

RCA Communications, Inc., Riverhead. L. I., N. Y.

Summary

The equation for the carrier and side frequencies of a frequency or phase-modulated wave is resolved for the case of two applied modulating tones. It is shown that when more than one modulating tone is applied, the amplitude of the carrier is proportional to the product of the zero-order Bessel Functions of all the modulation indexes involvcd. The amplitudes of the side frequencies are proportional to the products of Bessel Functions equal in number to the number of tones applied and having orders respectively equal to the orders of the frequencies involved in the side frequency. Beat side frequencies are produced which have higher-order amplitude and do not appreciably widen the band width occupied by frequency modulation.

(4 pages)

* Decimal Classification: R630.1.

† *RCA REVIEW*, July, 1938.

SOME RECENT DEVELOPMENTS IN RECORD REPRODUCING SYSTEMS*†

By

G. L. BEERS‡ AND C. M. SINNETT#

RCA Manufacturing Co., Inc., Camden, N. J.

Summary

Several factors of importance in obtaining satisfactory reproduction of sound from lateral-cut phonograph records are considered. An experimental record reproducing system employing the principles of frequency modulation is described and data are supplicd on the measured and calculated performance characteristics of the system. Curves are included showing the vertical force required for satisfactory tracking with the experimental frequency-modulation pickup as compared with other pickups of conventional design.

(9 pages; 18 figures; Appendix)

* Decimal Classification: R681.843.

† *Proc. I.R.E.*, April, 1943.

‡ Now Asst. Director of Engineering in Charge of Advanced Development, RCA Victor Division, Camden, N. J.

Now with the Home Instrument Department, RCA Victor Division, Camden, N. J.

FREQUENCY MODULATION OF RESISTANCE-CAPACITANCE OSCILLATORS*†

By

MAURICE ARTZT

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

A method is described for direct-frequency modulation of a resistance-capacitance oscillator which is simpler and more stable then the beat oscillators formerly used. Spurious amplitude modulation is reduced to a negligible value without the use of limiters and without introducing appreciable harmonics in the output wave. Balanced control tubes prevent transients or signal frequencies from appearing in the output. Curves are given which provide an easy method of choosing the network and constants for any desired condition. The device is especially suited to facsimile transmission by sub-carrier frequency modulation.

(6 pages; 9 figures)

* Decimal Classification: R148.5.

† *Proc. I.R.E.*, July, 1944.

AN F-M CALIBRATOR FOR DISC RECORDING HEADS*†

By

H. E. ROYS

Engineering Products Department, RCA Victor Division,
Camden, N. J.

Summary

Since the beginning of disc recording, a device has been needed which would permit the calibration of a recording head under actual cutting conditions. Such a devise is described in this paper. First, however, various former methods (such as sunlight pattern, microscope, and photoelectric cell) are discussed. The paper then treats in detail with the FM method and describes the FM calibrator. A comparison of FM and optical calibrators is included together with discussions on change in frequency response due to cutting load, change in recording level with groove width, distortion, and monitoring.

(6 pages; 10 figures)

* Decimal Classification: R681.843.

† *Broadcast News*, June, 1946.

FREQUENCY-MODULATION MOBILE RADIO-TELEPHONE SERVICES*†

By

H. B. Martin

Assistant Chief Engineer, Radiomarine Corporation of America, New York, N. Y.

Summary

The proposed use of frequencies in the Very High Frequency ranges of 30-44 and 152-162 megacycles for Common-carrier General Mobile Radiotelephone Communication is discussed with reference to propagational and equipment advantages. Consideration is limited to the radio link for vehicular and marine mobile service. Frequency modulation and its advantages for mobile communication are discussed briefly. The use of separate transmitting and receiving frequencies is the basis for all recommendations and allocation plans which are traced from their origin by the Radio Technical Planning Board to the most recent Industry suggestions to the Federal Communications Commission. Geographical considerations for service areas of fixed stations are based on Radio Manufacturers Association recommendations for transmitter powers, antenna heights, desired and undesired signal levels, etc. The Urban and Highway proposed services, including Marine, are compared on an operational and equipment basis and the most important mobile unit performance requirements are given. References from Radio Technical Planning Board (RTPB), Radio Manufacturers Association (RMA), Federal Communications Commission (FCC) and technical publications are provided.

(12 pages; 4 tables)

* Decimal Classification: R540.

† *RCA REVIEW*, June, 1946.

A FREQUENCY-MODULATED MAGNETRON FOR SUPER-HIGH FREQUENCIES*†

By

G. R. Kilgore, Carl I. Shulman, and J. Kurshan

Research Department, RCA Laboratories Division, Princeton, N. J.

Summary

This paper is based on a wartime requirement for a 25-watt, 4000-megacycle continuous-wave oscillator capable of electronic frequency modulation with a deviation of at least 2.5 megacycles. A satisfactory solution was found in the addition of frequency control to a continuous-wave magnetron by the introduction of electron beams into the magnetron cavities in a manner described by Smith and Shulman. This method is referred to as "spiral-beam" control.

A brief account is given of the method of designing a continuous-wave magnetron for the specified power and frequency. The problem of adding

spiral-beam frequency control to this magnetron is considered in detail, and a procedure is presented for obtaining the optimum design consistent with negligible amplitude modulation and reasonable cathode-current densities. The unusual features of construction of this magnetron are described, including a method of mechanical tuning and the method of adding grid-controlled beams for frequency modulation.

Performance data on the continuous-wave magnetron over a wide range of operating conditions indicate that the required 25 watts output can be attained at 850 volts with 50 per cent efficiency. Experimental results on "spiral-beam" frequency control demonstrate that the required 2.5-megacycle frequency deviation with no amplitude modulation can be achieved by electron beams introduced into two out of twelve cavities. A deviation of 4 megacycles is attained under conditions allowing some amplitude modulation. Still greater deviation is predicted by using more beams.

(8 pages; 10 figures)

---

* Decimal Classification: R355.912.1.

† *Proc. I.R.E.*, July, 1947.

A 1-KILOWATT FREQUENCY-MODULATED MAGNETRON FOR 900 MEGACYCLES*†

By

J. S. Donal, Jr., R. R. Bush, C. L. Cuccia, and H. G. Hegbar

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

The method of Smith and Shulman has been used for the frequency modulation of a 1-kilowatt continuous-wave magnetron. This tube is of the "Vane" type, having twelve resonant cavities, and it is mechanically tunable over a range from about 720 to 900 megacycles by a cylindrical element which varies the interstrap capacitance. At the applied magnetic field required for frequency modulation without change in amplitude, 1 kilowatt output at 900 megacycles is obtained with an anode voltage of 2.5 kilovolts and an efficiency of about 55 per cent; the efficiency rises with decreasing frequency or with increasing magnetic field.

At 900 megacycles, electron beams in nine of the magnetron resonant cavities give a frequency deviation of 3.5 megacycles (a total frequency swing of 7 megacycles) at an output of 1 kilowatt, rising to 4 megacycles at an output of 750 watts. The frequency deviation is reduced when the tube is tuned to lower frequencies. The modulator power required would be very low, since the grid-cathode capacitance of the frequency-modulation guns is small and the grids draw no current.

It would be practicable to increase the frequency deviation of this tube by about 15 per cent through an increase in beam current, and by an additional 20 per cent through the use of eleven beams. A change in the type of beam cathode would effect an even greater deviation.

(6 pages; 6 figures; 1 table)

---

* Decimal Classification: R355.912.1.

† *Proc. I.R.E.*, July, 1947.

APPENDIX I

FREQUENCY MODULATION

A Bibliography of Technical Papers by RCA Authors 1936 — 1947

This listing includes some 75 technical papers on FREQUENCY MODULATION and closely related subjects, selected from those written by RCA Authors and published during the period 1936-1947.

Papers are listed chronologically except in cases of multiple publication. Papers which have appeared in more than one journal are listed once, with additional publication data appended.

A List of Sources is given on the following page.

LIST OF SOURCES

Broadcast News — *Communications*

Electronics — *FM and Television*

RCA REVIEW

Jour. of the Society of Motion Picture Engineers

Proceeding of the Institute of Radio Engineers

RCA Licensee Bulletins
(RCA Licensee Bulletins are not published and are issued only as a service to licensees of the Radio Corporation of America.)

FREQUENCY MODULATION BIBLIOGRAPHY

"Frequency Modulation Propagation Characteristics", M. G. Crosby, *Proc. I.R.E.* (June) 1936

"Frequency Modulation Noise Characteristics", M. G. Crosby, *Proc. I.R.E.* (April) 1937

"Carrier and Side Frequency Relations with Multi-Tone Frequency or Phase Modulation", M. G. Crosby, *RCA REVIEW* (July) 1938

"Frequency Modulation", S. W. Seeley, *RCA Licensee Bulletin LB-505* (January) 1940

"The Service Range of Frequency Modulation", M. G. Crosby, *RCA REVIEW* (January) 1940

"A Cathode Ray Frequency Modulation Generator", R. E. Shelby, *Electronics* (February) 1940

"Comparative Characteristics of Frequency and Amplitude Modulation Receivers", C. N. Kimball and S. W. Seeley, *RCA Licensee Bulletin LB-515* (March) 1940

"Signal Generators for Frequency Modulated Waves", J. A. Rankin and D. E. Foster, *RCA Licensee Bulletin LB-514* (March) 1940

"Sound Broadcasting on Ultra-High Frequencies with Special Reference to Frequency Modulation", A. F. Van Dyck, *RCA Licensee Bulletin LB-518* (March) 1940

'Frequency Modulation Receiver Considerations", J. A. Rankin and D. E. Foster, *RCA Licensee Bulletin LB-510* (May) 1940

"Reactance-Tube Frequency Modulators", M. G. Crosby, *RCA REVIEW* (July) 1940

"Dual I-F Amplifier for FM-AM Reception", G. Mountjoy, *RCA Licensee Bulletin LB-535* (October) 1940

"Frequency Modulation Field Tests", R. F. Guy, *RCA REVIEW* (October) 1940

Radio (January) 1941

"Spurious Responses in FM Receivers", D. E. Foster and J. A. Rankin, *RCA Licensee Bulletin LB-543* (December) 1940

"A Transmitter for Frequency Modulated Broadcast Service Using a New Ultra-High-Frequency Tetrode", A. K. Wing and J. E. Young, *RCA REVIEW* (January) 1941

"Band Width and Readability in FM", M. G. Crosby, *RCA REVIEW* (January) 1941

"Impulse Noise in FM Reception", V. D. Landon, *Electronics* (February) 1941

"Frequency Modulation", S. W. Seeley, *RCA REVIEW* (April) 1941

"Drift Analysis of the Crosby Frequency-Modulated Transmitter Circuit", E. S. Winlund, *Proc. I.R.E.* (July) 1941

"Duplex Transmission of Frequency-Modulated Sound and Facsimile", *RCA REVIEW* (July) 1941

"Intermediate-Frequency Values for Frequency-Modulated-Wave Receivers", D. E. Foster and J. A. Rankin, *Proc. I.R.E.* (October) 1941

"The Development of a Frequency-Modulated Police Receiver for Ultra-High-Frequency Use", H. E. Thomas, *RCA REVIEW* (October) 1941

"Generation and Detection of FM Waves", A. Barco, C. N. Kimball and S. W. Seeley, *RCA REVIEW* (January) 1942

"Frequency-Modulation Distortion in Loudspeakers", G. L. Beers and H. Belar, *Proc. I.R.E.* (April) 1943
Jour. Soc. Mot. Pic. Eng. (April) 1943
"Some Recent Developments in Record Reproducing Systems", G. L. Beers and C. M. Sinnett, *Proc. I.R.E.* (April) 1943
"FM & UHF", R. F. Guy, *Communications* (August) 1943
"Use of Subcarrier Frequency Modulation in Communication System", W. H. Bliss, *Proc. I.R.E.* (August) 1943
"FM Noise Level and AM Noise Level", *Broadcast News* (January) .. 1944
"An F-M Receiver Circuit", *RCA Licensee Bulletin LB-616* (February) 1944
"A New System for Crystal Control of F-M Carriers", G. Mountjoy, *RCA Licensee Bulletin LB-621* (June) 1944
"Frequency Modulation of Resistance-Capacitance Oscillators", M. Artzt, *Proc. I.R.E.* (July) 1944
"Coverage Curves for FM Antennas", R. D. Duncan, Jr., *Broadcast News* (August) 1944
"FM Audio Measurements with an AM Receiver", R. J. Newman, *Broadcast News* (August) 1944
"Antennas for FM Stations", J. P. Taylor, *Broadcast News* (August) 1944
"A Frequency-Dividing Locked-in Oscillator Frequency Modulation Receiver", G. L. Beers, *Proc. I.R.E.* (December) 1944
"Selecting a Site for an FM Station", J. P. Taylor, *Broadcast News* (January) 1945
"Square Loop Frequency Modulation Antenna", J. P. Taylor, *Electronics* (March) 1945
"A Pretuned Turnstile Antenna", G. H. Brown and J. Epstein, *Electronics* (June) 1945
"How to Determine the Area an FM Station Should Serve", J. P. Taylor, *Broadcast News* (June) 1945
"How to Determine the Required Transmitter Power of an FM Station", J. P. Taylor, *Broadcast News* (June) 1945
"Audio Frequency Response and Distortion Measuring Techniques for FM Transmitting Stations", R. J. Newman, *Broadcast News* (June) 1945
"Quench AVC for F-M Receivers", *RCA Licensee Bulletin LB-636* (July) 1945
"FM Antenna Coupler", J. P. Taylor, *Electronics* (August) 1945
"Ratio Detectors for F-M Receivers", S. W. Seeley, *RCA Licensee* Bulletin *LB-645* (September) 1945
"Frequency-Modulation Distortion Caused by Multi-path Transmission", M. S. Corrington, *Proc. I.R.E.* (December) 1945
"High Frequency R-F Circuits for AM-FM Receivers", E. I. Anderson, *RCA Licensee Bulletin LB-653* (December) 1945
"Multi-Unit Construction—A Feature of New FM Transmitters", J. L. Ciba, *Broadcast News* (January) 1946
"Push-Pull Frequency Modulated Circuit and Its Application to Vibratory Systems", A. Badmaieff, *Jour. Soc. Mot. Pic. Eng.* (January) 1946
"Range Prediction Chart for FM Stations", F. C. Everett, *Communications* (November) 1945
Broadcast News (January) 1946
"A New Exciter Unit for Frequency Modulated Transmitters", N. J. Oman, *RCA REVIEW* (March) 1946
"Balanced Phase Shift Discriminator", S. W. Seeley, *RCA Licensee Bulletin LB-666* (March) 1946

"Input Impedance of Several Receiving-Type Pentodes at F-M and Television Frequencies", F. Mural, *RCA Licensee Bulletin LB-661* (March) 1946

"The Transmission of a Frequency-Modulated Wave Through a Network", W. J. Frantz, *Proc. I.R.E.* (March) 1946

"Grounded-Grid Power Amplifiers", E. E. Spitzer, *Electronics* (April) 1946

"Frequency Modulation Mobile Radiotelephone Services", H. B. Martin, *RCA REVIEW* (June) 1946

"Development of an Ultra-Low Loss Transmission Line", E. O. Johnson, *RCA REVIEW* (June) 1946

"An FM Calibrator for Disc Recording Heads", H. E. Roys, *Broadcast News* (June) 1946

"A Discussion of FM Field Survey Techniques", *Broadcast News* (June) 1946

"How to Make a Field Survey of an FM Station", G. W. Klingaman, *Broadcast News* (June) 1946

"Characteristics of the Pylon FM Antenna", R. F. Holtz, *FM and Tele.* (September) 1946

"Determining the Population Served by an FM Station", *Broadcast News* (October) 1946

"Isolation Methods for FM Antennas Mounted on AM Towers", R. F. Holtz, *Broadcast News* (October) 1946

"Frequency-Modulation Distortion Caused by Common- and Adjacent-Channel Interference", M. S. Corrington, *RCA REVIEW* (December) 1946

"Twelve-Channel FM Converter", J. E. Young and W. A. Harris, *Electronics* (December) 1946

"Input Circuit Noise Calculations for F-M and Television Receivers", W. J. Stotze, *Communications* (February) 1947

"Explanation of the Ratio Detector as an Aid in FM Servicing", J. A. Cornell, *RCA Rad. Serv. News* (March-April) 1947

"Transmission Lines and Antennas for FM and Television", M. Kaufmann, *Radio Maintenance* (March and April) 1947

"The Ratio Detector", J. Avins, *RCA Licensee Bulletin LB-710* (May 26) 1947

"The Ratio Detector", S. W. Seeley and J. Avins, *RCA REVIEW* (June) 1947

"An Analysis of Modern Antennas for FM and Television Reception", M. Kaufman, *RCA Rad. Serv. News* (June-July) 1947

"A Non-Directional Antenna for Mobile Field Strength Measurements in the FM Band", B. W. Robins, *Broadcast News* (June) 1947

"Triplex Antenna for Television and FM", L. J. Wolf, *Electronics* (July) 1947

"Frequency Modulation and Control By Electron Beams", L. P. Smith and Carl I. Shulman, *Proc. I.R.E.* (July) 1947

"A Frequency-Modulated Magnetron for Super-High Frequencies", G. R. Kilgore, Carl I. Shulman and J. Kurshan, *Proc. I.R.E.* (July) 1947

"A 1-Kilowatt Frequency-Modulated Magnetron for 900 Megacycles", J. S. Donal, Jr., R. R. Bush, C. L. Cuccia, and H. R. Hegbar, *Proc. I.R.E.* (July) 1947

"Generalized Theory of Multitone Amplitude and Frequency Modulation", L. J. Giacoletto, *Proc. I.R.E.* (July) 1947

"Servicing FM Receivers", M. Kaufman, *Radio Maintenance*, "Part I —Basic Theory" (July) 1947
"Part II—Conventional Circuits" (August) 1947
"Part III—FM Detectors" (September) 1947

"An FM Quality Speaker", G. E. Rand, *Broadcast News* (September). **1947**

"An Analysis of FM Servicing", N. L. Chalfin, *RCA Rad. Serv. News* (October-November) 1947

"Variation of Bandwidth with Modulation Index in Frequency Modulation", M. S. Corrington, *Proc. I.R.E.* (October) 1947

"10.7 Megacycle FM Intermediate Frequency Signal Generator", W. R. Alexander, *RCA Licensee Bulletin LB-737* (November 10)..... 1947

"Ratio Detectors for FM Receivers", S. W. Seeley, *FM and Tele.* (December) 1947

"Slot Antennas", N. E. Lindenblad, *Proc. I.R.E.* (December) **1947**

APPENDIX II

FM STATION PLACEMENT AND FIELD SURVEY TECHNIQUES

The ever-increasing number of FM stations now in the planning stage has focused attention on the difficult problems of station placement and field survey work. It is considered both appropriate and timely to include herewith a bibilography of material relating to these problems with the assurance that this guide will assist to some degree in the solution of these problems.

"Selecting a Site for an FM Station"*, John P. Taylor‡, *Broadcast News* ..... (January) 1945.

> ***Summary***—*This paper treats with the problem of choosing an appropriate station location from a very practical point of view. First, the factors to be considered are discussed in some detail including technical requirements, importance of central location, height, and other practical considerations. Then, various types of FM sites are covered including tall buildings, mountain tops, AM towers and special structures.*

* Decimal Classification: R630.11 x R630.2
‡ Engineering Products Department, RCA Victor Division, Camden, N. J.

"How to Determine the Area an FM Station Should Serve"*, John P. Taylor‡, *Broadcast News* ..... (June) 1945.

> ***Summary***—*Details of the procedure to be followed in determining the "trade area" and "service area" of an FM station are covered in this paper. The problems dealt with for both the first and subsequent applicants for a particular area. Sources of "trade area" charts and maps are given, and the use of these maps is discussed. The problems of "limited trade area" and other classes of stations are also covered.*

* Decimal Classification: R630.11
‡ See previous paper.

"How to Determine the Required Transmitter Power of an FM Station"*, John P. Taylor‡, *Broadcast News* ..... (June) 1945.

> ***Summary***—*The method of calculating the transmitter power required to provide coverage of a specified service area when the antenna elevation and gain are known is discussed in this paper. Topics covered include: starting data, obtaining topographic maps, drawing the radials, other sources of elevation data, determining coverage for a given power, coverage in eight directions, effect of available transmitter power, and simultane-*

ous consideration of power and location. Appropriate charts and tables are also included.

---

* Decimal Classification: R630.11.

‡ See previous papers.

"Range Prediction Chart for FM Stations"*, F. C. Everett‡, *Communications* ..... (November) 1945.

Broadcast News (reprinted from *Communications*) ... (January) 1946.

***Summary**—The chart makes it possible to determine the signals that can be expected with various powers, antenna heights and distances. Heights indicated are those of transmitting antenna over average terrain to point in question and receiving antenna height of 30 feet. Several examples of the use of the chart are included.*

---

* Decimal Classification: R630.11.

‡ Engineering Department, National Broadcasting Co., Inc., New York, N. Y.

"How to Make a Field Survey of an FM Station"*, G. W. Klingaman‡, *Broadcast News* ..... (June) 1946.

***Summary**—This article discusses the mechanics involved in making a practical field survey for an FM station and presents typical data from surveys which have been made. Topics covered include: methods, equipment, making the survey, organizing the data, and discussion of the results.* **(See also next paper)**

---

* Decimal Classification: R270 x R630.11.

‡ Engineering Products Department, RCA Victor Division, Camden, N. J.

"A Discussion of FM Field Survey Methods"*, *Broadcast News* ..... (June) 1946.

***Summary**—This paper carries a reprint of Section 5 of the FCC's "Standard of Good Engineering Practise Concerning FM Broadcast Stations" and disqusses FM field survey methods in the light of these standards. The paper includes sections on methods of making mobile recordings, method of analyzing recordings, determining the 1000 μv/m contour, determining the 50 μv/m contour, determining the interference area, maps, exhibits, etc. Procedures outlined by the FCC differ slightly from those described in the preceding paper and these slight differences should be noted.*

---

* Decimal Classification: R270 x R630.11.

"Determining the Population Served by an FM Station"*, *Broadcast News* ..... (October) 1946.

Summary—*The method of using Minor Civil Division Maps and Population Bulletins in determining population within the practical service area of an FM station is discussed. This information is required when applying for FM Construction Permits. The various steps in the required procedure are clearly outlined and explained.*

* Decimal Classification: R630.11.

"A Non-Directional Antenna for Mobile Field Strength Measurements in the FM Band"*, B. W. Robins‡, *Broadcast News*..... (June) 1947.

Summary—*FCC "Standards" require the use of a non-directional receiving antenna when making field strength measurements of FM Broadcast Stations. Such an antenna, capable of being used with the Standard Field Intensity Meter is described in this paper. The antenna is of simple arrangement but must be tailor-made to fit conditions. Hence, it is not a regularly manufactured item. The experimental work on and with this antenna is described, however, so that station engineers may adapt its principles for their own specific purposes. The paper includes antenna specifications, mounting problems, and a number of representative patterns obtained through its use.*

* Decimal Classification: R630.11.

‡ Engineering Products Department, RCA Victor Division, Camden, N. J.