#2

INTERFERENCE REJECTION IN FM RECEIVERS

ELIE J. BAGHDADY

Loan Copy Only

TECHNICAL REPORT 252

SEPTEMBER 24, 1956

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U. S. Navy (Office of Naval Research), and the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-64637, Department of the Army Task 3-99-06-108 and Project 3-99-00-100.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

Technical Report 252 September 24, 1956

INTERFERENCE REJECTION IN FM RECEIVERS

Elie J. Baghdady

This report is based on a thesis submitted to the Department of Electrical Engineering, M.I.T., May 16, 1956, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

A new role is suggested for the amplitude limiter in FM receivers. By spreading out the spectrum which is necessary for the reproduction of the FM disturbance that is caused by the interference, the limiter makes it possible for a filter to reject an important portion of this spectrum without substantially affecting the spectrum that carries the message modulation. The conditions for the success of this operation are analyzed in terms of an ideal limiter followed by an idealized filter. The variation of the required minimum extent of linearity in the discriminator characteristic with the limiter bandwidth is determined. This is followed by a study of the effect upon the interference of a repeated cycle of amplitude limiting and spectrum filtering. The cascading of several narrow-band limiters is found to be an invaluable scheme for enhancing the capture capabilities of an FM receiver.

I. LIMITER AND DISCRIMINATOR BANDWIDTH REQUIREMENTS

INTRODUCTION

Essential to the interference rejection ability of a frequency-modulation receiver is the use of the proper bandwidths in its nonlinear sections. The weaker of two competing signals (whose amplitude may approach the amplitude of the stronger signal within arbitrary limits) can be suppressed by a frequency-modulation (FM) receiver if, other requirements being met, the limiter and discriminator bandwidths exceed certain minimum permissible values. A brief survey of the problem of interference rejection and the FM receiver design requirements set by previous investigators (1, 2) is made. This is followed by a study of the spectrum of the amplitude-limited resultant of two carriers differing in amplitude as well as in frequency. The properties exhibited by the spectral components lead to a simple criterion for interference suppression when only certain portions of the spectrum are passed by an ideal filter that follows the limiter. The criterion is tacitly based upon the assumption that it is the message carried by the stronger signal that we desire to get through, although the conditions for reliable capture of the weaker signal will be treated in a separate discussion. The interference rejection criterion is then used to calculate the minimum bandwidths required after the limiter in order to preserve the interference rejection ability of the receiver for capture ratios up to 0.98.

A narrow-band filter after the limiter will, in general, distort the pattern of the instantaneous-frequency perturbations caused by the interference. This distortion will vary with the bandwidth of the filter and with the position of the stronger of the two signals relative to the center frequency of the filter, as well as with the frequency of the weaker signal relative to the stronger one. The configurations leading to the largest instantaneous-frequency deviations (from the desired average frequency) with various values of limiter bandwidth are studied to determine the corresponding minimum necessary ranges over which the FM-to-AM detection characteristic of the receiver must be linear. The results of this study will reveal how the first stage of bandpass limiting will modify the character of the resultant signal passed. They will also show how the minimum requirement in discriminator bandwidth will vary with the value of limiter bandwidth. The effect upon other design considerations, such as the time-constant requirements of the limiter and discriminator circuits is taken up in Section II.

In this report the investigation is carried out by the Fourier method on a steady-state basis and in terms of an ideal limiter followed by an ideal bandpass filter. In a future discussion of the nature of rejectable interferences and of the theory of capture in frequency modulation, the results of the present study will be correlated with conclusions derived from a study of the dynamic steady-state response of a filter. This alternative approach deals, in general, with physical filters as energy storing systems, and stresses their inertia to fast frequency and amplitude changes rather than their frequency-selectivity properties. The Fourier approach, which utilizes idealized band-pass filters, has been chosen because it enables us to reach many important theoretical conclusions which are otherwise not so easily demonstrated.

Throughout our discussions the concept of instantaneous frequency is used frequently and our understanding of it is fully exploited. A brief statement of the mathematical description of this useful concept, together with a brief survey of the problem of interference rejection, the requirements imposed in receiver design, and the bearing of "narrow-band limiting" upon these requirements are taken up first. The term "frequency" is loosely used to mean "angular frequency." When cyclic frequency is meant, it will be specifically stated.

1.1 INSTANTANEOUS FREQUENCY AND THE PROBLEM OF INTERFERENCE REJECTION

The mathematical operations leading to the unambiguous formulation of the exceedingly useful concept of instantaneous frequency have been the subject of much discussion. Elaborate mathematics, using, among other means, Fourier and Hilbert integral transforms, has been harnessed for the purpose. But it is significant that in almost all of the publications of those who have used it effectively in physics and in electrical communication (such as Helmholtz, Rayleigh, Carson, Van der Pol, Armstrong, and others) the characteristic features used in introducing and utilizing the concept have almost invariably been simplicity and straightforwardness. This concept is now so well appreciated that it needs no special clarification, but it seems pertinent to start our discussion by a statement of how it is mathematically described in most practical problems.

The two most significant (and useful) ways of introducing the concept of instantaneous frequency follow. The first is best stated in the form:

a. If the real time function $f(t)$ that describes the signal or the vibration, is reducible to the forms $A(t) \cos \phi(t)$ or $A(t) \sin \phi(t)$, both of which are clearly included in the complex function

$$F(t) = A(t)\, e^{j\phi(t)} \tag{1}$$

where $A(t)$ and $\phi(t)$ are real functions of time (and $f(t)$ is the real or imaginary part of $F(t)$), and furthermore, if $A(t)$ contains none of the zeros of $f(t)$, then $\phi(t)$ is by definition the "instantaneous phase angle" of $f(t)$, and $(d/dt)\ \phi(t)$ is by definition the "instantaneous frequency." The amplitude function $A(t)$ is the "instantaneous amplitude."

This definition is unique and unambiguous in almost all practical situations in sinusoidal-carrier modulation. For, in most cases, $A(t)$ is bounded and usually not called upon to contribute to the zeros of the signal; the unmodulated carrier frequency is usually much larger than the extent of the significant spectra of the modulating functions in $A(t)$ and $\phi'(t)$; and the extent of the frequency swings about the mean unmodulated carrier frequency is usually a small fraction of that frequency.

The second definition essentially counts the density of zero crossings per unit interval of time. In a period of $2\pi/\omega_o$ seconds, for instance, the sinusoidal signal $\cos \omega_o t$ has two zeros. Therefore, in every second, this sinusoid has ω_o/π zeros, and the (angular) frequency can be said to be equivalent to the number of zeros in a time interval of π seconds. When these notions are extended to the case of a time function f(t), the definition (3) becomes:

b. The instantaneous frequency of f(t) is defined at the time t as the ratio of the number of zeros of f(t) in the interval of time between $t - \tau/2$ and $t + \tau/2$ to τ/π, or as the mean density of zero crossings averaged over τ/π seconds.

The two definitions yield the same result for an ordinary sinusoidal-carrier frequency-modulated signal, but the first one is the more common and it will be applied in our computations. Stumpers found the second definition more suitable for use in the analysis of frequency-modulation noise at arbitrary levels.

Most of the signals that will be analyzed in our study consist of a superposition of several sinusoidally varying time functions that have different frequencies and amplitudes. The quickest, as well as the most elegant, way of achieving the reduction of the sum to the form indicated in the first statement of the definition of instantaneous frequency follows.

1. Replace each sinusoidal component of amplitude $A_n(t)$ and phase angle $\phi_n(t)$ by the corresponding complex function indicated in Eq. 1, with the understanding that only the real or the imaginary part of this function is the quantity of physical significance.

2. Represent each complex function $F_n(t) = A_n(t) \exp[j\phi_n(t)]$ thus obtained by a directed rotating line (henceforth called "phasor") in an Argand diagram, using an arbitrary reference axis (labeled the "axis of reals") for the measurement of the phase angle $\phi_n(t)$. The rotation of the phasors is conventionally positive if it is counterclockwise.

3. Add the representative component phasors vectorially to obtain their resultant. The amplitude and phase functions of this resultant will then be those of the resultant signal.

Analytically, the addition indicated in step 3 leads to

$$F(t) = \sum_{n=1}^{k} A_n(t) \exp[j\phi_n(t)]$$

$$= A(t) \exp[j\phi(t)]$$

where

$$A(t) = \sqrt{[\mathrm{Re}\ F(t)]^2 + [\mathrm{Im}\ F(t)]^2}$$

$$\phi(t) = \mathrm{Im}[\ln F(t)]$$

Let us now apply these concepts to the case of two-signal interference (1, 2). Consider that two carriers of relative strengths 1 and a (where $a < 1$) and of frequencies p and p + r rad/sec fall within the linear passband of a frequency-modulation receiver. The signals are supposed to be unmodulated in amplitude or frequency or, at worst, to have a frequency modulation that is so slow relative to the frequency difference r that the signal frequencies are not appreciably changed during a period of $2\pi/r$ sec.

At the input to the first limiter stage, the resultant signal will be, if time is counted from the instant at which the two signals are momentarily in phase,

$$f(t) = \cos pt + a \cos(p+r)t$$

The corresponding complex function of time is

$$F(t) = e^{jpt} + ae^{j(p+r)t} = e^{jpt}[1 + ae^{jrt}]$$

Figure 1 is a phasor diagram representation of the linear superposition of the carriers. The instantaneous phase of the resultant is $\Phi = pt + \theta$, therefore the instantaneous frequency of the resultant signal is $d\Phi/dt = p + d\theta/dt$. Clearly, $d\theta/dt$ represents the instantaneous deviation of the frequency of the resultant signal from that of the stronger signal. In essence, the most important step toward achieving interference rejection is to make the instantaneous-frequency deviation of the resultant signal from the desired frequency p average out to zero, over one period of the frequency difference r, at every point in the receiver prior to FM-to-AM conversion. This process must then be such that the average direct voltage level at the output of the discriminator corresponds to that dictated by the desired frequency p. If r is beyond the audible range, then the preceding requirements are necessary and sufficient, since the interference will not pass through the de-emphasis circuit and audio filter. If, however, r is audible, then those requirements (though necessary) will not ensure complete rejection of the interference, although it can be shown that by special design, and with the help of the de-emphasis circuit and the audio filter, the disturbance that can get through can be greatly reduced, if not effectively eliminated. This question will be taken up in greater detail in Section II.

From Fig. 1 we have

$$d\theta/dt = d/dt \; \mathrm{Im}[\ln(1 + ae^{jrt})]$$

$$= \mathrm{Re}\left[\frac{rae^{jrt}}{1 + ae^{jrt}}\right]$$

or

$$d\theta/dt = r\frac{a \cos rt + a^2}{1 + 2a \cos rt + a^2} \tag{2}$$

A plot of $d\theta/dt$ versus t is shown in Fig. 2 for $a = 0.8$.

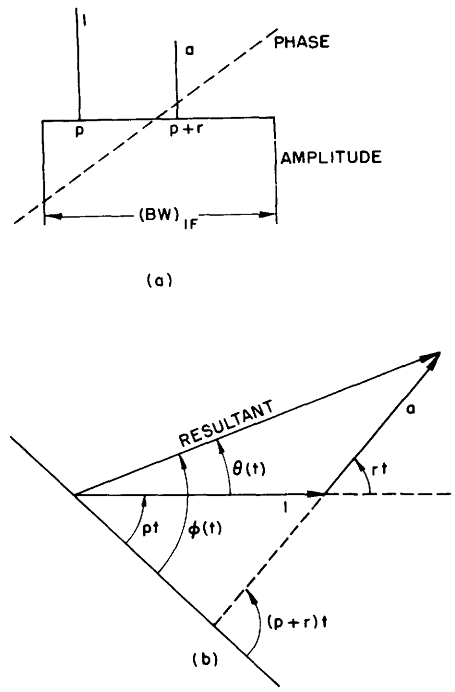

Fig. 1. Two-carrier interference: (a) resultant spectrum within the idealized i-f passband; (b) superposition of representative phasors.

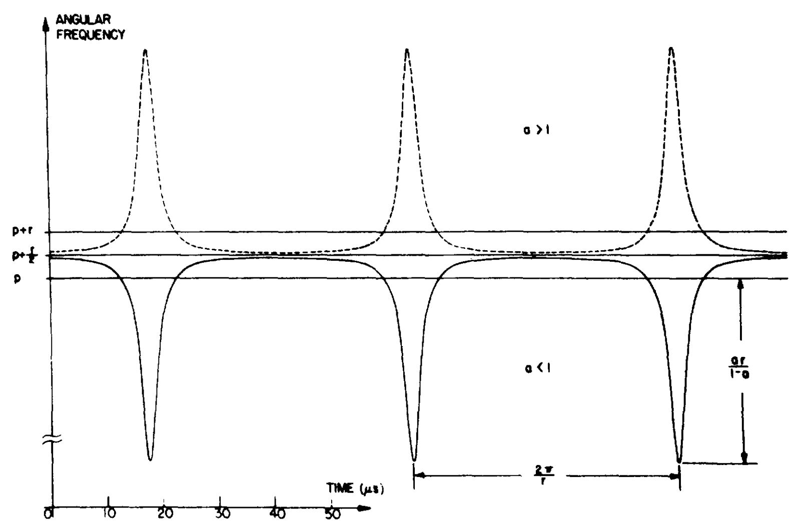

Fig. 2. Instantaneous-frequency disturbance caused by the interference, plotted for a = 0.8 and a = 1.25.

From Fig. 2 we find that the instantaneous-frequency deviation caused by the presence of the weaker signal is of such a value that the frequency of the resultant signal lingers near the average of the two carrier frequencies, $p + r/2$, during a large fraction of the frequency-difference cycle, attaining a maximum of $p + ar/(1+a)$, and then dips to a sharp minimum of $p - ar/(1-a)$ at $t = \pi/r$. This cycle of instantaneous variation recurs $r/2\pi$ times per second. Over one complete cycle, the average phase angle of the resultant signal is exactly the phase angle of the stronger signal, no net phase change being acquired from the instantaneous deviations in frequency. This means that the areas enclosed by the instantaneous-frequency deviation curve, above and below the frequency p, are exactly equal. Thus the average frequency of the resultant signal, over one period of the frequency difference r, is exactly the frequency of the stronger signal.

In addition to the instantaneous deviations in frequency, the interference also causes instantaneous-amplitude variations, with a ratio of maximum to minimum amplitude of $(1+a)/(1-a)$. The instantaneous-amplitude and -frequency variations of the resultant signal arise simultaneously, and, as long as no nonlinearities in response are encountered, the resultant signal will still be the result of a linear superposition of two signals, and the spectrum of the resultant will continue to be the sum of the spectra of the component signals. This will be true throughout the linear stages of the receiver, up to the first-limiter stage, and the passband need not exceed the frequency range in which the desired signal may be expected to fall.

However, when the resultant is passed through the limiter, the amplitude variations are completely eliminated, leaving behind the large excursions in instantaneous frequency. The spectrum, after limiting, is spread out with an "infinite" number of components on both sides of the frequency p of the stronger signal (and of harmonics of p). Thus, it becomes necessary to re-examine the bandwidth requirements after limiting, so that the average frequency of the signal at the input of the discriminator will still be the frequency of the stronger signal, as is required for the capture of this signal. The specification of the discriminator bandwidth should also be studied in relation to its possible dependence upon the value of the limiter bandwidth. It is with these questions that we are now chiefly concerned.

The work of Arguimbau and Granlund (1, 2) has indicated that interference, with arbitrary values of a in the range $0 < a < 1$, can be suppressed at the output if the receiver is designed with the following characteristics:

(a) In the linear sections, the stages preceding the limiter-discriminator section, the bandwidth should be sufficient to accommodate the desired stronger signal over the whole range of its frequency variations. Furthermore, these linear stages must have a constant gain over the whole passband to preserve the relative magnitudes of the signals that are passed; this gain should fall very steeply at the skirts to effect essentially complete rejection outside the passband and secure excellent selectivity.

(b) Since a frequency-modulation receiver should be completely insensitive to amplitude changes, the linear stages should be followed by a perfect rapid-acting limiter to cope with amplitude ratios of the order of $(1+a)/(1-a)$, (or 39/1 for $a = 0.95$) that may recur at a maximum rate equivalent to the intermediate-frequency (i-f) bandwidth in cycles/sec. If a capture ratio $\underline{a}$ (strength of weaker signal relative to the desired stronger signal) is desired, it is clear that the linear stages must provide enough gain to raise the value of the minimum amplitude $(1-a)x$ (expected minimum signal strength) to the level necessary to drive the limiter. The discriminator section should also be sufficiently rapid-acting to handle the sharp changes in instantaneous frequency (that may recur at a maximum rate equivalent to the i-f bandwidth in cycles/sec) and still preserve the average output dc level at the value dictated by the frequency p.

(c) For the requirements in the bandwidths of the limiter and the discriminator sections, Arguimbau and Granlund indicated that interference rejection will be fully achieved (with arbitrary values of $\underline{a}$) if the interference frequency spikes are fully accommodated within a passband in the nonlinear sections. If account is taken of the situation in which the stronger signal will have the higher frequency, then, from Fig. 2, the bandwidth required to accommodate the spikes is given by $[(1+a)/(1-a)]\,(BW)_{if}$, when r is assigned its maximum value of one i-f bandwidth, $(BW)_{if}$. A plot of the required bandwidth as a function of $\underline{a}$, calculated from $[(1+a)/(1-a)]\,(BW)_{if}$, is presented in Fig. 16.

Thus, it was thought that the key to interference rejection (1, 2) lay in the fast action of limiter and discriminator (to avoid diagonal clipping), and in the full accommodation of the instantaneous-frequency excursions within limiter and linear discriminator passbands (to preserve the equality of the areas enclosed by the $(d\theta/dt)$-curve above and below the frequency p). The physical basis for this argument can be traced to the behavior of networks that involve energy-storage elements when they are excited by variable-frequency sources. The response of such networks will follow a variable-frequency excitation, through essentially stationary states, provided the bandwidth is much larger than the rate at which the excitation frequency is varied; still better, provided the static amplitude-response characteristic is essentially a constant, or a linear function of frequency, over the whole range of the instantaneous-frequency excursions of the excitation. Under such conditions, the dynamic response is readily evaluated from the static characteristics on an instantaneous-frequency basis.

It was assumed, however, that if the limiter bandwidth was chosen equal to $[(1+a)/(1-a)]\,(BW)_{if}$, the quasi-static argument applied, and a linearity over the same range would be necessary in the discriminator characteristic. (This assumption will be shown to be invalid in section 1.6.) With a linear FM-to-AM conversion characteristic for the discriminator over the range of the spikes (thus extending over a bandwidth that is much larger than the spike repetition rate for values of $a > 0.8$) and with sufficiently low associated low-frequency circuit time constants, we can plot the instantaneous detected output as a function of the instantaneous frequency on a static basis (in the same way in which we handle the static tube characteristics in low-frequency electronic circuit problems). However, if we deal with a relatively narrow-band discriminator, we have no assurance that we can plot the instantaneous detected output as a function of instantaneous frequency because the narrow-band detector is likely to be too sluggish to follow the rapid spike variations and, thus, the quasi-static reasoning is likely to break down.

It becomes important to determine whether or not the bandwidth given by $[(1+a)/(1-a)]r$ is a necessary requirement in the nonlinear sections. This is contingent upon the over-all role played by the limiter bandwidth. Granlund (1) performed a Fourier analysis of the resultant of two carriers after limiting with the intention of "determining whether [or not the bandwidth specified by the extent of the spikes] is a reasonable estimate of the extent of the spectrum after limiting. Thus the result was to be used as a guide in determining limiter and discriminator bandwidths." A good portion of our treatment in section 1.2 will parallel Granlund's analysis, and some of his results (particularly the tables of spectral amplitudes) will be repeated here for the sake of completeness.

Finally, aside from being of theoretical interest, the question of whether or not "wideband" limiting and detecting is necessary has important practical and economic implications in communication by frequency modulation, and in frequency-modulation receiver design. Some of the more obvious considerations are:

a. Wideband discriminators are more expensive to construct than the narrow-band types. This is also true of limiters.

b. Wideband discriminators require critical adjustments that become more unreliable in time and with changes of ambient temperature and humidity.

c. Wideband discriminators are considerably less efficient FM-to-AM converters than are the narrow-band types, and this can have detrimental effects upon the quality of reception at the low-modulation levels.

d. A narrow-band limiter yields a stronger signal at its output than does a wideband limiter. Furthermore, the fact that the audio-signal level is higher at the output of a narrow-band discriminator than at the output of a wideband discriminator decreases the demand on the number of audio stages that are necessary to bring the signal strength up to the desired level at the loud-speaker.

e. In video applications of frequency modulation, widebanding demands prohibitive bandwidths to effect a reasonable degree of interference rejection.

I.2 THE TWO-PATH INTERFERENCE SPECTRUM AFTER LIMITING

Consider two frequency-modulated carriers of relative constant amplitudes 1 and $\underline{a}$, where $a < 1$, that have such frequencies that they fall simultaneously within the ideal intermediate-frequency passband of the receiver. For simplicity, assume that the modulation is so slow that the frequencies of the modulated carriers (henceforth called signals) do not change appreciably during several cycles of the frequency difference. Let the instantaneous frequencies be momentarily p and p + r rad/sec, the former being that of the stronger signal. Consider the resultant wave to be passed through an ideal limiter that is followed by an ideal wideband filter. A simple analysis shows that the structure of the unfiltered amplitude-limited resultant signal includes a fundamental carrier frequency of p rad/sec with associated sidebands, plus other carriers at harmonic frequencies of p (only odd harmonics with symmetrical limiting) with associated sidebands. The wideband filter will thus be assumed to be sufficiently selective that only the spectral components centered about the frequency p are of significance, with $r \ll p$, and with the harmonics of p and their associated sidebands completely rejected or negligible. Thus, with the input (to the ideal selective limiter) described by $A(t) \cos \phi(t)$, the output signal will be

$$e(t) = \cos \Phi(t) = \cos(pt + \theta) \tag{3a}$$

where $\Phi(t)$ and θ are as shown in Fig. 1.

The assumptions strip the problem of unnecessary computational complexities and make it easier to "see the forest for the trees." In the light of standard FM practice, it is readily appreciated that the assumptions correspond rather well with most important practical situations. Furthermore, the assumption of a modulation that is slow in comparison with the frequency difference is realistic, since, with wideband FM, the maximum allowable frequency deviation is often much larger than the audio frequencies that are of importance, and so the frequency difference r rad/sec will be supersonic most of the time. In a later discussion, the problem in which the frequency difference is within the audio range will be given special attention and the assumption regarding the relative frequencies of the modulation and the frequency difference will be reconsidered.

In Eq. 3a, we note that if we expand the cosine of the sum we get

$$e(t) = \cos pt \cos \theta - \sin pt \sin \theta \tag{3b}$$

From Fig. 1 we note that, with

$$g(t) = (1 + 2a \cos rt + a^2)^{-1/2}$$

$$\cos \theta = g(t) \cdot (1 + a \cos rt) \quad \text{and} \quad \sin \theta = g(t) \cdot a \sin rt$$

so that

$$e(t) = g(t) \left[\cos pt + a(\cos pt \cos rt - \sin pt \sin rt)\right]$$

or

$$e(t) = g(t) \left[\cos pt + a \cos(p+r)t\right] \tag{4}$$

Equation 4 could have been written directly by normalizing the instantaneous phasor-amplitude scales in Fig. 1 by dividing by $(1 + 2a \cos rt + a^2)^{1/2}$. This shows that the resultant constant-amplitude signal at the output of the ideal limiter can be expressed as the sum of two amplitude-modulated waves with the same carrier frequencies and the same instantaneous relative amplitudes as the two input signals. The resultant

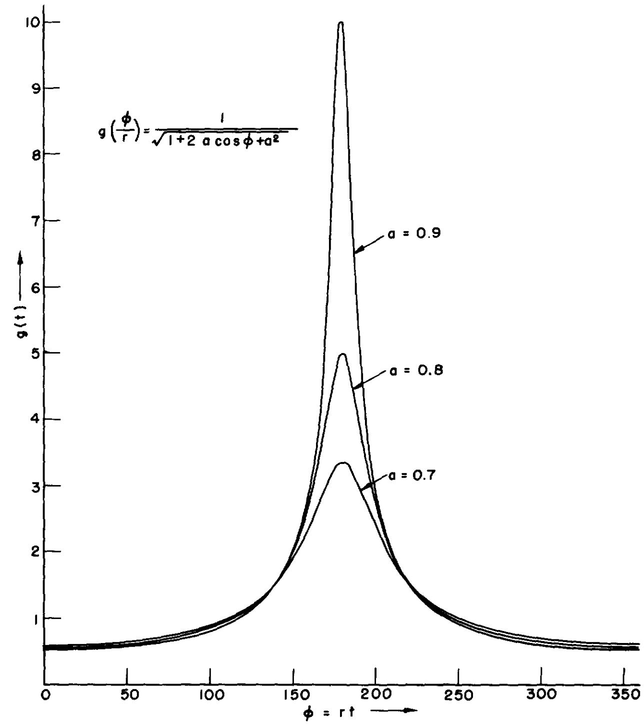

Fig. 3. The amplitude-modulation function g(t) introduced by amplitude limiting the resultant of the two input signals.

amplitude at any instant remains, of course, constant. The amplitude-limiting process can, therefore, be interpreted as amplitude modulation of the resultant signal by a function that is given by the reciprocal of its instantaneous amplitude. Plots of the amplitude-modulation function g(t) appear in Fig. 3.

Next, if we note that the amplitude-modulation function

$$g(t) = (1 + 2a \cos rt + a^2)^{-1/2}$$

is an even periodic function of $\phi = rt$, we can write

$$g(t) = \sum_{n=0}^{\infty} a_n \cos n\phi$$

where

$$a_o = 1/\pi \int_0^{\pi} g(\phi/r)\, d\phi = 1/\pi\ G_o(a)$$

and

$$\alpha_n = 2/\pi \int_0^\pi g(\phi/r) \cos n\phi \, d\phi$$

$$= (2/\pi) \int_0^\pi \frac{\cos n\phi}{(1 + 2a \cos \phi + a^2)^{1/2}} \, d\phi$$

or

$$\alpha_n = (2/\pi) \, G_n(a)$$

where

$$G_n(a) = \int_0^\pi \frac{\cos n\phi}{(1 + 2a \cos \phi + a^2)^{1/2}} \, d\phi \tag{5}$$

Thus

$$g(t) = (1 + 2a \cos rt + a^2)^{-1/2}$$

$$= (1/\pi) \, G_o(a) + (2/\pi) \sum_{n=1}^{\infty} G_n(a) \cos nrt$$

Substitution in Eq. 4 yields, after some trigonometric manipulations,

$$e(t) = (1/\pi) \, [G_o(a) + aG_1(a)] \cos pt$$

$$+ (1/\pi) \sum_{n=1}^{\infty} [G_n(a) + aG_{n-1}(a)] \cos(p + nr)t$$

$$+ (1/\pi) \sum_{n=1}^{\infty} [G_n(a) + aG_{n+1}(a)] \cos(p - nr)t$$

which can be expressed in the final form

$$e(t) = \sum_{n=-\infty}^{\infty} A_n \cos(p - nr)t$$

$$= \text{Re} \left[e^{jpt} \sum_{n=-\infty}^{\infty} A_n \, e^{-jnrt} \right] \tag{6}$$

with the definitions

$$\left.\begin{aligned} A_o &= (1/\pi)\,[G_o(a) + aG_1(a)] \\ A_{-n} &= (1/\pi)\,[G_n(a) + aG_{n-1}(a)] \\ A_n &= (1/\pi)\,[G_n(a) + aG_{n+1}(a)] \end{aligned}\right\} \tag{7}$$

The auxiliary function $G_n(a)$ is readily recognized as an elliptic integral. A fruitful analysis of the function $G_n(a)$, which is well known in celestial mechanics (4), has been cited by Granlund (1). For completeness, some of the steps involved in this analysis are outlined and the parts that are important for our purposes are presented and expanded.

First, we note that $G_n(a)$ can be expressed in the form

$$G_n(a) = (1/2)\ \mathrm{Re} \int_{-\pi}^{\pi} \frac{e^{jn\phi}\, d\phi}{(1 + 2a\cos\phi + a^2)^{1/2}}$$

$$= (1/2) \int_{-\pi}^{\pi} \frac{e^{jn\phi}\, d\phi}{(1 + 2a\cos\phi + a^2)^{1/2}}$$

since the contribution from the odd imaginary part vanishes. With $z = e^{j\phi}$, $G_n(a)$ reduces to

$$G_n(a) = (1/2j) \oint \frac{z^{n-1/2}}{[(1 + az)(a+z)]^{1/2}}\, dz$$

wherein the path of integration is a complete circuit of the unit circle in the z-plane. By a straightforward contour integration, we get

$$G_n(a) = (-1)^n \int_0^a \frac{x^{n-1/2}}{[(1 - ax)(a-x)]^{1/2}}\, dx \tag{8}$$

where x is a dummy variable of integration.

Finally, the substitution of $x = a \sin^2\theta$ yields $G_n(a)$ in the form

$$G_n(a) = 2(-a)^n \int_0^{\pi/2} \frac{\sin^{2n}\theta\, d\theta}{(1 - a^2 \sin^2\theta)^{1/2}} \tag{9a}$$

or

$$G_n(a) = (-a)^n \int_0^{\pi} \frac{\sin^{2n}\theta\, d\theta}{(1 - a^2 \sin^2\theta)^{1/2}} \tag{9b}$$

The last integral on the right is given in references 5 and 8 and can be expressed as

$$I_n = \int_0^{\pi} \frac{\sin^{2n}\theta\, d\theta}{(1 - a^2 \sin^2\theta)^{1/2}}$$

$$= \frac{(1;2;n)}{(2;2;n)}\, \pi \sum_{k=0}^{\infty} \frac{(1;2;k)\left([2n+1];2;k\right)}{(2;2;k)\left([2n+2];2;k\right)} a^{2k} \qquad a^2 < 1 \tag{10a}$$

in which we have used the notation (7)

$$(m;d;v) = m[m+d][m+2d]\ldots[m+(v-1)d]$$

$$= \frac{d^v \Gamma\left(\frac{m}{d} + v\right)}{\Gamma\left(\frac{m}{d}\right)} \qquad v = 1, 2, \ldots \tag{10b}$$

When simplified, the expression for I_n becomes

$$I_n = \sum_{k=0}^{\infty} \frac{\Gamma(k + 1/2)\, \Gamma(k + n + 1/2)}{\Gamma(k+1)\, \Gamma(k + n + 1)} a^{2k} \tag{10c}$$

Substitution in Eq. 9b yields

$$G_n(a) = (-a)^n \sum_{k=0}^{\infty} C_1(k, n)\, a^{2k} \tag{11a}$$

where

$$C_1(k, n) \equiv \frac{\Gamma(k + 1/2)\, \Gamma(k + n + 1/2)}{\Gamma(k+1)\, \Gamma(k + n + 1)} \tag{11b}$$

Similar expressions can be found for $A_n(a)$ and $A_{-n}(a)$ by substituting from Eqs. 9 in Eqs. 7. This yields the equations

$$A_n(a) = (2/\pi)(-a)^n \int_0^{\pi/2} \sin^{2n}\theta\, (1 - a^2 \sin^2\theta)^{1/2}\, d\theta \tag{12}$$

$$A_{-n}(a) = (2/\pi)(-a)^n \left\{ \int_0^{\pi/2} \frac{\sin^{2n}\theta\, d\theta}{(1 - a^2 \sin^2\theta)^{1/2}} - \int_0^{\pi/2} \frac{\sin^{2(n-1)}\theta}{(1 - a^2 \sin^2\theta)^{1/2}}\, d\theta \right\}$$

$$A_{-n}(a) = (1/\pi)(-a)^n \left[I_n - I_{n-1}\right] \tag{13}$$

where I_n is defined in Eq. 10a. The integral

$$J_n = 2\int_0^{\pi/2} \sin^{2n}\theta\, (1 - a^2 \sin^2\theta)^{1/2}\, d\theta$$

can be easily evaluated from another integral (8) and the result can be reduced to the form

$$J_n = -\sum_{k=0}^{\infty} \frac{1}{2} \cdot \frac{\Gamma(k - 1/2)\,\Gamma(k + n + 1/2)}{\Gamma(k+1)\,\Gamma(k + n + 1)}\, a^{2k} \qquad a^2 < 1$$

Compare this expression with expression 10c for I_n. Substitution in Eq. 12 leads to

$$A_n(a) = (-1/2\pi)(-a)^n \sum_{k=0}^{\infty} C_2(k, n)\, a^{2k} \tag{14a}$$

where

$$C_2(k, n) \equiv \frac{\Gamma(k - 1/2)\,\Gamma(k + n + 1/2)}{\Gamma(k+1)\,\Gamma(k + n + 1)} \tag{14b}$$

The value of expressions 11, 13, and 14 in the numerical evaluation of $G_n(a)$, $A_{-n}(a)$, and $A_n(a)$ is best brought out by studying the convergence properties of the infinite series that are involved, and by safely estimating the necessary number of terms that is required in each summation to meet a certain prescribed tolerance in the computed values of the desired functions. The details of this study will not be presented here. Only steps and results are outlined. In this study, Stirling's asymptotic formula for the gamma function is used to simplify the expressions for $C_1(k, n)$ and $C_2(k, n)$ in Eqs. 11b and 14b. It follows immediately that

$$C_1(k, n) < [k(k+n)]^{-1/2} \cdot e < e/k \tag{15a}$$

and

$$C_2(k, n) < [k^3(k+n)]^{-1/2}\, e^2 < e^2/k^2 \tag{15b}$$

for all positive integral n, e being the base of natural logarithms. The series in Eq. 14 is thus seen to converge much more rapidly than that in Eq. 11, the latter converging, in fact, only for $a < 1$, which is the only range of significance in our discussions. The number of terms, N, (estimated conservatively by a rough estimate of the remainder) that must be added to meet a prescribed tolerance, ϵ, in the computed value of the series, can be obtained from the formulas

$$\epsilon = \frac{a^{2N+n}}{[N(N+n)]^{1/2}} \cdot \frac{a^2}{1 - a^2} \quad \text{(for each sum in Eq. 13)}$$

$$\epsilon = \frac{a^{2N+n}}{[N^3(N+n)]^{1/2}} \cdot \frac{a^2}{1 - a^2} \quad \text{(for Eq. 14)}$$

In each case, the error ϵ is about $(1 - a^2)^{-1}$ multiplied by the first neglected term in the sum. Estimates for N, computed from these expressions for various prescribed

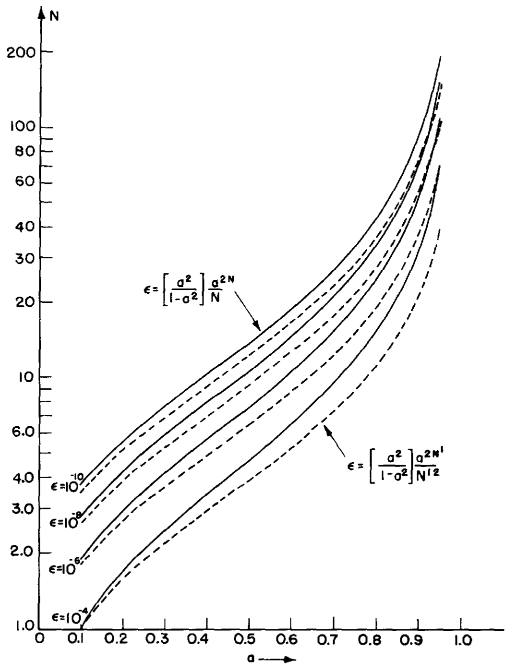

Fig. 4. Conservative estimates of the number of terms N and N' that is needed in series 11 and 14 to meet the prescribed tolerance ϵ on finite approximants.

tolerances ϵ, are plotted against a in Fig. 4 for the worst possible situation, namely, that with $n = 0$. The computation of the coefficients C_1 and C_2 is greatly facilitated by the availability of excellent tables (9) for log $\Gamma(x)$ with x ranging over large values. Admittedly, some of the estimates shown in Fig. 4 are not encouraging, cognizant as we are of the high degree of safety insured by these estimates.

For an alternative approach to the evaluation of $G_n(a)$, $A_n(a)$, and $A_{-n}(a)$, we go back to Eq. 8, and with the substitution $x = au^2$ we obtain

$$G_n(a) = 2(-a)^n \int_0^1 \frac{u^{2n}\,du}{[(1 - a^2u^2)(1 - u^2)]^{1/2}} \tag{16}$$

The elliptic integral on the right is of the general type (6) represented by

$$I_k = \int \frac{u^k}{[R(u)]^{1/2}}\,du$$

for which a recursion formula can be found in the following way. First, the expression for $d/du\,[u^{k-1}\sqrt{R(u)}]$ is formed, and then both sides are integrated between the limits 0 and 1. The result for $G_n(a)$ is given by (see also ref. 1)

$$G_{n+1}(a) + \frac{2n}{2n+1} \cdot \frac{(1+a^2)}{a} G_n(a) + \frac{2n-1}{2n+1} G_{n-1}(a) = 0 \tag{17}$$

for $n > 1/2$.

The restriction on n is inconsequential, since, from Eq. 9a, we have

$$G_o(a) = 2 \int_0^{\pi/2} \frac{d\theta}{(1 - a^2 \sin^2 \theta)^{1/2}}$$

$$= 2\, K(a) \tag{18}$$

and

$$G_1(a) = (2/a) \left[\int_0^{\pi/2} (1 - a^2 \sin \theta)^{1/2}\, d\theta - \int_0^{\pi/2} \frac{d\theta}{(1 - a^2 \sin^2 \theta)^{1/2}} \right]$$

$$= (2/a)\, [E(a) - K(a)] \tag{19}$$

where K(a) and E(a) are the complete elliptic integrals of the first and second kind.

Equations 17, 18, and 19, with the help of a table of complete elliptic integrals, reduce the task of computing $G_n(a)$, for any integer n, to a fairly systematic procedure. Granlund (1) used ten-place tables of complete elliptic integrals to evaluate $G_o(a)$ and $G_1(a)$, as given by Eqs. 18 and 19 for several values of a. These, then, together with the recursion formula, Eq. 17, and the expressions for the spectral-component amplitudes given by Eqs. 7, were used to calculate and tabulate these amplitudes up to reasonably large values of n. Granlund's table, which has been expanded to include the values for a = 0.85, is presented as our Table I.

Equation 17 is readily recognized as a linear difference equation with variable coefficients. The task of developing the general expression for $G_n(a)$ by solving this difference equation directly is not pleasant. However, for large values of n, the coefficients become approximately constant, and the solution to the resulting constant-coefficient difference equation shows that $G_n(a)$ is asymptotically approximated by a constant multiplied by a^n. It can also be shown (using recursion formulas for A_n and A_{-n} derived with the help of Eqs. 17 and 7) that $A_n(a)$ and $A_{-n}(a)$ tend asymptotically to expressions of the form (constant) $\times\, a^n$.

It is convenient at this point to summarize the important properties displayed by the side-frequency components with amplitudes A_n and A_{-n}. It is noted that the spectral component that has the frequency of the stronger of the two input signals is A_o. The component A_{-1} has the frequency of the weaker signal. The amplitudes of the side-frequency components are not symmetrically distributed about the center-frequency component A_o. This lack of symmetry conforms to our physical expectations. For, on an instantaneous-frequency basis, the instantaneous frequency of the resultant signal (as shown in Fig. 2) places this signal on one side of the center frequency much longer than it does on the other. This means that the power in the composite signal

Table I. Two-signal Interference Spectrum at the Output of a Limiter.

$a = 0.2$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.17373 56950	0.98992 37219	
1	-0.31899 30122	-0.09848 72762	0.10050 76601
2	0.04793 05440	0.01474 77722	-0.00505 09599
3	-0.00799 52557	-0.00245 58489	0.00050 63843
4	0.00139 98939	0.00042 95516	-0.00006 33937
5	-0.00025 20885		0.00000 88778

$\sum A_n = 1.00006\ 64892$

$a = 0.3$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.21609 72398	0.97710 53305	
1	-0.48810 10335	-0.14483 86397	0.15174 72794
2	0.11025 67567	0.03245 83936	-0.01151 43997
3	-0.02761 90193	-0.00809 82530	0.00173 73378
4	0.00725 86910	0.00249 78153	-0.00032 69090
5	-0.00196 14240		0.00006 88133

$\sum A_n = 1.00083\ 67685$

$a = 0.4$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.27999 97318	0.95871 22063	
1	-0.67029 12677	-0.18756 91886	0.20426 21993
2	0.20256 32068	0.05584 88876	-0.02086 62636
3	-0.06777 18792	-0.01854 55201	0.00421 86894
4	0.02377 35234	0.00647 60031	-0.00106 16361
5	-0.00857 13987		0.00029 85781

$\sum A_n = 1.00177\ 39554$

$a = 0.5$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.37150 07096	0.93421 54575	
1	-0.87315 25820	-0.22518 58510	0.25865 79046
2	0.33142 07333	0.08337 99297	-0.03347 20536
3	-0.13894 99174	-0.03451 73819	0.00851 81156
4	0.06102 07278	0.01504 20514	-0.00269 10653
5	-0.02752 94594	-0.00675 10369	0.00094 88514
6	0.01264 09032	0.00308 83458	-0.00035 77251
7	-0.00587 71571	-0.00143 18211	0.00014 11050
8	0.00275 79172	0.00067 04154	-0.00005 75063
9	-0.00130 34901	-0.00031 63008	0.00002 40224
10	0.00061 96033	0.00015 01323	-0.00001 02311
11	-0.00029 58978	-0.00007 16100	0.00000 44257
12	0.00014 18569	0.00003 42952	-0.00000 19391
13	-0.00006 82306	-0.00001 64807	0.00000 08588
14	0.00003 29099	0.00000 79429	-0.00000 03837
15	-0.00001 59127	-0.00000 38378	0.00000 01726
16	0.00000 77118		-0.00000 00779

$\sum A_n = 0.99999\ 87240$

$a = 0.6$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.50150 76058	0.90277 99276	
1	-1.10890 13617	-0.25585 71623	0.31576 44262
2	0.50850 39668	0.11282 68940	-0.04992 27199
3	-0.25674 63756	-0.05582 63749	0.01539 21943
4	0.13560 44105	0.02912 17643	-0.00587 07213
5	-0.07352 61496	-0.01565 78319	0.00249 44344
6	0.04055 93664	0.00858 49521	-0.00113 20129
7	-0.02264 82400	-0.00477 17658	0.00053 71097
8	0.01276 21591	0.00267 91634	-0.00026 31738
9	-0.00724 21982	-0.00151 59387	0.00013 21296
10	0.00413 28938		-0.00006 76170

$\sum A_n = 0.99942\ 76771$

$a = 0.7$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.69138 79968	0.86304 06833	
1	-1.40009 38939	-0.27713 89930	0.37683 99771
2	0.75633 72415	0.14095 48384	-0.07121 49883
3	-0.44787 79380	-0.07927 26736	0.02596 07594
4	0.27690 78411	0.04902 08188	-0.01165 22794
5	-0.17557 77096	-0.03066 67364	0.00581 16316
6	0.11319 33084	0.01957 76500	-0.00309 11354
7	-0.07384 04385	-0.01267 62485	0.00171 72428
8	0.04859 54706	0.00829 35002	-0.00098 44804
9	-0.03220 09592	-0.00546 94786	0.00057 80094
10	0.02145 44079	0.00362 98539	-0.00034 57684
11	-0.01435 84080	-0.00242 12943	0.00020 99819
12	0.00964 52682	0.00162 19453	-0.00012 91120
13	-0.00649 96812	-0.00109 03403	0.00008 02162
14	0.00439 18231	0.00073 51859	-0.00005 02782
15	-0.00297 45264		0.00003 17513

$\sum A_n = 1.00190\ 02387$

$a = 0.8$

n	$G_n(a)$	A_{+n}	A_{-n}
0	3.99060 55554	0.81254 96100	
1	-1.79738 20863	-0.28533 48346	0.44407 48727
2	1.12622 03322	0.16277 17893	-0.09921 25241
3	-0.76857 20932	-0.10559 41474	0.04214 55571
4	0.54604 78694	0.07265 55225	-0.02190 28413
5	-0.39724 22671	-0.05169 45793	0.01260 38073
6	0.29354 86956	0.03758 08077	-0.00771 74608
7	-0.21935 63826	-0.02773 17860	0.00492 82563
8	0.16529 30092	0.02068 95656	-0.00324 42452
9	-0.12536 85272	-0.01556 54878	0.00218 54775
10	0.09558 51314	0.01178 80316	-0.00149 91410
11	-0.07318 99224	-0.00897 50382	0.00104 34780
12	0.05624 25105	0.00686 34248	-0.00073 51136
13	-0.04335 05320	-0.00526 80037	0.00052 31348
14	0.03350 07627	0.00405 61305	-0.00037 54984
15	-0.02594 75660		0.00027 15324

$\sum A_n = 1.00188\ 02967$

a = 0.85

n	$G_n(a)$	A_{+n}	A_{-n}
0	4.21987 09352	0.78183 80350	
1	-2.07488 74214	-0.28261 30658	0.48128 54628
2	1.39650 85767	0.16880 36615	-0.11686 61160
3	-1.01905 43927	-0.11534 70926	0.05346 90254
4	0.77256 56635	0.08383 96327	-0.02980 35362
5	-0.59903 02230	-0.06311 69204	0.01835 07531
6	0.47146 18465	0.04859 90692	-0.01200 46891
7	-0.37503 92562	-0.03800 98203	0.00818 16187
8	0.30073 86870	0.03006 99857	-0.00574 38003
9	-0.24267 18129	-0.02399 71966	0.00412 40455
10	0.19680 28193	0.01928 27864	-0.00301 38286
11	-0.16026 37169	-0.01558 06083	0.00223 41151
12	0.13095 97558	0.01264 67479	-0.00167 57117
13	-0.10732 80301	-0.01030 45846	0.00126 93442
14	0.08818 26152	0.00842 34301	-0.00096 96389
15	-0.07261 13282	-0.00690 48833	0.00074 60849
16	0.05990 47029	0.00567 37914	-0.00057 77089
17	-0.04950 58370	-0.00467 20785	0.00044 98229
18	0.04097 41993	0.00385 44224	-0.00035 19750
19	-0.03395 90286	-0.00318 51594	0.00027 66243
20	0.02817 94766	0.00263 60282	-0.00021 82644

$\sum A_n = 1.00109\ 78085$

a = 0.9

n	$G_n(a)$	A_{+n}	A_{-n}
0	4.56109 82768	0.74592 55110	
1	-2.46411 57458	-0.27345 33054	0.52230 60032
2	1.78337 42781	0.16923 60803	-0.13825 14988
3	-1.39078 16128	-0.12081 07532	0.06819 95602
4	1.12360 38200	0.09211 80049	-0.04077 53791
5	-0.92689 61917	-0.07292 84983	0.02684 85624
6	0.77531 61744	0.05914 47989	-0.01874 53959
7	-0.65500 81206	-0.04877 38870	0.01361 61626
8	0.55753 38172	0.04071 56760	-0.01017 74784
9	-0.47735 74983	-0.03430 63100	0.00777 40624
10	0.41064 56077	0.02911 77756	-0.00604 02932
11	-0.35463 26866	-0.02485 97281	0.00475 82109
12	0.30725 94972	0.02132 71632	-0.00379 10455
13	-0.26695 35975	-0.01837 06010	0.00304 93928
14	0.23248 96137	0.01587 81488	-0.00247 28299
15	-0.20289 65978	-0.01376 42143	0.00201 93753
16	0.17739 44925	0.01196 21272	-0.00165 91729
17	-0.15534 92905	-0.01041 91580	0.00137 05637
18	0.13624 06003	0.00909 30583	-0.00113 75635
19	-0.11963 76835	-0.00794 95998	0.00094 81996
20	0.10518 14212	0.00696 07856	-0.00079 33855
21	-0.09257 05204	-0.00610 35244	0.00066 61458
22	0.08155 08144	0.00535 86332	-0.00056 10702
23	-0.07190 68574	-0.00471 00743	0.00047 39238
24	0.06345 52474	0.00414 43641	-0.00040 13647
25	-0.05603 92705		0.00034 07355

$\sum A_n = 1.00209\ 68939$

a = 0.95

n	$G_n(a)$	A_{+n}	A_{-n}
0	5.18002 24618	0.70201 44040	
1	-3.13113 59644	-0.25469 63862	0.56973 82095
2	2.45366 70194	0.16040 69563	-0.16581 14861
3	-2.05235 12672	-0.11794 68863	0.08869 14479
4	1.77032 65218	0.09325 43616	-0.05710 70797
5	-1.55511 50572	-0.07687 16475	0.04032 83151
6	1.38275 33202	0.06509 40585	-0.03011 40200
7	-1.24026 76887	-0.05616 02396	0.02334 73825
8	1.11982 64133	0.04911 98511	-0.01859 81753
9	-1.01632 82631	-0.04341 19769	0.01512 18935
10	0.92625 84374	0.03868 25148	-0.01249 47493
11	-0.84708 81401	-0.03469 59015	0.01045 88274
12	0.77693 44746	0.03128 87109	-0.00884 87788
13	-0.71435 58845	-0.02834 35952	0.00755 40877
14	0.65822 30042	0.02577 38383	-0.00649 83237
15	-0.60763 37928	-0.02351 37835	0.00562 71016
16	0.56185 58561	0.02151 26740	-0.00490 07777
17	-0.52028 61025	-0.01973 05577	0.00428 98499
18	0.48242 18182	0.01813 54894	-0.00377 19655
19	-0.44783 94714	-0.01670 15704	0.00332 99212
20	0.41617 88846	0.01540 75460	-0.00288 54109
21	-0.38713 12119	-0.01423 57818	0.00262 24687
22	0.36042 96679	0.01317 15044	-0.00233 79808
23	-0.33584 22805	-0.01220 22287	0.00208 99922
24	0.31316 61562	0.01131 73202	-0.00187 29386
25	-0.29222 28907	-0.01050 76560	0.00168 22543
26	0.27285 48587	0.00976 53593	-0.00151 41643
27	-0.25492 21870		0.00136 55267

$\sum A_n = 1.00541\ 78050$

a = 0.98

n	$G_n(a)$	A_{+n}	A_{-n}
0	6.04196 08912	0.66850 94960	
1	-4.02222 07846	-0.23527 98956	0.60443 89257
2	3.35006 85643	0.14679 01498	-0.18834 64439
3	-2.94787 11306	-0.10809 34049	0.10669 62204
4	2.66151 60040	0.08611 88178	-0.07238 29373
5	-2.43976 09782	-0.07184 73944	0.05364 30799
6	2.25923 03385	0.06176 86669	-0.04193 26866
7	-2.10732 48451	-0.05423 30535	0.03397 03133
8	1.97647 62068	0.04836 04982	-0.02823 47685
9	-1.86178 28785	-0.04363 96739	0.02392 53820
10	1.75988 24486	0.03975 01985	-0.02058 34363
11	-1.66837 09356	-0.03648 21314	0.01792 52596
12	1.58546 83056	0.03369 16872	-0.01576 75475
13	-1.50981 91312	-0.03127 69528	0.01398 64754
14	1.44036 70264	0.02916 35899	-0.01249 54845
15	-1.37627 23537	-0.02729 60398	0.01123 23067
16	1.31685 64444	0.02563 19025	-0.01015 10494
17	-1.26156 27018	-0.02413 82232	0.00921 71764
18	1.20992 88134	0.02278 89814	-0.00840 42195
19	-1.16156 64458	-0.02156 33357	0.00769 15737
20	1.11614 61517	0.02044 43850	-0.00706 29670
21	-1.07338 59407	-0.01941 82678	0.00650 53908
22	1.03304 25043	0.01847 34973	-0.00600 83275
23	-0.99490 43886	-0.01760 04697	0.00556 31861
24	0.95878 66145	0.01679 10861	-0.00516 28865
25	-0.92452 63897	-0.01603 84642	0.00480 15431
26	0.89197 96616	0.01533 67154	-0.00447 42275
27	-0.86101 83154	-0.01468 07672	0.00417 67837
28	0.83152 78828	0.01406 62248	-0.00390 56834
29	-0.80340 56473	-0.01348 92594	0.00365 79147
30	0.77655 90704	0.01294 65176	-0.00343 08915
31	-0.75090 44753	-0.01243 50481	0.00322 23827
32	0.72636 59382	0.01195 22426	-0.00303 04526
33	-0.70287 43475	-0.01149 57858	0.00285 34164
34	0.68036 66034	0.01106 36141	-0.00268 98004
35	-0.65878 49332	-0.01065 38820	0.00253 83107
36	0.63807 63017	0.01026 49329	-0.00239 78070
37	-0.61819 19019	-0.00989 51898	0.00226 72811
38	0.59908 69859	0.00954 34804	-0.00214 57518
39	-0.58071 96510	-0.00920 84944	0.00203 25981
40	0.56305 13395		-0.00192 70221

$\sum A_n = 0.99450\ 20805$

Note: See reference 1, pp. 43-45; values for a = 0.1 are omitted and values for a = 0.85 have been added.

will not be equally shared by the two sidebands. Since the instantaneous frequency of the composite signal lingers in the neighborhood of the mean of the two carrier frequencies (that is, $p + (1/2)\,r$ rad/sec) during the major portion of the difference-frequency cycle, more signal power should reside in each of the two components that have frequencies closest to that frequency (namely A_o and A_{-1}) than in any of the other components. This is indeed confirmed by the computed values for the amplitudes. The magnitude of A_o is larger than that of A_{-1}, and this may be appreciated by noting that the instantaneous frequency of the composite signal always puts it on the A_o side of the mean frequency $(p + (1/2)\,r)$ rad/sec.

From the choice of time reference ($t = 0$ when the two signals are in phase) we have alternately positive and negative real values for the spectral amplitudes. The signs at $t = 0$ or $2m\pi/r$, where m is any integer, are so distributed that the $A_{\pm n}$'s alternate in sign, starting with A_{+1} negative, A_o and A_{-1} positive. However, it is readily seen that at $t = q\pi/r$, where q is an odd integer, all the A_{+n}'s line up in the same positive direction as A_o, while all the A_{-n}'s line up opposite in phase to A_o.

Figure 5 shows the input and output spectra superimposed upon a plot of the instantaneous-frequency deviation of the resultant signal from the frequency of the stronger signal for $a = 0.8$.

Thus, it is seen that the limiting process, by eliminating the amplitude variations of the resultant, spreads out the resultant spectrum over an "infinite" band. The instantaneous frequency of the resultant signal after limiting (but with essentially all of the significant side-frequencies centered about p rad/sec passed unaltered) appears as the spike trains of Figs. 2 and 5. However, it must be borne in mind that the amplitude of the resultant will remain substantially constant, and the instantaneous-frequency variations will follow, essentially, the spike pattern given by Eq. 2, only when most .

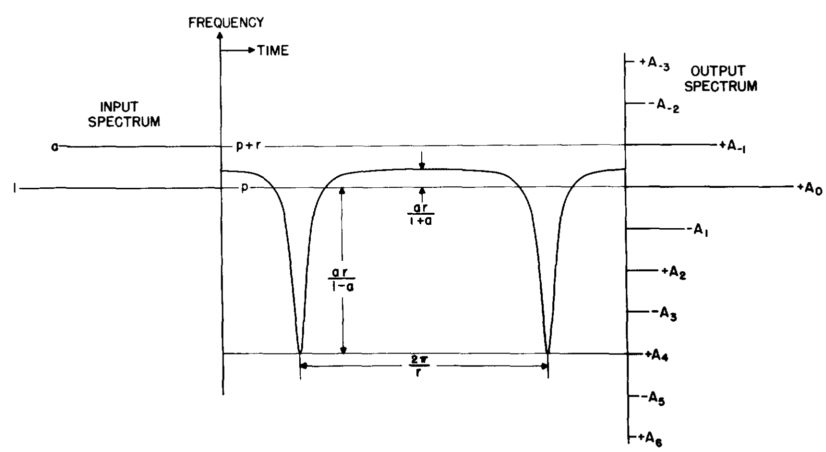

Fig. 5. Instantaneous-frequency disturbance caused by the interference. Input and computed output spectra are superimposed to clarify the notations and locations of the spectral components.

or all of the sideband components of significant strength (centered about the frequency p) are passed by the filter that follows the limiter.

We shall next determine the effect of eliminating some or most of the significant sideband components upon the possibility of rejecting the disturbance arising from the simultaneous presence of the weaker signal. This will spotlight the means of providing the proper bandwidths in the limiter-discriminator sections for the preservation of the interference rejection ability of the FM receiver.

1.3 A USEFUL THEOREM

The mathematical formulation of the general criterion which we shall use in determining whether or not interference can be suppressed when an ideal bandpass filter (that may reject major portions of the output spectrum) is inserted after the limiter, will derive largely from the properties of the spectral amplitude components, $A_{\pm n}$. The most important consequence of these properties can be appreciated by examining the behavior of the locus described during a period of $2\pi/r$ seconds, by the end point of the phasor that represents the resultant signal at the output of the filter. This behavior is indicated by the following theorem.

THEOREM 1. If, at the output of the limiter, an ideal filter is inserted that will pass: (a) an arbitrary number of components from both sidebands simultaneously or (b) an aribtrary number of components from the upper sideband, along with A_o only, then, over a period of $2\pi/r$ seconds, the terminus of the resultant phasor that represents the signal at the output of the filter will cross a reference axis along which a phasor representing A_o lies, only at $rt = m\pi$, when m is an integer or zero.

The ideal filter is characterized by a constant amplitude response within the passband and sharp cutoffs at its edges and by a linear phase characteristic (or constant time delay) over the passband.

Figure 6 is a phasor diagram illustrating the linear superposition of several spectral components that fall within the passband of the ideal filter. The plane of the figure can be imagined as rotating clockwise at an angular velocity of p rad/sec. A_o will be stationary, and the n^{th}-component phasor will rotate at nr rad/sec about its origin.

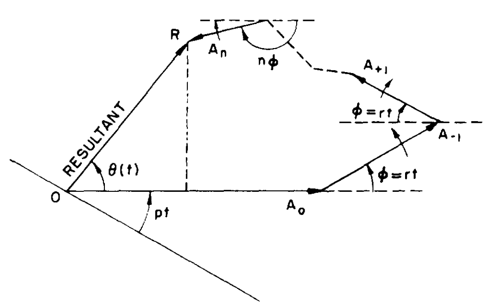

Fig. 6. Linear superposition of phasors representing the spectral components passed by the ideal-limiter filter.

For the axis of reference we choose arbitrarily the line along which A_o lies and call the origin of A_o "O". We want to show that the path traced by the point R during one complete cycle of the frequency difference r will cross the reference axis only at $rt = 0$ and $rt = \pi$. First, we shall demonstrate a few helpful lemmas.

Let us translate the assertion of the theorem into a more specific mathematical statement. We note from Fig. 6 that the locus of R crosses the reference axis only when Y, the instantaneous length of the vertical component of the resultant phasor, vanishes. But

$$\overline{OR} = \sum_{n=-N}^{M} A_n e^{-jn\phi}$$

$$= \sum_{n=-N}^{M} A_n \cos n\phi - j \sum_{n=-N}^{M} A_n \sin n\phi \tag{20}$$

where $\phi \equiv rt$, N is the number of upper sideband components that is passed, and M is the number of lower sideband components that is passed. Therefore, the locus traced by R crosses the reference axis for values of ϕ that are the roots of

$$Y = -\sum_{n=-n}^{M} A_n \sin n\phi \tag{21}$$

When components from both sidebands are passed, Eq. 21 can be reduced to the form

$$y = \frac{Y}{A_{-1} - A_1} = \sin \phi + \frac{A_{-2} - A_2}{A_{-1} - A_1} \sin 2\phi + \frac{A_{-3} - A_3}{A_{-1} - A_1} \sin 3\phi + \ldots$$

the sum terminating with the term contributed by the last component (in either or both sidebands) that is passed.

From Table I we find that the expression for y can be rewritten as

$$y = \sin \phi - b_1 \sin 2\phi + b_2 \sin 3\phi - b_3 \sin 4\phi + \ldots + b_{q-1} \sin q\phi \tag{22a}$$

where

$$b_o = 1$$

$$b_{\pm 1} = \frac{|A_{-2}| + |A_2|}{|A_{-1}| + |A_1|} \ldots \tag{22b}$$

$$b_{\pm(n-1)} = \frac{|A_{-n}| + |A_n|}{|A_{-1}| + |A_1|}$$

Similarly, we can show that expressions of the form of Eq. 22a exist for the special cases in which only upper sideband components or only lower sideband components are passed. For the former, $b_{-(n-1)} = |A_{-n}|/|A_{-1}|$; for the latter, $b_{+(n-1)} = |A_n|/|A_1|$.

By direct substitution from Eqs. 7 the expression for $b_{\pm n}$ takes the alternative useful form

$$b_{\pm n} = (-1)^n \frac{G_n - G_{n+2}}{G_o - G_2} \tag{22c}$$

In terms of Eq. 22a, theorem 1 states, in effect, that, when the magnitude of the coefficient of sin $n\phi$ is given by either (a) $b_{\pm n}$ or (b) b_{-n}, then, in the interval $0 \leqslant \phi < 2\pi$, y will have zeros only at $\phi = 0$ and $\phi = \pi$.

Admittedly, the zeros of the finite sum in Eq. 22a would best be placed in evidence by expressing this sum in a convenient closed form. However, any attempt to do this would meet with discouragement in view of the formidable appearance of the expressions for the coefficients of the sine terms. But the following lemmas are quite helpful.

LEMMA 1. If a finite sum of harmonically related sine terms (each weighted by an appropriate coefficient, the n^{th} term being given by $a_n \sin n\phi$) is to vanish only at $\phi = 0$ or $\phi = \pi$, in the range $0 \leqslant \phi < 2\pi$, then the zeros of the sum must not be produced by mutual cancellation among the terms, but only by the vanishing of the individual terms simultaneously and separately. This can only be ensured by special restrictions on the magnitudes of the weighting coefficients (a_n's).

The truth of this statement is best illustrated by referring to Eq. 22a. If only the first term in the sum is present, then $y = \sin \phi$ with zeros only at $\phi = 0$ and $\phi = \pi$. If only the first two terms are present, then $y = \sin \phi - b_1 \sin 2\phi$, whose zeros are those of $\sin \phi$ only if $b_1 \leqslant 1/2$. If only the first three terms are present, then $y = \sin \phi - b_1 \sin 2\phi + b_2 \sin 3\phi$, and with b_1 assigned its highest permissible value of 1/2, the zeros of y will be those of $\sin \phi$ only if $b_2 < 0.933$. The illustration grows in complication as more terms are dragged in, but the pattern is obvious. The coefficients $b_1, b_2, b_3, \ldots$ must obey certain restrictions on their magnitudes in order for the zeros of y to be identical with those of $\sin \phi$; that is, in order for the zeros of y not to be brought about by the various component terms canceling one another, but only by the simultaneous vanishing of the individual terms.

LEMMA 2. Given the two finite sums of harmonically related sine terms

$$y_1 = \sum_{n=1}^{q} a_n \sin n\phi \text{ and } y_2 = \sum_{n=1}^{q} b_n \sin n\phi$$

in which $a_1 = b_1$; otherwise the terms of the first sum dominate those of the second sum (that is, $|a_n| \geqslant |b_n|$), and corresponding coefficients (e.g., a_n and b_n) have the same sign. If y_1 has zeros only where $\sin \phi$ has zeros, then the zeros of y_2 must likewise be only those of $\sin \phi$.

Clearly, if the magnitudes of the various coefficients in the expression for y_1 are within the bounds imposed upon them by the condition that the zeros of y_1 be those of $\sin\phi$, then with the same restrictions on the coefficients of y_2 (since $a_1 = b_1$ and a_n and b_n have same signs) and with $|a_n| \geqslant |b_n|$, $n \neq 1$, the magnitudes of the coefficients of y_2 are certainly within the proper bounds to ensure that the zeros of y_2 be those of $\sin\phi$.

LEMMA 3. In the finite sum of sine terms given by Eq. 22a

$$y = \sin\phi - b_1 \sin 2\phi + b_2 \sin 3\phi - b_3 \sin 4\phi + \ldots + (-1)^{q-1} b_{q-1} \sin q\phi$$

where

$$b_{n-1} = \frac{|A_{-n}(a)| + |A_n(a)|}{|A_{-1}(a)| + |A_1(a)|} \quad \text{(for case a of theorem 1)}$$

and

$$b_{n-1} = |A_{-n}(a)| \big/ |A_{-1}(a)| \quad \text{(for case b of theorem 1)}$$

the n^{th} coefficient, $b_n (n\neq 0)$, is dominated by the corresponding coefficient, $(1/2)a^n$, in the similar sum,

$$S = \sin\phi - (1/2)\,a \sin 2\phi + (1/2)\,a^2 \sin 3\phi - (1/2)\,a^3 \sin 4\phi + \ldots$$
$$+ (-1)^{q-1} (1/2)\,a^{q-1} \sin q\phi \tag{23}$$

That is to say,

$$b_{\pm(n-1)} = \frac{|A_{-n}(a)| + |A_n(a)|}{|A_{-1}(a)| + |A_1(a)|} < (1/2)\,a^{n-1} \tag{24a}$$

and

$$b_{-(n-1)} = |A_{-n}(a)| \big/ |A_{-1}(a)| < (1/2)\,a^{n-1} \tag{24b}$$

where, as before, $\underline{a}$ lies between 0 and 1, and n is a positive integer different from 1.

From Eqs. 22c and 11a, it is readily seen that

$$b_{\pm n} = f(n, a) \cdot a^n \tag{25}$$

where $f(n, a)$ is a complicated function with no factorable powers of $\underline{a}$. That $f(n,a) < 1/2$ for all values of $\underline{a}$ and all values of n, is quite obvious from Figs. 7 and 8. An analytical demonstration is also possible, but it is too involved to be worthy of reproduction. Similar statements may also be made for b_{-n}, but not for b_{+n}, as is obvious from Figs. 7 and 8.

LEMMA 4. The finite sum S, given by Eq. 23, has zeros only where $\sin\phi$ has zeros, for all values of $\underline{a}$ between 0 and 1, and for all $q = 1, 2, 3, \ldots$.

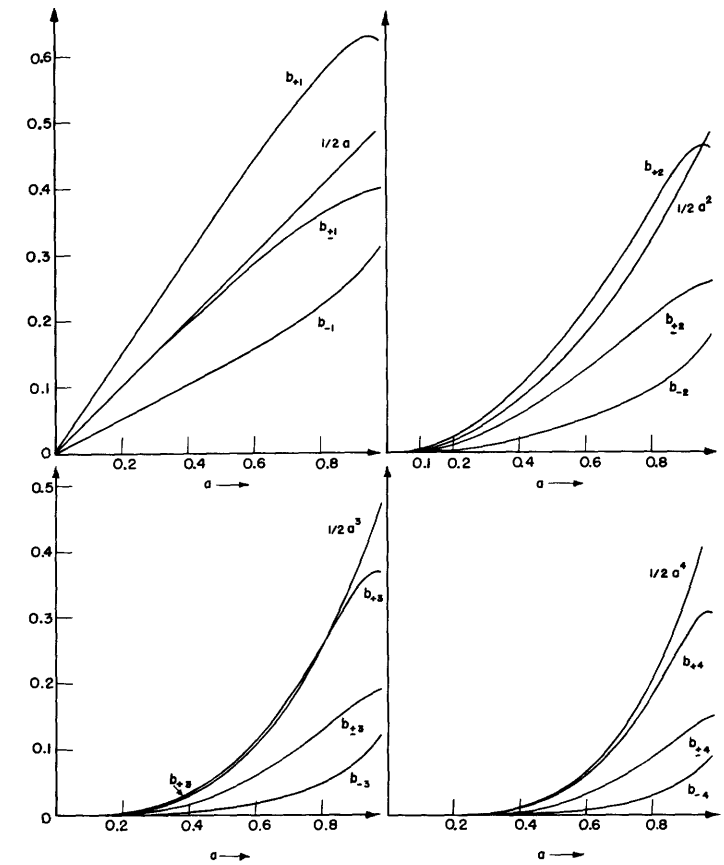

Fig. 7. Graphical demonstration of lemma 3.

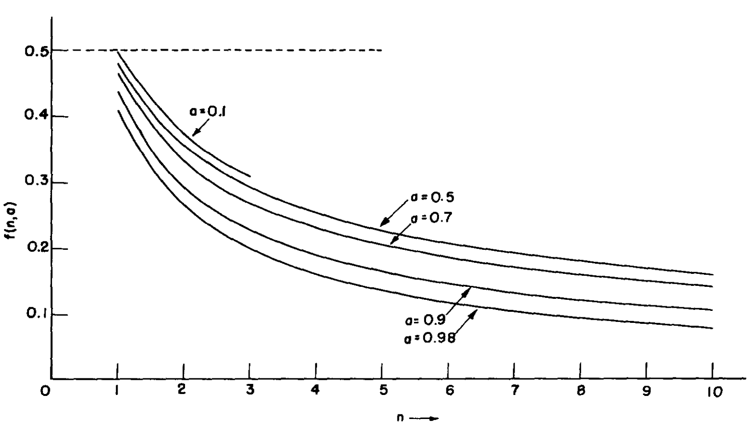

Fig. 8. Graphical demonstration of lemma 3.

This is obviously true when $S = \sin\phi$. It is also true for $S = \sin\phi - (1/2)\, a \sin 2\phi$, since the necessary restriction is that $(1/2) a$ be $\leq 1/2$; that is, for $a \leq 1$. When S is made up of the first three terms, it can be readily shown to be true for all $a < (7/4)^{1/2}$, which includes the range $a < 1$. Finally, Eq. 23 can be expressed in closed form as follows. First, write

$$2S = \sin\phi + \sin\phi - a \sin 2\phi + a^2 \sin 3\phi + \dots + (-a)^{q-1} \sin q\phi$$

$$= \sin\phi - (1/a) \sum_{n=1}^{q} (-a)^n \sin n\phi \tag{23a}$$

If $\sin n\phi$ is replaced by its value in terms of complex exponentials, and the standard formula for the sum of a finite geometric progression is used, it is readily established that, with $a < 1$, we have

$$2S = \sin\phi + \frac{\sin\phi - (-a)^q \left[\sin(q+1)\phi + a \sin q\phi\right]}{1 + 2a\cos\phi + a^2} \tag{23b}$$

As q is increased, a^q approaches zero and the zeros of S become more and more obviously those of $\sin\phi$. Therefore, we may conclude that lemma 4 is true for all $a < 1$, and all positive integers q. (Another argument based on Eq. 23b and making use of phasors is also possible.)

The argument that proves theorem 1 is now obvious. The sum in Eq. 22a must vanish only at the zeros of $\sin\phi$. But this sum is exactly similar to the sum in Eq. 23a, in that they are both made up of the same number of harmonically related weighted sine terms; the coefficient of the first term, $\sin\phi$, is the same in both, and the coefficient of $\sin n\phi$ has the same sign in both. Furthermore, by lemma 4, the finite sum in Eq. 23a vanishes only when $\sin\phi$ vanishes, and by lemma 3 the n^{th} coefficient, $(1/2)a^n$, in Eq. 23a, dominates the n^{th} coefficient, b_n, in Eq. 22a (only when $b_n = b_{\pm n}$ or b_{-n}; that is, for conditions a and b of theorem 1). Therefore, by lemma 2, the finite sum in Eq. 22a can vanish only at the zeros of $\sin\phi$, and this proves theorem 1.

In Fig. 9 plots of typical y's are shown for arbitrarily chosen values of a, N, and M in order to illustrate the demonstration given above. Perhaps they also provide an independent demonstration which is per se satisfactory to engineers.

In conclusion, the theorem cannot be extended to include the situation in which A_o is accompanied by lower sideband components only, for a greater than approximately 0.69, and for all values of q. The quoted upper limit on a can be read directly off the plot of b_{+1} in Fig. 7, since, for $a > 0.69$, b_{+1} exceeds the maximum permissible value of 0.5. Furthermore, the plots of Figs. 7 and 8 show that b_{+n} cannot be said to be bounded by $(1/2)a^n$ for all n and all $a \leq 1$, and so the argument presented above does not apply. Actually, the most serious violation of the conditions for this argument is the fact that b_1 does exceed 0.5 for $a > 0.69$, for otherwise the remaining coefficients

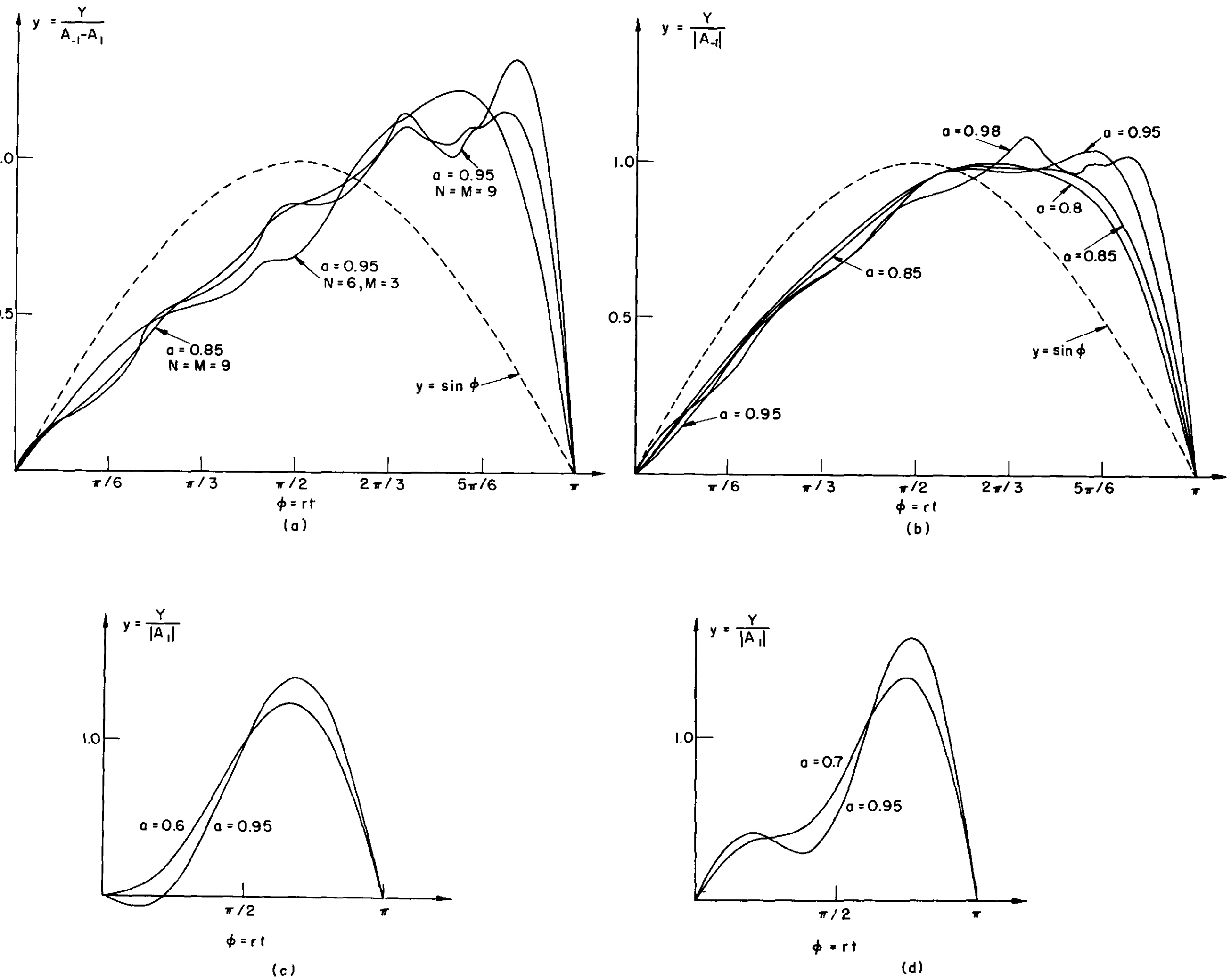

Fig. 9. Illustration of the behavior of the trigonometric sum in Eq. 22a: (a) case of components from both sidebands; (b) case of components from the upper sideband with N = 6; (c) case of components from the lower sideband with M = 2; (d) same as Fig. 9c with M = 3.

b_{+n} are not large enough to exceed the more liberal bounds that apply to them when b_1 is within its own bounds. Indeed, for the range $a < 0.69$, for which $b_{+1} < 0.5$, the corresponding finite sum, Eq. 22a, will have zeros at 0 or π only. This is illustrated in Fig. 9c and d. From these plots we may also conclude that the theorem holds for all values of $\underline{a}$ when q is odd; it only breaks down for even values of q in the range $0.69 < a < 1$.

The importance of theorem 1 will best be appreciated from the discussions of the two following sections.

1.4 A CRITERION FOR INTERFERENCE REJECTION

If the limiter bandwidth is narrowed down to pass only a portion of the power in each sideband, the interference will still be suppressed only if, over a period of $2\pi/r$ sec, the average frequency of the resultant of the passed components is still equal to the frequency of the stronger of the two carriers. It is clear that the minimum value that the limiter bandwidth can have is equivalent to one intermediate-frequency bandwidth. The conditions for this, or any other value of limiter bandwidth, to be permissible will now be determined.

At the output of the limiter, the component that has the frequency of the stronger signal is A_o. From Fig. 6 we find that the average frequency of the resultant, $\overline{OR}$, will be the frequency of A_o if and only if, over a period of $2\pi/r$ sec, the net phase deviation, θ, is zero. It is readily appreciated that, since the locus of the point R traces a closed path during a complete period of r, the net value of the phase deviation θ will be zero only if this closed path does not enclose the origin, O.

Now, the closed path traced by R will enclose the origin, O, if, at any instant of time, the resultant of the passed sideband components opposes A_o in phase and exceeds it in magnitude. Or, in terms of the resultant phasor $\overline{OR}$, the locus of R will enclose the origin O only if $\overline{OR}$ can assume a negative real value at any instant during the frequency-difference cycle. Obviously, $\overline{OR}$ becomes real only when the path of the terminal point R crosses the axis on which A_o lies (this axis being chosen as the axis of reals). But theorem 1 states that this can occur only when $rt = 0$ or π, and at no other instant during the cycle. Consequently, loci of R as shown in Fig. 10, for instance, are ruled out completely.

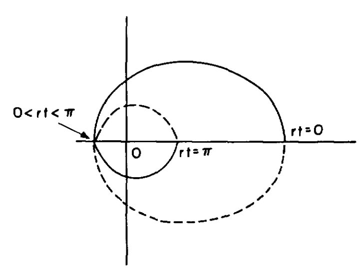

Fig. 10. Type of locus for the end point, R, of the resultant phasor which is ruled out by theorem 1.

Now, the n^{th} upper and lower sideband components are $A_{-n} \exp(jnrt)$ and $A_n \exp(-jnrt)$. Furthermore, Table I reveals that the spectral terms in each sideband alternate in sign, A_{-n} and A_n being positive for n odd and even, respectively. As a consequence, we find that, since (with $rt = \pi$)

$$e^{jn\pi} = e^{-jn\pi} = \begin{cases} -1, & \text{for } n \text{ odd} \\ +1, & \text{for } n \text{ even} \end{cases}$$

the signs are so distributed that, at $rt = \pi$, all the upper sideband components line up in phase opposition to A_o, while all the lower sideband components line up in phase aiding A_o.

Finally, at $rt = 0$, $\exp(\pm jnrt) = 1$, for all n. Consequently, the components in each sideband are so oriented that every other component aids or opposes A_o directly, A_{-n} and A_n aiding A_o for n odd and even, respectively. In this mutual cancellation among the terms, with A_{-1} heavily weighting the positively oriented components, it is very unlikely that the passed components will subtract from the magnitude of A_o.

We conclude, therefore, that the only critical instant of time to consider, during a frequency-difference cycle, is that corresponding to $t = \pi/r$. The following theorem can therefore be stated as the criterion for the loss or preservation of the desired average frequency (hence for the possibility of rejecting the interference) when the ideal limiter is followed by an ideal narrow-band filter.

THEOREM 2. If arbitrary numbers N and M of upper and lower sideband components fall within the passband of the ideal filter that follows the limiter, the average frequency of the resultant of all the passed components, including the component A_o, will be exactly the frequency of A_o if and only if

$$\sum_{n=-N}^{M} A_n e^{-jn\pi} = \sum_{n=-N}^{M} (-1)^n A_n < A_o \qquad n \neq 0$$

This important inequality can also be expressed in the more convenient form

$$\sum_{n=0}^{M} |A_n| > \sum_{n=1}^{N} |A_{-n}| \tag{26}$$

PROOF. At the output of the ideal filter the resultant signal is given by

$$e(t) = \sum_{n=-N}^{M} A_n \cos(p - nr)t$$

The corresponding complex function of time is

$$E(t) = e^{jpt} \sum_{n=-N}^{M} A_n e^{-jnrt}$$

which may be expressed as

$$E(t) = e^{jpt} F(t) \tag{27}$$

Over a period of $2\pi/r$ sec, the net phase shift of E(t) is $2\pi p/r$ if and only if, over $2\pi/r$ sec, the complex function

$$F(t) \equiv \sum_{n=-N}^{M} A_n e^{-jnrt}$$

introduces no net phase change.

For convenience, let us shift our time reference from the instant at which $rt = 0$ to the instant at which $rt = \pi$. For this purpose we substitute $\tau + \pi/r$ for t to get

$$H(\tau) = F(\tau + \pi/r) = \sum_{n=-N}^{M} (-1)^n A_n e^{-jnr\tau}$$

or

$$H(\tau) = -\sum_{n=1}^{N} |A_{-n}| e^{jnr\tau} + \sum_{n=0}^{M} |A_n| e^{-jnr\tau}$$

Now let $z = e^{jr\tau}$, in order to obtain

$$h(z) = -\sum_{n=1}^{N} |A_{-n}| z^n + \sum_{n=0}^{M} |A_n| z^{-n} \tag{28}$$

As $\exp(jr\tau)$ covers one complete cycle of variation over a period of $2\pi/r$ sec, z traverses the unit circle in the z-plane once counterclockwise, and h(z) traces some closed path C' in the h(z)-plane, as shown in Fig. 11. In tracing C' counterclockwise h(z) will sustain a net phase shift given by $2\pi(Z-P)$, where Z and P are the numbers of zeros and poles of h(z), within the unit circle in the z-plane, each zero or pole being counted in accordance with its multiplicity. But, from a well-known theorem in function theory (10), if a function f(z) is analytic, except for possible poles within and on a given contour, the number of times that the plot of f(z) encircles the origin of the h(z)-plane counterclockwise while z itself traverses a prescribed contour once counterclockwise is equal to the number of zeros, Z, diminished by the number of poles, P, of f(z) within the contour in the z-plane (each pole or zero being counted according to its multiplicity).

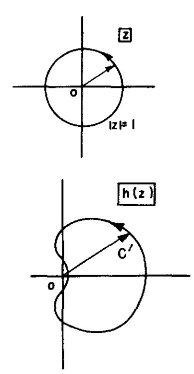

Fig. 11. Illustration for the proof of theorem 2.

Therefore, if h(z) is to acquire no net phase shift in tracing the path C' once, the quantity Z - P must be zero or, equivalently, the path C' must not encircle the origin of the h(z)-plane. This condition is rather obvious from an examination of Fig. 11. It is also readily appreciated that if, while z traverses the unit circle and h(z) describes the path C', h(z) never assumes a negative real value, then C' will never encircle the origin of the h(z)-plane.

Now, on the unit circle, Eq. 28 can be written in the form

$$h(z)\big|_{|z|=1} = -\sum_{n=1}^{N} |A_{-n}| \cos n\phi + \sum_{n=0}^{M} |A_n| \cos n\phi$$

$$-j\left[\sum_{n=1}^{N} |A_{-n}| \sin n\phi + \sum_{n=0}^{M} |A_n| \sin n\phi\right]$$

and this is recognized to be equivalent to Eq. 20, with the reference axis shifted by π radians. The roots of the imaginary component of $h\big(|z| = 1\big)$ are then exactly the roots of Eq. 22a, $\phi = 0$ and $\phi = \pi$ in the range $0 \leq \phi < 2\pi$. Therefore, $h\big(|z| = 1\big)$ becomes real only when z = 1 or -1, corresponding to $\phi = 0$ and $\phi = \pi$, and its real values are given by

$$\begin{aligned} h(-1) &= -\sum_{n=1}^{N} (-1)^n |A_{-n}| + \sum_{n=0}^{M} (-1)^n |A_n| \\ &= \sum_{n=-N}^{M} A_n \\ &= A_o + \big(|A_{-1}| - |A_1| - |A_{-2}|\big) + \big(|A_2| - |A_3|\big) \\ &\quad + \big(|A_4| - |A_5|\big) + \dots + \big(|A_{M-1}| - |A_M|\big) \\ &\quad + \big(|A_{-3}| - |A_{-4}|\big) + \big(|A_{-5}| - |A_{-6}|\big) + \dots \\ &\quad + \big(|A_{N-1}| - |A_N|\big) \end{aligned} \tag{29}$$

and

$$h(1) = -\sum_{n=1}^{N} |A_{-n}| + \sum_{n=0}^{M} |A_n| \tag{30}$$

It is readily ascertained from Table I that all of the terms in parenthesis in the expression for h(-1) are positive. Consequently, h(1) is the minimum real value that h(z) can assume on the unit circle. If this minimum real value is positive, h(z) will never become negative real for $|z| = 1$; hence the path C' traced by h(z) in the h(z)-plane (as z traces the unit circle in the z-plane) will never encircle the origin of the h(z)-plane. From Eq. 30, the condition for the minimum real part, h(1), to be positive is

$$\sum_{n=0}^{M} |A_n| > \sum_{n=1}^{N} |A_{-n}|$$

This is recognized as the inequality stated in theorem 2, and thus it completes the formal proof.

We shall next apply the criterion of theorem 2 to the determination of the minimum permissible values of limiter bandwidth for the suppression of interference.

1.5 MINIMUM PERMISSIBLE LIMITER BANDWIDTHS

At the outset, we recognize that with a narrow-band filter whose bandwidth can at best be equal to, but never less than, the i-f bandwidth, the possible configurations of accomodated side-frequency components are resolved into three situations. First, there is the limiting case in which only A_o and an arbitrary number of lower sideband components are passed, to the complete exclusion of all of the upper sideband components. Second, a limiting case arises when it is the lower sideband components that are not passed by the ideal filter. Third, the general case occurs when some components from both sidebands are simultaneously passed (with A_o, of course). It is needless to point out that the remarkable simplification in the approach that the use of the concept of ideal filters makes possible, will be best manifested by the following analysis. For instance, with an ideal filter, we are able to draw sharp lines of demarcation between the three possible situations, and thus reduce our problem to three simpler problems. The results and experience are not only needed for the analysis of section 1.6, but they also serve as an invaluable guide to a clearer understanding of the nature of the problem, and to the selection of actual design figures.

Case A. Consider first the situation in which only an arbitrary number, M, of lower sideband components is passed, along with A_o, while all of the upper sideband components fall outside the passband. Although this situation is possible only when the ideal filter has one i-f bandwidth that is not well centered about the intermediate frequency, it will be treated for the sake of completeness.

At $t = \pi/r$, all of the lower sideband components line up in phase, aiding A_o. Thus the resultant phasor can never be negative at this instant of time. At $t = 0$, we have

$$F(0) = \sum_{n=0}^{M} A_n$$

$$= \left(|A_o| - |A_1|\right) + \left(|A_2| - |A_3|\right) + \left(|A_4| - |A_5|\right) + \dots + \left(|A_{M-1}| - |A_M|\right)$$

All the terms in parenthesis on the right are positive numbers; thus F(0) is also always positive real. This completes the check for odd values of M, since this case is covered by theorem 1. However, for even values of M, we must investigate the positive-realness of F(t) at an additional instant in the cycle, given by $rt = \phi_1$, where $0 < \phi_1 < \pi$. Here

$$F(\phi_1/r) = |A_o| \cdot [1 - |A_1/A_o| \cos \phi_1 + |A_2/A_o| \cos 2\phi_1 - |A_3/A_o| \cos 3\phi_1 + \dots + |A_M/A_o| \cos M\phi_1]$$

It is a simple matter to show that the coefficients in the finite series in brackets are dominated by the corresponding coefficients in the series

$$\begin{aligned} z(\phi) &= 1 - a \cos\phi + a^2 \cos 2\phi - a^3 \cos 3\phi \\ &\quad + \dots + (-a)^n \cos n\phi + \dots + a^M \cos M\phi \\ &= \sum_{n=0}^{M} (-a)^n \cos n\phi \end{aligned}$$

If $\cos n\phi$ is replaced by its value in terms of complex exponentials, and the resulting finite geometric progressions are summed in the usual manner, $z(\phi)$ can be expressed in the closed form (with $a^2 < 1$)

$$\begin{aligned} z(\phi) &= \sum_{n=0}^{M} (-a)^n \cos n\phi \\ &= \frac{1 + a \cos \phi + a^{M+1} \cos(M+1)\phi + a^{M+2} \cos M\phi}{1 + 2a \cos \phi + a^2} \end{aligned}$$

Since $a < 1$, it is evident that as M becomes large

$$z(\phi) \to \frac{1 + a \cos \phi}{1 + 2a \cos \phi + a^2}$$

and this quantity can never go negative for any real value of ϕ. For the lower values of M, a close examination of the numerator in the expression for $z(\phi)$ reveals that $z(\phi)$ can never go negative. Actually, the preceding argument is independent of ϕ and of whether M is even or odd. It can, therefore, be used to establish theorem 3 without the help of theorem 1.

We conclude that at no instant of time will any arbitrary number of lower sideband components produce a resultant that opposes A_o in phase and exceeds it in magnitude. Furthermore, this holds for all $a < 1$.

THEOREM 3. If only A_o and an arbitrary number of lower sideband components fall within the ideal-filter passband, the average frequency of the resultant signal at the output of the filter is still the frequency of the stronger signal.

Granlund (1) proved this theorem in the following way. Over a period of $2\pi/r$ sec, the quantity

$$F(t) = \sum_{n=0}^{M} A_n e^{-jnrt}$$

must add no phase shift to the resultant signal. To show that it does not, let $z = \exp(-jrt)$ and write

$$f(z) = \sum_{n=0}^{M} A_n z^n$$

As $\exp(-jrt)$ covers one cycle of variation over a period $2\pi/r$, z traverses the unit circle in the z-plane clockwise. Since $f(z)$ has no poles within the unit circle, the net phase change that $f(z)$ sustains while z traverses the unit circle is simply 2π multiplied by the number of zeros of $f(z)$ within the unit circle, each zero being counted in accordance with its multiplicity. But $f(z)$ has as many zeros within the unit circle as

$$f(-z) = \sum_{n=0}^{M} (-1)^n A_n z^n$$

$$= \sum_{n=0}^{M} |A_n| z^n$$

The zeros of a polynomial of this kind (characterized by positive real coefficients that decrease with n), according to Hurwitz (11), lie within the annular ring

$$\left[\frac{|A_n|}{|A_{n+1}|}\right]_{min} < |z| < \left[\frac{|A_n|}{|A_{n+1}|}\right]_{max}$$

where $n = 0, 1, 2, \ldots, M-1$. Since $|A_n|$ decreases monotonically with n, this ring lies outside the unit circle, and so $f(-z)$ has no zeros within the unit circle. This completes the proof of theorem 3.

Case B. Consider the situation in which only an arbitrary number, N, of upper sideband components is passed, along with A_o, to the complete exclusion of the lower sideband components. This is possible only with an ideal filter of one i-f bandwidth, if filters whose passbands are not well centered about the intermediate frequency are excluded from consideration. Theorem 2 applies, and the inequality

$$A_o > \sum_{n=1}^{N} |A_{-n}| \tag{31}$$

must be satisfied. This conclusion can also be reached in the following interesting manner. As before, we require that

$$F(t) = \sum_{n=0}^{N} A_{-n}\, e^{jnrt} \tag{32}$$

shall not introduce any net phase shift over a period of $2\pi/r$ sec. If, for convenience, we substitute $\tau + \pi/r$ for t to shift the time reference from $t = 0$ to $t = \pi/r$, we can write

$$H(\tau) = F(\tau + \pi/r)$$

$$= A_o - \sum_{n=1}^{N} |A_{-n}|\, e^{jnrt}$$

If we set $z = \exp(jrt)$, we obtain

$$h(z) = A_o - \sum_{n=1}^{N} |A_{-n}|\, z^n \tag{33}$$

As before, $h(z)$ will acquire a net phase shift, as z traverses the unit circle once, if and only if $h(z)$ has zeros within the unit circle (it has no poles there). Such zeros can exist only if, for $|z| \leq 1$, the right-hand side of Eq. 33 vanishes. Since the summation term is analytic within and on the unit circle in the z-plane, we have from the principle of the maximum modulus (10) that this term assumes its maximum value of

$$\sum_{n=1}^{N} |A_{-n}|$$

on the circle itself. Therefore, if

$$A_o > \sum_{n=1}^{N} |A_{-n}|$$

$h(z)$ cannot have any zeros within the unit circle.

In view of the complexity of the expressions for the A_n's, the criterion is best applied graphically. Figure 12a is a plot of the sum of all the tabulated A_{-n} amplitudes that are significant over the whole range of a taken from Table I. Superimposed on this plot are plots of $A_o(a)$ and of $\sum_{n=0}^{M} |A_n(a)|$ for several values of M. Figure 12b

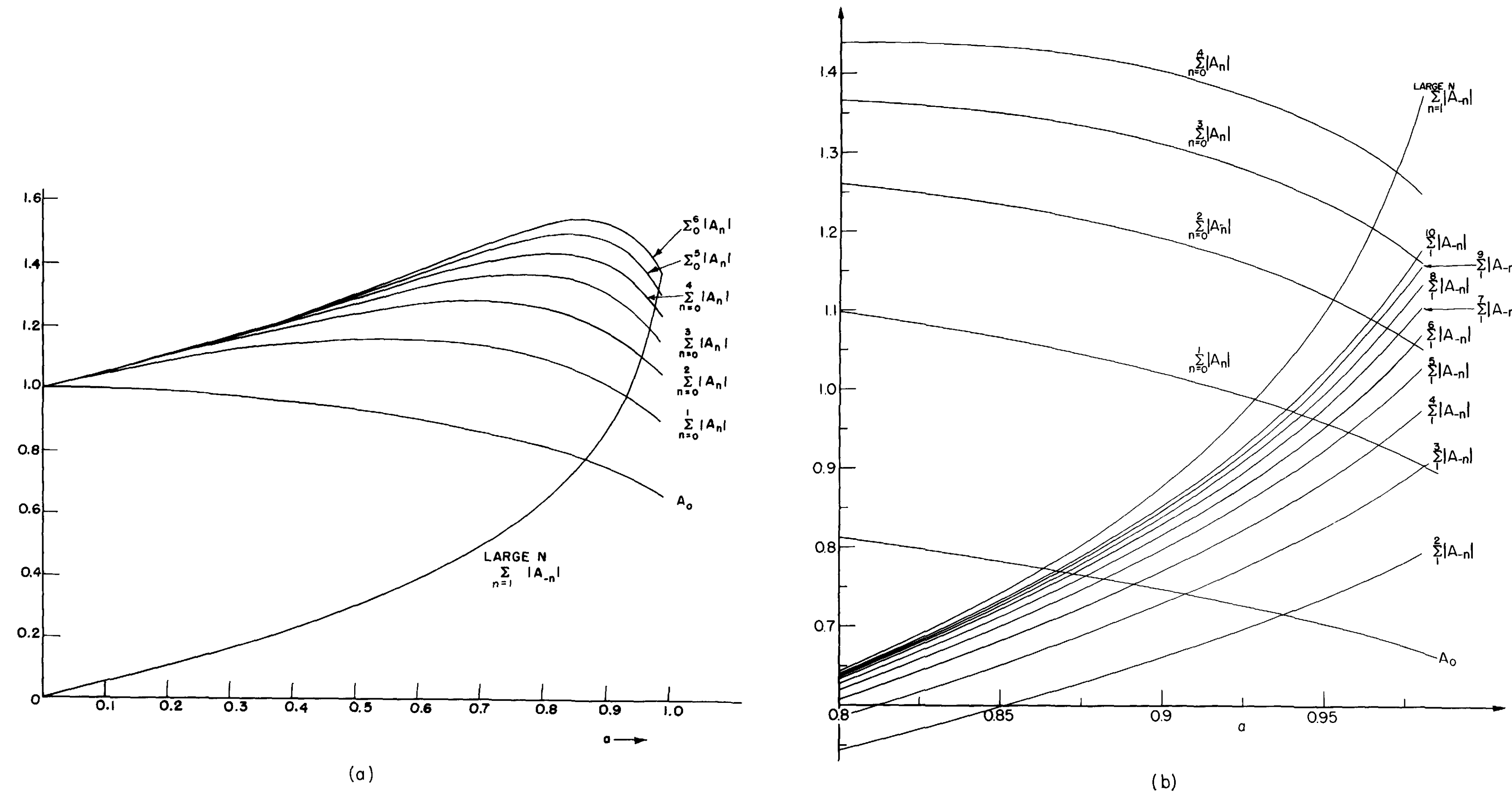

Fig. 12. Variation, with a, of: A_o, the spectral component that has the frequency of the stronger carrier; $\sum_{n=1}^{N} |A_{-n}|$, the sum of the magnitudes of the first N upper sideband components (the curve *for which the* upper limit of the summation is marked "large N" represents, essentially, the sum of the magnitudes of all the upper sideband components); $\sum_{n=0}^{M} |A_n|$, the sum of the magnitudes of A_o and the first M lower sideband components. (a) Over-all appearance of curves. (b) Enlarged *region of intersections.*

shows an enlarged view of the region of intersections in Fig. 12a. From these plots it is evident that the magnitude of A_o is greater than the sum of the magnitudes of (effectively) all of the upper sideband components for $a \leqslant 0.863$.

THEOREM 4. If only A_o and an arbitrary number of upper sideband components fall within the ideal filter passband, then the average frequency of the resultant signal at the output of the filter will still be the frequency, p, of the stronger signal for values of $a \leqslant 0.863$.

For $a \geqslant 0.863$, the average frequency of the resultant signal is $p + r$, the frequency of the weaker (interfering) signal, if more than a few upper sideband components are passed. This is illustrated in Fig. 13 by a plot of the path traced by the end point of the resultant phasor over a period of $2\pi/r$ sec, for $a = 0.95$, when only A_o, A_{-1}, and A_{-2} are passed. The encirclement of the origin, O, by the traced path signifies a gain of 2π radians, over the phase of A_o, by the resultant signal every $2\pi/r$ sec. The resultant has, therefore, an average frequency of $p + r$ radians/second.

Figure 12b also shows plots of $\sum_{n=1}^{N} |A_{-n}(a)|$ for various values of N. The intersections of these plots with the plot for $A_o(a)$ determines the value of $\underline{a}$ up to which a certain number, N, of upper sideband components can be passed (with A_o only) before the desired average frequency, p, is lost.

We can conclude, therefore, that the bandwidth of the limiter need not exceed the bandwidth of the i-f section (for interference rejection) for interference ratios up to $a = 0.863$. For values of $a > 0.863$, bandwidths greater than that of the i-f section are required.

The minimum permissible limiter bandwidths, for $a > 0.863$, can be determined as follows. As before, let N be the number of upper sideband components passed, and M be the number of lower sideband components passed.

(a) Let the worst situation that must be handled satisfactorily by the filter be one in which $M = 0$ and $N = N_{max}$. Clearly, this implies that the situation in which $N = N_{max} + 1$ can arise only along with $M = 1$.

(b) Determine the minimum ideal-filter bandwidth, in units of one i-f bandwidth, for which the situation in step (a) is the limiting situation. This can most conveniently be done by first drawing a diagram like the one in Fig. 14 (drawn for $N_{max} = 4$). It is evident from this diagram that for $N = N_{max}$, $M = 0$ to arise, the frequency difference, r, should be greater than some value, r_ℓ, that is given by $r_\ell = (BW)_{if}/N_{max}$. For this value of r, $N = N_{max} + 1$, $M = 1$ arises, and so the limiter bandwidth should be

$$\begin{aligned}(BW)_{lim} &= r_\ell(M+N) = r_\ell(2 + N_{max}) \\ &= (BW)_{if}[1 + 2/N_{max}] \qquad (34)\end{aligned}$$

Clearly, this is the minimum limiter bandwidth required here, since smaller values of bandwidth will allow situations to arise in which $N > N_{max}$ and $M = 0$, while larger

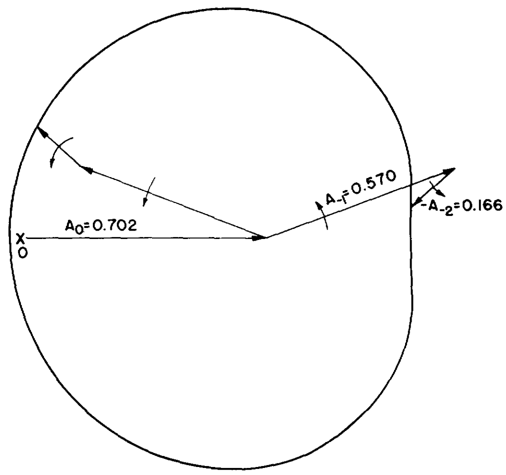

Fig. 13. Path traced by the end point of the resultant phasor over a period of $2\pi/r$ sec, for the case of a = 0.95, when only A_0, A_{-1}, and A_{-2} are passed.

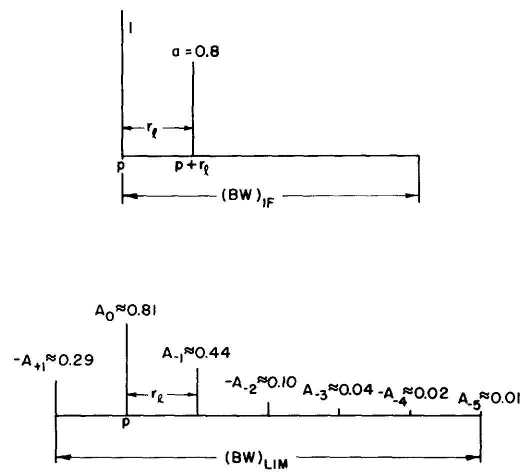

Fig. 14. The idealized passband of the limiter filter provides the largest available space for the upper sideband components when the stronger signal falls infinitesimally to the right of the lower cutoff frequency of the idealized i-f filter passband.

values will have limiting configurations in which $N < N_{max}$, and $M = 0$, the value $N = N_{max}$ arising only along with some nonzero M.

(c) Determine from Fig. 12b up to what value of $\underline{a}$ the inequality

$$A_o > \sum_{n=1}^{N_{max}} |A_{-n}|$$

is satisfied. Then up to this value of $\underline{a}$, the minimum required limiter bandwidth is the value that was found in step (b).

Table II

M	N_{max}	Min $\dfrac{(BW)_{lim}}{(BW)_{if}}$ Required	a_{max}
0	2	2	0.937
0	3	1 2/3	0.906(5)
0	4	1.5	0.891
0	5	1.4	0.882
0	6	1 1/3	0.877
0	7	1 2/7	0.873
0	8	1.25	0.870(5)
0	9	1 2/9	0.868(6)
0	10	1.2	0.867
0	11	1 2/11	0.865(6)
0	12	1 1/6	0.865(2)

Table II summarizes the results of calculations that cover the requirements for the range $0.863 < a \leqslant 0.937$. These results are also plotted in Figs. 15 and 16. The transition in the requirements from one range of values of $\underline{a}$ to the next takes place in steps. This can be justified in the following way. Let $a = a_{max}$ mark the end of a range in which the requirement is set by the configuration $M = 0$, $N = N_{max}$. This means that immediately beyond $a = a_{max}$ the requirement is set by $M = 0$, $N = N_{max} - 1$. The ideal filter response will either pass, or completely reject, a spectral component in the neighborhood of its cutoff frequencies. Therefore, the transition from one region to the next must occur in a step.

It is evident that for $a > 0.937$ the bandwidth of the limiter must be so chosen that at least one or more lower sideband components are passed at all instants of time, regardless of the value of the frequency difference r, if interference is to be suppressed.

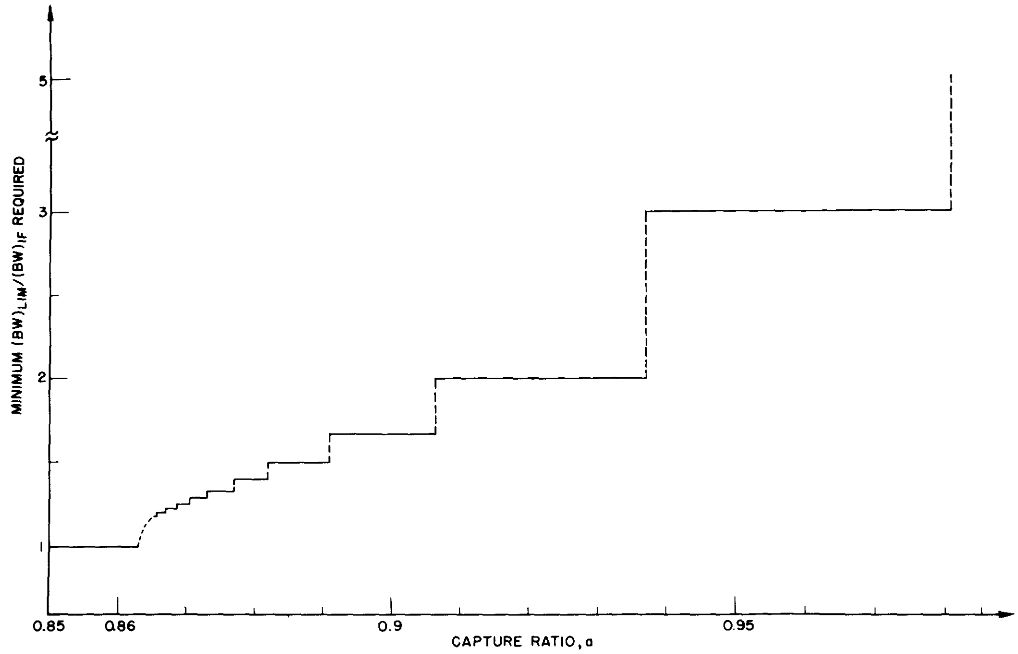

Fig. 15. Calculated minimum values of the bandwidth of the ideal-limiter filter for keeping the average instantaneous frequency of the resultant signal equal to the frequency of the stronger signal.

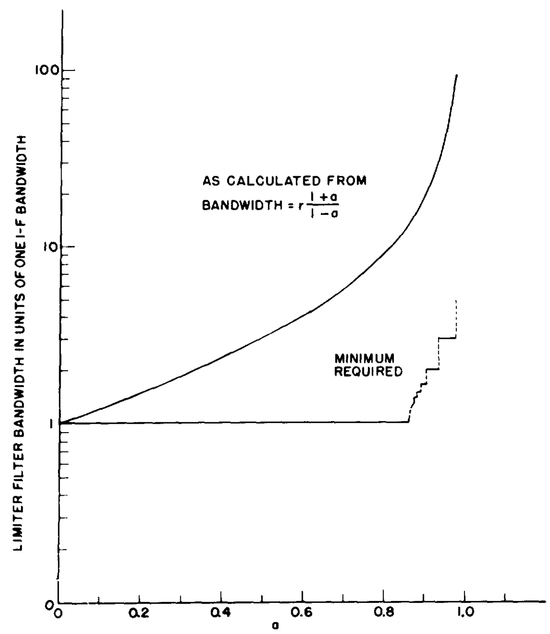

Fig. 16. The "sufficient" requirements for limiter bandwidth specified by previous investigators, superimposed upon the calculated necessary minimum requirements.

Case C. Consider, finally, the situation in which components from both sidebands fall within the ideal-filter passband. Configurations falling in this category will evidently decide the minimum limiter bandwidth requirements in the range $a > 0.937$, and here the criterion of theorem 2 applies directly.

A rough indication of the relative numbers of upper and lower sideband components that must be accommodated in limiting situations, in order to preserve the possibility of suppressing the interference, is indicated in Table III. From the criterion of theorem 2, it is clear that the interference rejection ability is enhanced by the presence (at all times) of some lower sideband components within the ideal-filter passband. The ratios N/M, presented in Table III, suggest that for $a \leqslant 0.98$, interference can always be rejected if, in the worst possible situation, the ideal-filter bandwidth is sufficient to accommodate about twice as many upper sideband components as lower sideband components. This rule of thumb is helpful only in the range $0.937 < a \leqslant 0.98$, where M must never be zero to ensure interference rejection.

Table III

Capture ratio a	Number of lower sideband components passed = M	Maximum permissible number of upper sideband components = N	N/M
0.98	1	3	3
	2	5	2.5
	3	9	3
	4	14	3.5
0.95	1	6	6
	2	N arbitrary	--

For a more careful determination of the minimum requirements in limiter bandwidth for the range $a > 0.937$, we first extend the reasoning used in case B to the present situation with $M \neq 0$. Thus, if the worst configuration of passed components is to arise only with $M = M_{min}$ and $N = N_{max}$, it is clear that the bandwidth employed must be so chosen that N can assume the value $N_{max} + 1$ only if $M = M_{min} + 1$ arises simultaneously with that value. As is evident from a diagram similar to the one in Fig. 14, if for the worst situation to arise the frequency difference, r, must exceed a limiting value of r_m rad/sec, then the necessary bandwidth is given by

$$\frac{(BW)_{lim}}{(BW)_{if}} = 2 \frac{r_m}{(BW)_{if}} (M_{min} + 1) + 1$$

But

$$r_m = \frac{r_m(M_{min} + 1) + (BW)_{if}}{N_{max} + 1}$$

or

$$\frac{r_m}{(BW)_{if}} = \frac{1}{N_{max} - M_{min}}$$

Therefore,

$$\frac{(BW)_{lim}}{(BW)_{if}} = \frac{N_{max} + M_{min} + 2}{N_{max} - M_{min}} \tag{35}$$

From Table III we find that the situations in which $M_{min} = 1$, $N_{max} = 3$ and $M_{min} = 2$, $N_{max} = 5$, are both limiting situations that the limiter filter must handle. Both configurations require that the limiter bandwidth be three times the i-f bandwidth, which can be verified by direct substitution in Eq. 35. It is also readily appreciated that this value of bandwidth is, in fact, the limiter bandwidth required to make it impossible for the configuration $M = 0$, $N = 2$ to arise in the range $a > 0.937$. We also notice from Fig. 12b that the plot of

$$\sum_{n=0}^{1} |A_n|$$

intersects the plot of

$$\sum_{n=1}^{3} |A_{-n}|$$

at $a = 0.9807$, while the intersection of

$$\sum_{n=0}^{2} |A_n|$$

with

$$\sum_{n=1}^{5} |A_{-n}|$$

occurs at a higher value for $\underline{a}$. Thus, in the range $0.937 < a \leqslant 0.9807$, the configuration $M = 1$, $N = 3$ is the most critical one. The minimum limiter bandwidth required in this range is set, by this situation, at three times the i-f bandwidth.

For values of $\underline{a}$ just exceeding $a = 0.9807$, $N = 3$ should only arise with $M = 2$. This corresponds to a limiting situation described by $M_{min} = 1$, $N_{max} = 2$, and thus requires a minimum limiter bandwidth of five i-f bandwidths. However, a little check reveals that with a limiter bandwidth of $5(BW)_{if}$, $M = 2$, $N = 3$ can arise only when the frequency difference r = one i-f bandwidth, whereas $M = 1$, $N = 2$ cannot arise at all, since r would have to be larger than one i-f bandwidth. But $M = 2$, $N = 4$ can arise here, and it is even more serious than $M = 2$, $N = 3$. Consequently, the upper limit on a, for the present range, is defined by $M = 2$, $N = 4$.

It should be borne in mind that Eq. 35 will give the correct answer only when the values corresponding to the worst (or limiting) situation are substituted for M_{min} and N_{max}. On the basis of the argument leading to this equation, in the "worst situation" $N_{max} + 1$ must be accompanied by $M_{min} + 1$; that is, it should be possible through an infinitesimal change in the value of the frequency difference, r, to restore the situation M_{min}, N_{max}. Thus, for the calculation of the required limiter bandwidth in the range that is just above $a = 0.9807$, substitution of $M_{min} = 2$, $N_{max} = 4$ in Eq. 35 does not yield the right answer, since this configuration does not satisfy the indicated criterion.

Up to this point, we have been carrying on the discussion in terms of the case in which the weaker signal has the higher frequency, $p + r$. The results can be easily carried over to the case in which the weaker signal has the lower frequency. There, r is simply replaced by $-r$, and so the line spectra that appeared in the upper sideband previously will now form the lower sideband, while those that appeared in the lower sideband will now constitute the upper sideband. The steps in the previous analysis may thus be retraced with the terms "upper" and "lower" interchanged throughout. Therefore, the swapping of either relative signal strengths, or relative signal frequencies, will not affect the conclusions reached in connection with the limiter bandwidth requirements.

In fact, it is to take care of such an alternative situation that the limiter-filter and the i-f amplifier amplitude-response characteristics have been centered about the same frequency throughout the analysis (see Fig. 14) and no effort has been made to allow the limiter selectivity to discriminate between the two sidebands.

In either situation, however, the effect that each of the two sidebands will have on the loss or preservation of the desired average frequency, is easily distinguishable. The sideband that is on the same side as the weaker signal (relative to the stronger signal) will always contribute the components that will try to offset the average frequency in favor of the frequency of the weaker signal. A physical feeling for this conclusion may also be acquired through a better appreciation of the basic relationship between the constituent spectral components and the characteristics of the resultant wave. The degree of correlation that seems to exist between the amplitude distribution of the components, on the frequency scale, and the instantaneous-frequency pattern of the resultant wave, has already been pointed out in our discussion of the spectrum. Inasmuch as the spectral components are basically the "building blocks" of the resultant signal, the components that tend to pull the instantaneous frequency of the resultant signal toward the frequency of the weaker signal must logically be those that lie on the same side relative to the frequency p (of the stronger signal) as the frequency of the weaker signal. The components in the opposite sideband provide the balancing necessary (to aid the component A_o) to preserve the desired average value, p, of the frequency of the resultant signal.

The preceding study embodies the first step in a novel switch in the basic approach to the limiter selectivity problem, and in the philosophy of the limiter's share of the task of interference suppression in FM. The more impressive aspects of this new line of thought will be covered by later discussions.

At this point, it suffices to say that the discussion of this section emphasizes the minimum basic requirement that the limiter filter must satisfy; that is, no matter in what way it alters the spectrum delivered to it by the limiter, it must always preserve the desired average instantaneous frequency in the resultant signal that it delivers to the succeeding stages by accommodating proper portions of each sideband. Thus the reason that the large values of limiter bandwidth formerly prescribed and tested (1) have enabled the achievement of high capture ratios (better than 0.95), is primarily that such bandwidths will allow a considerable number of components from the helpful sideband to be present within the passband at all times. Consequently, the bandwidth values specified by the formula

$$(BW)_{lim} = \frac{1+a}{1-a} r \tag{36}$$

although they are helpful, are not necessary for interference rejection. Equation 36 is plotted in Fig. 16 for comparison with the results of our computations.

These computations also emphasize the fact that the desired average frequency of the resultant signal can be preserved without necessarily providing the bandwidth value dictated by the extent of the frequency spikes. In fact, as far as the limiter bandwidth is concerned, it will become apparent, later on, that there is no special significance to the bandwidth value specified by Eq. 36. This value will even be found to fall short of satisfying the conditions for a physical filter to follow the instantaneous frequency of the resultant signal through quasi-stationary states. The basic condition that the limiter bandwidth must satisfy is merely to be able, in the worst situations possible, to pass portions of the sidebands that will result in a signal whose average instantaneous-frequency deviation from the frequency p of the stronger signal is zero, over a period of the frequency difference r. Since, in general, only a finite number of significant components will be passed, the resultant signal at the output of the limiter filter will exhibit instantaneous variations in amplitude, as well as frequency. If an amplitude-insensitive discriminator follows the filter, the conversion of the instantaneous-frequency pattern (of the resultant signal delivered by the filter), into a variable direct-voltage level, is then achieved. The discriminator characteristic must, therefore, be linear over a frequency band that is sufficiently wide to accommodate the instantaneous variations in frequency (above and below the level corresponding to the frequency p), in order to preserve the average direct-voltage level of the output at the value dictated by the frequency of the stronger signal.

1.6 DISCRIMINATOR BANDWIDTH REQUIREMENTS

In general, the resultant signal at the output of the limiter narrow-band filter will exhibit instantaneous variations of frequency and envelope. The resultant signal will be of constant amplitude only if it is the sum of all the spectral components on both sides of the central component A_o. When the limiter bandwidth always passes the bulk of the components of significant amplitude, the instantaneous-amplitude variations will usually be insignificant. However, in the narrow-band limiter case, situations in which only a few significant components are passed are likely to occur at all times, and, in a sense, the resultant signal will behave as it would in the presence of multisignal interference.

The narrow-band limiter case will thus, in general, call for a limiter stage to follow the narrow-band filter, when amplitude-sensitive frequency detectors are employed. Even if this second stage of limiting requires theoretically a very wide bandwidth in order to deliver a constant-amplitude signal to the discriminator, it is significant to note (in anticipation of our results) that the combination of one narrow-band limiter, followed by a relatively wideband limiter, will still serve the desirable purpose of reducing, by a sizable amount, the required minimum discriminator bandwidth, in addition to protecting an amplitude-sensitive discriminator from variations in the resultant signal amplitude. If we then assume that an amplitude-insensitive discriminator is used, such a device will only respond to the instantaneous-frequency variations of the resultant signal at its input, and convert these variations into a variable direct-voltage level. For the average value of this voltage level to correspond with that dictated by the frequency of the stronger signal, the instantaneous-frequency swings must be accommodated fully over a linear range of the FM-to-AM conversion characteristic. The required discriminator bandwidth will, thus, be dictated by the expected maximum swing in the instantaneous frequency.

At this point, the question may be raised as to why the discriminator FM-to-AM conversion characteristic must be linear (over the range of the maximum instantaneous-frequency variations) and not some other form of curve which might produce (in response to the sharp and large instantaneous-frequency deviations caused by the interference) a conveniently distorted replica of the undesirable frequency variations that will minimize the over-all effects of the disturbance. Such a curved characteristic might be considerably less expensive to construct and maintain than a straight-line characteristic, and, by smoothing out the sharper variations in the incoming instantaneous-frequency pattern, might even ease up the fast-action requirements on the amplitude-detecting parts of the circuit. Such a question, although it sounds reasonable, really overlooks some fundamental considerations that enter into the mechanism of FM-to-AM conversion. Fundamentally, this conversion takes place when the amplitude-limited FM wave is impressed upon a filter whose amplitude-versus-frequency steady-state characteristic varies with frequency. The type of variation with frequency that this steady-state characteristic must exhibit is dictated by the basic requirement that (at least over the range of the frequency excursions caused by the expected message) the resultant amplitude variations be linearly related to the instantaneous-frequency variations of the message-bearing signal for final undistorted reproduction of the message. Fortunately, the linear variation of the steady-state amplitude characteristic with frequency also ensures that the filter will follow the impressed variable-frequency excitation which carries the expected message through quasi-stationary states over the entire extent of the linearity; thus no noticeable distortion attributable to FM transients can arise. Under interference conditions, the same considerations apply, and now the extent of the linearity in the steady-state amplitude-versus-frequency characteristic must encompass the range of the maximum frequency deviations that the circuit must handle. The necessity of this requirement can be appreciated from the fact that it is not possible to produce an amplitude-response characteristic which is linear within, and has nonlinearity outside the message modulation band so that it will translate an arbitrarily situated (and perhaps arbitrarily distorted) frequency-spike pattern into an amplitude variation that (despite the further distortion in the nonlinear region, and lack of it in the linear region) will still average out to zero (over one cycle of the frequency difference) about a value that corresponds to the level dictated by the frequency of the stronger signal. Moreover, a nonlinear steady-state amplitude characteristic always leads to deviations from the steady-state amplitude response which are larger the greater the degree of nonlinearity and the higher the repetition rate, rate of change, and extent of the instantaneous-frequency variations of the excitation. The result is increased disturbance and no capture. The only characteristic that will satisfy all of the fundamental performance requirements is, therefore, one which varies linearly with frequency over the entire range of instantaneous-frequency excursions that must be handled successfully.

As a first step toward the determination of the variation of the minimum requirements in discriminator bandwidth as a function of the limiter bandwidth we shall study the variation of the instantaneous-frequency pattern of the resultant signal with the number of significant components passed from each sideband. Note that with every prescribed value of limiter bandwidth we may associate certain possible configurations of passed components. For each configuration, the number of components passed (with the desired component, A_o) from each sideband will, generally, depend upon the value of the limiter bandwidth that is used, the frequency difference between the two signals, and the positions of the signals relative to the center frequency of the i-f passband. In view of the insight gained, thus far, about the effect of each sideband upon the character of the resultant signal, it will be appreciated that, of all the different possible configurations, a few can always be spotted by inspection and expected to produce instantaneous-frequency deviations (from the frequency of A_o) of such magnitude that they require greater discriminator bandwidths than the remaining legion of possibilities. Among those few cases, the one configuration that will impose the greatest requirement in discriminator bandwidth can then be determined by direct computation. As a consequence, our problem is thus reduced to one of spotting the most critical situation that may arise with every value of limiter bandwidth that is proposed, and stating the value of discriminator bandwidth dictated by this case as the one that is demanded by the particular limiter bandwidth that is assumed.

Accordingly, we now consider an arbitrary configuration of passed components, and determine by direct computation how the magnitude of the greatest resultant instantaneous deviation (from the frequency of the stronger signal) is affected by the number of components passed from each sideband. This can be done as follows.

As before, if N and M upper and lower sideband components are passed, then the resultant phasor $\overline{OR}$ (in Fig. 6) will be given by

$$\overline{OR} = \sum_{n=-N}^{M} A_n e^{-jnrt}$$

The instantaneous phase angle, θ, which $\overline{OR}$ makes with A_o, will be

$$\theta = \mathrm{Im}\left\{\ln \sum_{n=-N}^{M} A_n e^{-jnrt}\right\} \tag{37}$$

The time derivative of θ will be, therefore, the instantaneous-frequency deviation, from the frequency of the stronger signal, that the resultant signal will experience. Plots of

$$\frac{d\theta}{dt} = y_{M-N}$$

for various values of M and N, and for values of $\underline{a}$ between 0.8 and 0.95, are shown in Fig. 17a, b, and c.

It is clear, from the properties of the interference spectrum (and the discussion of section 1.5), that the situations in which only lower sideband components pass with A_o present no serious discriminator-bandwidth problem, since the maximum deviations in instantaneous frequency that they engender are comparatively small (see Fig. 17c). The most serious situations arise when only upper sideband components are passed. The simultaneous presence of lower sideband components and upper sideband components within the ideal-filter passband results in reduced frequency-spike magnitudes. These conclusions are clearly illustrated by Fig. 17, and will presently be reinforced by the derived expression for the spike magnitude.

The frequency spikes occur at $t = \pi/r$ sec or any odd multiple thereof. From Eq. 37 we have

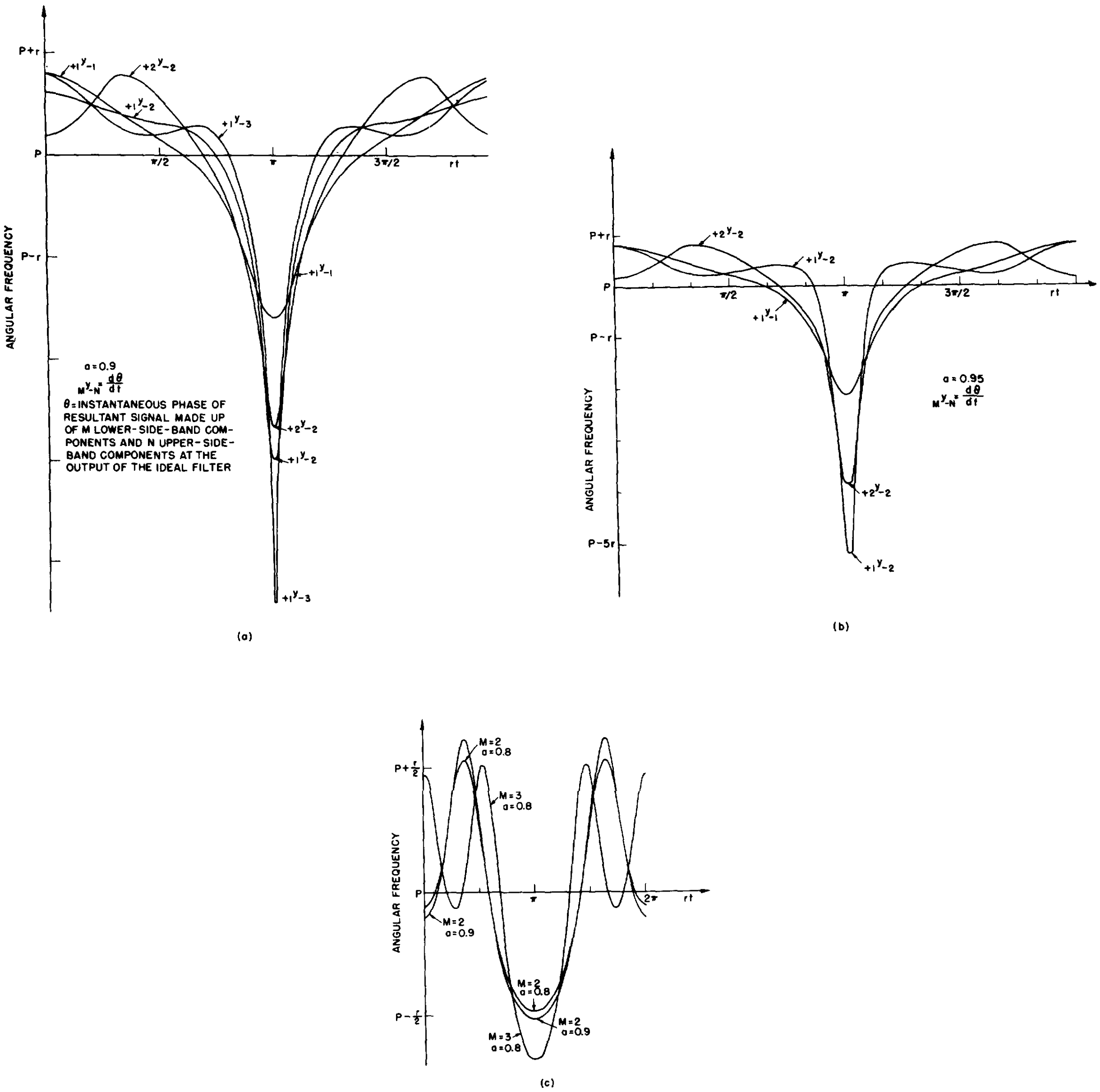

Fig. 17. Variation of the instantaneous frequency of the resultant signal with time when only a few important sideband components are passed by the ideal-limiter filter.

$$\frac{d\theta}{dt} = \mathrm{Im}\left\{\frac{\sum_{n=-N}^{M} (-jnr)\, A_n\, e^{-jnrt}}{\sum_{n=-N}^{M} A_n\, e^{-jnrt}}\right\}$$

or

$$\frac{d\theta}{dt} = \mathrm{Re}\left\{\frac{-r \sum_{n=-N}^{M} nA_n\, e^{-jnrt}}{\sum_{n=-N}^{M} A_n\, e^{-jnrt}}\right\} \tag{38}$$

whence, at $t = \pi/r$, from the properties displayed by the A_n's, we can write

$$-[\Delta\omega] = \left.\frac{d\theta}{dt}\right|_{t=\pi/r} = -r\,\frac{\sum_{n=1}^{M} n|A_n| + \sum_{n=1}^{N} n|A_{-n}|}{\sum_{n=0}^{M} |A_n| - \sum_{n=1}^{N} |A_{-n}|}$$

where $[\Delta\omega]$ is the magnitude of the frequency spike. This expression demonstrates clearly the effect that the components from each sideband will have upon the magnitude of the frequency spike. Finally, we have

$$\underset{M-N}{\lambda} = \frac{[\Delta\omega]}{r} = \frac{\sum_{n=1}^{M} n|A_n| + \sum_{n=1}^{N} n|A_{-n}|}{\sum_{n=0}^{M} |A_n| - \sum_{n=1}^{N} |A_{-n}|} \tag{39}$$

Table IV presents values of $[\Delta\omega]/r$ for the more serious configurations that will be encountered in the course of the present investigation. Given the magnitude of the frequency spike that arises with a given configuration of components, the discriminator bandwidth that is necessary to accommodate the whole spike pattern can be calculated readily. Thus, with due consideration to the possibility of an interchange of the relative signal magnitudes or frequencies, the required discriminator bandwidth is given by by

$$\begin{aligned}(BW)_{disc} &= 2[\Delta\omega] + r \\ &= r\left[2\,\frac{[\Delta\omega]}{r} + 1\right]\end{aligned} \tag{40}$$

where r is the frequency difference between the two signals, in rad/sec. Equation 40

Table IV

$$\lambda_{M-N} = [\Delta\omega]/r$$

a	λ_{0-1}	λ_{0-2}	λ_{0-3}	λ_{0-4}	λ_{0-5}	λ_{1-1}	λ_{1-2}
0.7	0.77506	1.2512	1.5349	1.7058	1.8106	0.8567	1.1506
0.8	1.2051	2.3860	3.3855	4.1736	4.7738	1.1156	1.6730
0.85	1.6014	3.8929	6.7234	9.9066	13.240	1.3099	2.1395
0.9	2.3357	9.3568	58.44			1.6009	2.9883
0.95						2.1305	5.2270

a	λ_{1-3}	λ_{2-2}	λ_{2-3}	λ_{2-4}	λ_{2-5}	λ_{3-4}	λ_{3-5}
0.7	1.3124	1.2943	1.4325	1.5120	1.5600	1.6469	1.6913
0.8	2.0573	1.7472	2.0435	2.2460	2.3886	2.3509	2.4751
0.85	2.8052	2.1024	2.5715	2.9265	3.1990	2.9392	3.1637
0.9	4.3935	2.6715	3.5127	4.2437	4.8764	3.9654	4.4347
0.95	10.375	3.8704	5.9510	8.3612	11.1185	6.5735	8.0629

a	λ_{4-5}	λ_{4-6}	λ_{5-6}	λ_{5-7}	λ_{6-7}	λ_{6-8}	λ_{7-8}
0.7	1.8146	1.8410	1.9435	1.9597	2.0419	2.0521	2.1165
0.8	2.6108	2.6923	2.8305	2.8870	3.0176	3.0577	3.1765
0.85	3.2594	3.4136	3.5413	3.6525	3.7909	3.8736	4.0123
0.9	4.3685	4.7001	4.7333	4.9813	5.0663	5.2578	5.3721
0.95	7.1312	8.1812	7.6419	8.4368	8.1165	8.7457	8.5616

a	λ_{7-9}	λ_{8-9}	λ_{8-10}	λ_{9-10}	λ_{9-11}	λ_{10-11}	λ_{10-12}
0.7	2.1231	2.1726	2.1770	2.2148	2.2176	2.2461	2.2480
0.8	3.2056	3.3113	3.3326	3.4252	3.4437	3.5213	3.5332
0.85	4.0753	4.2090	4.2576	4.3836	4.4217	4.5385	4.5688
0.9	5.5239	5.6538	5.7766	5.9140	6.0147	6.1547	6.2380
0.95	9.0746	8.9815	9.4093	9.3795	9.7426	9.7585	10.071

Table V

$$\beta_{M-N} = 2\lambda_{M-N} + 1$$

a	β_{0-1}	β_{0-2}	β_{0-3}	β_{0-4}	β_{0-5}	β_{1-1}	β_{1-2}
0.7	2.5501	3.5024	4.0698	4.4116	4.6212	2.7134	3.3012
0.8	3.4102	5.7720	7.7710	9.3472	10.5476	3.2312	4.3460
0.85	4.2028	8.7858	14.4468	20.813	27.480	3.6198	5.2790
0.9	5.6714	19.7136	117.88			4.2018	6.9766
0.95						5.2610	11.454

a	β_{1-3}	β_{2-2}	β_{2-3}	β_{2-4}	β_{2-5}	β_{3-4}	β_{3-5}
0.7	3.6248	3.5886	3.8650	4.0240	4.1200	4.2938	4.3826
0.8	5.1146	4.4944	5.0870	5.4920	5.7773	5.7018	5.9502
0.85	6.6104	5.2048	6.1430	6.8530	7.3980	6.8784	7.3274
0.9	9.7870	6.3430	8.0254	0.4874	10.7527	8.9308	9.8694
0.95	22.470	8.7408	12.902	17.7224	23.2370	14.147	17.126

a	β_{4-5}	β_{4-6}	β_{5-6}	β_{5-7}	β_{6-7}	β_{6-8}	β_{7-8}
0.7	4.6292	4.6820	4.8870	4.9194	5.0838	5.1042	5.2330
0.8	6.2216	6.3846	6.6610	6.7740	7.0352	7.1154	7.3530
0.85	7.5188	7.8272	8.0826	8.3050	8.5818	8.7472	9.0246
0.9	9.7370	10.400	10.4666	10.9626	11.1326	11.5156	11.7442
0.95	15.262	17.362	16.284	17.874	17.233	18.491	18.123

a	β_{7-9}	β_{8-9}	β_{8-10}	β_{9-10}	β_{9-11}	β_{10-11}	β_{10-12}
0.7	5.2462	5.3452	5.3540	5.4296	5.4352	5.4922	5.4960
0.8	7.4112	7.6226	7.6652	7.8504	7.8874	8.0426	8.0664
0.85	9.1506	9.4180	9.5152	9.7672	9.8434	10.0770	10.1376
0.9	12.0478	12.3076	12.5532	12.8280	13.0294	13.3094	13.476
0.95	19.149	18.963	19.819	19.759	20.485	20.517	21.142

can be written in the more convenient form

$$\frac{(BW)_{disc}}{(BW)_{if}} = \frac{r}{(BW)_{if}} \beta_{M-N} = \delta \cdot \beta_{M-N} \tag{41}$$

The use of Table IV in conjunction with the expression

$$\beta_{M-N} = 2 \frac{[\Delta\omega]}{r} + 1$$

produces the values given in Table V.

In Table VI, we have given a set of possible values of $(BW)_{lim}/(BW)_{if}$, together with the most serious situations that may arise with such limiter-bandwidth values, and the corresponding maximum values δ_m that the normalized frequency difference, $\delta = r/(BW)_{if}$, can assume and still make it possible for these situations to arise. The tabulated values of δ_m can be calculated in the following way. From Fig. 14 we find that N upper sideband components can pass as long as

$$r < r_m = \frac{\frac{(BW)_{lim} - (BW)_{if}}{2} + (BW)_{if}}{N}$$

whence

$$\delta_m = \frac{r_m}{(BW)_{if}} = \frac{1}{N}\left[\frac{1}{2}\left\{\frac{(BW)_{lim}}{(BW)_{if}} - 1\right\} + 1\right]$$

$$= \frac{1}{2N}\left[\frac{(BW)_{lim}}{(BW)_{if}} + 1\right] \tag{42}$$

Clearly, the maximum value that δ_m may assume is unity.

In the determination of the more critical situations that may arise with the different prescribed values of limiter bandwidth, we first allot the maximum space in the limiter passband to the components from the upper sideband. Then we weigh a particular situation in the light of the maximum value of δ_m beyond which the situation cannot arise, and the value of β_{M-N} associated with it. The situation that requires the largest value of

$$\frac{(BW)_{disc}}{(BW)_{if}} = \delta_m \cdot \beta_{M-N} \tag{43}$$

dictates the requirement in discriminator bandwidth.

Finally, values from Tables V and VI have been used in conjunction with Eq. 43 to construct Table VII. The entries followed by double daggers in Table VII are taken as

Table VI

$$\delta_m = \delta_{max} = r_m/(BW)_{if}$$

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	δ_m	M	N	δ_m	M	N	δ_m
1	0	1	1						
1.4	0	2	0.6	0	5	0.24	0	3	0.4
								4	0.3
1.5	0	2	0.625	0	4	0.3125	0	3	0.417
1.667	0	2	0.667	0	3	0.444			
2	0	2	0.75	1	3	0.5			
2.25	0	2	0.8125	1	3	0.5417			
2.5	0	2	0.875	1	3	0.583			
2.8	0	2	0.95	1	3	0.633			
3	1	2	1	1	3	0.667	2	4	0.5
3.5	1	2	1	1	3	0.75	2	4	0.5625
4	1	2	1	1	3	0.833	2	4	0.625
4.5	1	2	1	1	3	0.9167	2	4	0.6875
5	2	3	1	2	4	0.75	3	5	0.6
5.5	2	3	1	2	4	0.8125	3	5	0.65
6	2	3	1	2	4	0.875	3	5	0.7
6.5	2	3	1	2	4	0.9375	3	5	0.75
7	3	4	1	3	5	0.8	4	6	0.667
7.5	3	4	1	3	5	0.85	4	6	0.708
8	3	4	1	3	5	0.9	4	6	0.75
8	3	4	1	3	5	0.9	4	6	0.75
8.5	3	4	1	3	5	0.95	4	6	0.791
9	4	5	1	4	6	0.833	5	7	0.714
9.5	4	5	1	4	6	0.875	5	7	0.75
10	4	5	1	4	6	0.9167	5	7	0.7857
10.5	4	5	1	4	6	0.958	5	7	0.821
11	5	6	1	5	7	0.857	6	8	0.75
11.5	5	6	1	5	7	0.893	6	8	0.781
12	5	6	1	5	7	0.929	6	8	0.8125
12.5	5	6	1	5	7	0.964	6	8	0.834
13	6	7	1	6	8	0.875	7	9	0.778
13.5	6	7	1	6	8	0.906	7	9	0.806
14	6	7	1	6	8	0.9375	7	9	0.833
14.5	6	7	1	6	8	0.969	7	9	0.861
15	7	8	1	7	9	0.889	8	10	0.8
15.5	7	8	1	7	9	0.917	8	10	0.825
16	7	8	1	7	9	0.944	8	10	0.85
16.5	7	8	1	7	9	0.972	8	10	0.875
17	8	9	1	8	10	0.9	9	11	0.818
17.5	8	9	1	8	10	0.925	9	11	0.841
18	8	9	1	8	10	0.95	9	11	0.864
18.5	8	9	1	8	10	0.975	9	11	0.887
19	9	10	1	9	11	0.909	10	12	0.833
19.5	9	10	1	9	11	0.932	10	12	0.854
20	9	10	1	9	11	0.954	10	12	0.875
20.5	9	10	1	9	11	0.977	10	12	0.896
21	10	11	1	10	12	0.916	11	13	0.846

Table VII

a = 0.7

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$
1	0	1	2.55012 ‡						
1.5	0	2	2.1890	0	4	1.3786			
1.9	0	2	2.550 ‡						
2	0	2	2.6268 ‡	1	3	1.8124			
2.25	0	2	2.8457 ‡	1	3	1.964			
2.5	0	2	3.0646 ‡	1	3	2.11			
2.8	0	2	3.3273 ‡	1	3	2.29			
3	1	2	3.3012 ‡	1	3	2.42	2	4	2.0120
3.5	1	2	3.3012 ‡	1	3	2.7186	2	4	2.2635
4	1	2	3.3012 ‡	1	3	3.02	2	4	2.5150
4.46	1	2	3.3012 ‡	1	3	3.3012 ‡			
4.5	1	2	3.3012	1	3	3.323 ‡	2	4	2.7665
5	2	3	3.8650 ‡	2	4	3.018	3	5	2.6206
5.5	2	3	3.8650 ‡	2	4	3.2695	3	5	2.8487
6	2	3	3.8650 ‡	2	4	3.5210	3	5	3.0678
6.5	2	3	3.8650 ‡	2	4	3.7725	3	5	3.2870
6.684	2	3	3.8650 ‡	2	4	3.8650 ‡			
7	3	4	4.2938 ‡	3	5	3.5061	4	6	3.12
7.5	3	4	4.2938 ‡	3	5	3.7252	4	6	3.31
8	3	4	4.2938 ‡	3	5	3.9433	4	6	3.5115
8.5	3	4	4.2938 ‡	3	5	4.1635	4	6	3.70
8.8	3	4	4.2938 ‡	3	5	4.2938 ‡			
9	4	5	4.6292 ‡	4	6	3.90	5	7	3.51
9.5	4	5	4.6292 ‡	4	6	4.0968	5	7	3.6896
10	4	5	4.6292 ‡	4	6	4.292	5	7	3.865
10.5	4	5	4.6292 ‡	4	6	4.485	5	7	4.038
10.9	4	5	4.6292 ‡	4	6	4.6292 ‡			
11	5	6	4.8870 ‡	5	7	4.22	6	8	3.83
11.5	5	6	4.8870 ‡	5	7	4.39	6	8	3.9864
12	5	6	4.8870 ‡	5	7	4.57	6	8	4.15
12.5	5	6	4.8870 ‡	5	7	4.74	6	8	4.26
12.9	5	6	4.8870 ‡	5	7	4.8870 ‡			
13	6	7	5.0838 ‡	6	8	4.4662	7	9	4.08
13.5	6	7	5.0838 ‡	6	8	4.62	7	9	4.23
14	6	7	5.0838 ‡	6	8	4.7852	7	9	4.37
14.5	6	7	5.0838 ‡	6	8	4.95	7	9	4.52
15	7	8	5.2330 ‡	7	9	4.66	8	10	4.2832
15.5	7	8	5.2330 ‡	7	9	4.81	8	10	4.4171
16	7	8	5.2330 ‡	7	9	4.95	8	10	4.5509
16.5	7	8	5.2330 ‡	7	9	5.10	8	10	4.6848
17	8	9	5.3452 ‡	8	10	4.8186	9	11	4.45
17.5	8	9	5.3452 ‡	8	10	4.9525	9	11	4.57
18	8	9	5.3452 ‡	8	10	5.0863	9	11	4.70
18.5	8	9	5.2452 ‡	8	10	5.2202	9	11	4.82
19	9	10	5.4296 ‡	9	11	4.94	10	12	4.58
19.5	9	10	5.4296 ‡	9	11	5.07	10	12	4.69
20	9	10	5.4296 ‡	9	11	5.19	10	12	4.8090
20.5	9	10	5.4296 ‡	9	11	5.31	10	12	4.92
21	10	11	5.4922 ‡	10	12	5.03			

Table VII (cont'd.)

a = 0.8

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$
1	0	1	3.4102 ‡						
1.5	0	2	3.6075 ‡	0	4	2.9210			
2	0	2	4.3290 ‡	1	3	2.5573			
2.25	0	2	4.6898 ‡	1	3	2.771			
2.5	0	2	5.0505 ‡	1	3	2.98			
2.8	0	2	5.4834 ‡	1	3	3.24			
3	1	2	4.3460 ‡	1	3	3.41	2	4	2.7460
3.5	1	2	4.3460 ‡	1	3	3.8360	2	4	3.0893
4	1	2	4.3460 ‡	1	3	4.26	2	4	3.4325
4.1	1	2	4.3460 ‡	1	3	4.3460 ‡			
4.5	1	2	4.3460	1	3	4.589 ‡	2	4	3.7758
5	2	3	5.0870 ‡	2	4	4.1190	3	5	3.5701
5.5	2	3	5.0870 ‡	2	4	4.4623	3	5	3.8676
6	2	3	5.0870 ‡	2	4	4.8055	3	5	4.1651
6.41	2	3	5.0870 ‡	2	4	5.0870 ‡			
6.5	2	3	5.0870	2	4	5.1488 ‡	3	5	4.4627
7	3	4	5.7018 ‡	3	5	4.7601	4	6	4.26
7.5	3	4	5.7018 ‡	3	5	5.0577	4	6	4.52
8.0	3	4	5.4018 ‡	3	5	5.3552	4	6	4.7885
8.5	3	4	5.7018 ‡	3	5	5.6527	4	6	5.05
8.58	3	4	5.7018 ‡	3	5	5.7018 ‡			
9	4	5	6.2216 ‡	4	6	5.32	5	7	4.84
9.5	4	5	6.2216 ‡	4	6	5.5865	5	7	5.0805
10	4	5	6.2216 ‡	4	6	5.853	5	7	5.322
10.5	4	5	6.2216 ‡	4	6	6.12	5	7	5.56
10.69	4	5	6.2216 ‡	4	6	6.2216 ‡			
11	5	6	6.6610 ‡	5	7	5.81	6	8	5.3366
11.5	5	6	6.6610 ‡	5	7	6.05	6	8	5.56
12	5	6	6.6610 ‡	5	7	6.29	6	8	5.781
12.5	5	6	6.6610 ‡	5	7	6.53	6	8	5.93
12.76	5	6	6.6610 ‡	5	7	6.6610 ‡			
13	6	7	7.0352 ‡	6	8	6.2260	7	9	5.77
13.5	6	7	7.0352 ‡	6	8	6.45	7	9	5.97
14	6	7	7.0352 ‡	6	8	6.6707	7	9	6.17
14.5	6	7	7.0352 ‡	6	8	6.89	7	9	6.38
14.8	6	7	7.0352 ‡	6	8	7.0352			
15	7	8	7.3530 ‡	7	9	6.59	8	10	7.1322
15.5	7	8	7.3530 ‡	7	9	6.80	8	10	6.3238
16	7	8	7.3530 ‡	7	9	7.00	8	10	6.5154
16.5	7	8	7.3530 ‡	7	9	7.20	8	10	6.7071
16.86	7	8	7.3530 ‡	7	9	7.3530 ‡			
17	8	9	7.6226 ‡	8	10	6.8987	9	11	6.45
17.5	8	9	7.6226 ‡	8	10	7.0903	9	11	6.63
18	8	9	7.6226 ‡	8	10	7.2819	9	11	6.81
18.5	8	9	7.6226 ‡	8	10	7.4736	9	11	7.00
19	9	10	7.8504 ‡	9	11	7.17	10	12	6.72
19.5	9	10	7.8504 ‡	9	11	7.35	10	12	6.89
20	9	10	7.8504 ‡	9	11	7.52	10	12	7.0581
20.5	9	10	7.8504 ‡	9	11	7.71	10	12	7.23
21	10	11	8.0426 ‡	10	12	7.39			

Table VII (cont'd.)

a = 0.85

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$
1	0	7	5.76 ‡						
1.4	0	2	5.2715	0	5	6.5952 ‡			
1.5	0	2	5.4911	0	4	6.5041 ‡			
1.667	0	2	5.86	0	3	6.41 ‡			
2	0	2	6.5894 ‡	1	3	3.3052			
2.5	0	2	7.6876 ‡	1	3	3.85			
3	1	2	5.4790 ‡	1	3	4.41	2	4	3.4265
3.5	1	2	5.2790 ‡	1	3	4.9576	2	4	3.8548
4	1	2	5.2790	1	3	5.51 ‡	2	4	4.2831
4.5	1	2	5.2790	1	3	6.060 ‡	2	4	4.7114
5	2	3	6.1430 ‡	2	4	5.1398	3	5	4.3964
5.5	2	3	6.1430 ‡	2	4	5.5681	3	5	4.7628
6	2	3	6.1430 ‡	2	4	5.9964	3	5	5.1292
6.17	2	3	6.1430 ‡	2	4	6.1430 ‡			
6.5	2	3	6.1430	2	4	6.4247 ‡	3	5	5.4956
7	3	4	6.8784 ‡	3	5	5.8619	4	6	5.22
7.5	3	4	6.8784 ‡	3	5	6.2283	4	6	5.54
8	3	4	6.8784 ‡	3	5	6.6947	4	6	5.8704
8.39	3	4	6.8784 ‡	3	5	6.8784 ‡			
8.5	3	4	6.8784	3	5	6.9610 ‡	4	6	6.19
9	4	5	7.5188 ‡	4	6	6.52	5	7	5.93
9.5	4	5	7.5188 ‡	4	6	6.8488	5	7	6.2288
10	4	5	7.5188 ‡	4	6	7.175	5	7	6.525
10.5	4	5	7.5188 ‡	4	6	7.50	5	7	6.82
10.53	4	5	7.5188 ‡	4	6	7.5188 ‡			
11	5	6	8.0826 ‡	5	7	7.12	6	8	6.5604
11.5	5	6	8.0826 ‡	5	7	7.42	6	8	6.83
12	5	6	8.0826 ‡	5	7	7.72	6	8	7.107
12.5	5	6	8.0826 ‡	5	7	8.01	6	8	7.30
12.62	5	6	8.0826 ‡	5	7	8.0826 ‡			
13	6	7	8.5818 ‡	6	8	7.6538	7	9	7.12
13.5	6	7	8.5818 ‡	6	8	7.92	7	9	7.38
14	6	7	8.5818 ‡	6	8	8.2005	7	9	7.62
14.5	6	7	8.5818 ‡	6	8	8.48	7	9	7.88
14.7	6	7	8.5818 ‡	6	8	8.5818 ‡			
15	7	8	9.0246 ‡	7	9	8.13	8	10	7.6122
15.5	7	8	9.0246 ‡	7	9	8.39	8	10	7.8500
16	7	8	9.0246 ‡	7	9	8.64	8	10	8.0879
16.5	7	8	9.0246 ‡	7	9	8.89	8	10	8.3258
16.75	7	8	9.0246 ‡	7	9	9.0246 ‡			
17	8	9	9.4180 ‡	8	10	8.5637	9	11	8.05
17.5	8	9	9.4180 ‡	8	10	8.8016	9	11	8.28
18	8	9	9.4180 ‡	8	10	9.0394	9	11	8.50
18.5	8	9	9.4180 ‡	8	10	9.2773	9	11	8.73
18.80	8	9	9.4180 ‡	8	10	9.4180 ‡			
19	9	10	9.7672 ‡	9	11	8.95	10	12	8.44
19.5	9	10	9.7672 ‡	9	11	9.17	10	12	8.66
20	9	10	9.7672 ‡	9	11	9.39	10	12	8.8704
20.5	9	10	9.7672 ‡	9	11	9.62	10	12	9.08
21	10	11	10.077 ‡	10	12	9.29			

Table VII (cont'd.)

a = 0.9

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$
1.667	0	2	13.1	0	3	52.34 ‡			
2	0	2	14.7 ‡	1	3	4.8935			
2.5	0	2	17.249 ‡	1	3	5.71			
3	1	2	6.9766 ‡	1	3	6.53	2	4	4.7437
3.28	1	2	6.9766 ‡	1	3	6.9766 ‡			
3.5	1	2	6.9766	1	3	7.34 ‡	2	4	5.3367
4	1	2	6.9766	1	3	8.15 ‡	2	4	5.9296
4.5	1	2	6.9766	1	3	8.972 ‡	2	4	6.5226
5	2	3	8.0254 ‡	2	4	7.1156	3	5	5.9216
5.5	2	3	8.0254 ‡	2	4	7.7085	3	5	6.4151
5.77	2	3	8.0254 ‡	2	4	8.0254 ‡	3	5	
6	2	3	8.0254	2	4	8.3015 ‡	3	5	6.9086
6.5	2	3	8.0254	2	4	8.8944 ‡	3	5	7.4021
7	3	4	8.9308 ‡	3	4	7.8955	4	6	6.94
7.5	3	4	8.9308 ‡	3	5	8.3890	4	6	7.36
8	3	4	8.9308 ‡	3	5	8.8825	4	6	7.8002
8.05	3	4	8.9308 ‡	3	5	8.9308 ‡			
8.5	3	4	8.9308	3	5	9.3759 ‡	4	6	8.23
9	4	5	9.7370 ‡	4	6	8.66	5	7	7.83
9.5	4	5	9.7370 ‡	4	6	9.1002	5	7	8.2220
10.0	4	5	9.7370 ‡	4	6	9.534	5	7	8.613
10.24	4	5	9.7370 ‡	4	6	9.7370 ‡			
10.5	4	5	9.7370	4	6	9.96 ‡	5	7	9.00
11	5	6	10.467 ‡	5	7	9.39	6	8	8.6367
11.5	5	6	10.467 ‡	5	7	9.79	6	8	8.99
12	5	6	10.467 ‡	5	7	10.2	6	8	9.3564
12.37	5	6	10.467 ‡	5	7	10.467 ‡			
12.5	5	6	10.467	5	7	10.6 ‡	6	8	9.60
13	6	7	11.133 ‡	6	8	10.076	7	9	9.37
13.5	6	7	11.133 ‡	6	8	10.433	7	9	9.71
14	6	7	11.133 ‡	6	8	10.796	7	9	10.0
14.46	6	7	11.133 ‡	6	8	11.133 ‡			
14.5	6	7	11.133 ‡	6	8	11.2 ‡	7	9	10.4
15	7	8	11.744 ‡	7	9	10.7	8	10	10.043
15.5	7	8	11.744 ‡	7	9	11.0	8	10	10.356
16	7	8	11.744 ‡	7	9	11.4	8	10	10.670
16.5	7	8	11.744 ‡	7	9	11.73	8	10	10.984
16.54	7	8	11.744 ‡	7	9	11.744 ‡			
17	8	9	12.308 ‡	8	10	11.298	9	11	10.7
17.5	8	9	12.308 ‡	8	10	11.612	9	11	11.0
18	8	9	12.308 ‡	8	10	11.926	9	11	11.3
18.5	8	9	12.308 ‡	8	10	12.239	9	11	11.6
18.61	8	9	12.308 ‡	8	10	12.308 ‡			
19	9	10	12.828 ‡	9	11	11.8	10	12	11.2
19.5	9	10	12.828 ‡	9	11	12.1	10	12	11.5
20	9	10	12.828 ‡	9	11	12.4	10	12	11.792
20.5	9	10	12.828 ‡	9	11	12.7	10	12	12.1
20.66	9	10	12.828 ‡	9	11	12.828 ‡			
21	10	11	13.309 ‡	10	12	12.3			

Table VII (cont'd.)

a = 0.95

$\frac{(BW)_{lim}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$	M	N	$\frac{(BW)_{disc}}{(BW)_{if}}$
3	1	2	11.454	1	3	15.0 ‡	2	4	8.8612
3.5	1	2	11.454	1	3	16.853 ‡	2	4	9.9689
4	1	2	11.454	1	3	18.7 ‡	2	4	11.077
4.5	1	2	11.454	1	3	20.60 ‡	2	4	12.184
5	2	3	12.902	2	4	13.292 ‡	3	5	10.276
5.5	2	3	12.902	2	4	14.399 ‡	3	5	11.132
6	2	3	12.902	2	4	15.507 ‡	3	5	11.988
6.5	2	3	12.902	2	4	16.615 ‡	3	5	12.845
7	3	4	14.147 ‡	3	5	13.701	4	6	11.6
7.26	3	4	14.147 ‡	3	5	14.147 ‡			
7.5	3	4	14.147	3	5	14.557 ‡	4	6	12.3
8	3	4	14.147	3	5	15.413 ‡	4	6	13.022
8.5	3	4	14.147	3	5	16.270 ‡	4	6	13.7
9	4	5	15.262 ‡	4	6	14.5	5	7	12.8
9.5	4	5	15.262 ‡	4	6	15.192	5	7	13.406
9.55	4	5	15.262 ‡	4	6	15.262 ‡			
10	4	5	15.262	4	6	15.92 ‡	5	7	14.044
10.5	4	5	15.262	4	6	16.6 ‡	5	7	14.7
11	5	6	16.284 ‡	5	7	15.3	6	8	13.868
11.5	5	6	16.284 ‡	5	7	16.0	6	8	14.4
11.76	5	6	16.284 ‡	5	7	16.284 ‡			
12	5	6	16.284	5	7	16.6	6	8	15.024
12.5	5	6	16.284	5	7	17.2 ‡	6	8	15.4
13	6	7	17.233 ‡	6	8	16.180	7	9	14.9
13.5	6	7	17.233 ‡	6	8	16.8	7	9	15.4
13.91	6	7	17.233 ‡	6	8	17.233 ‡			
14	6	7	17.233	6	8	17.335 ‡	7	9	16.0
14.5	6	7	17.233	6	8	17.9 ‡	7	9	16.5
15	7	8	18.123 ‡	7	9	17.0	8	10	15.855
15.5	7	8	18.123 ‡	7	9	17.6	8	10	16.351
16	7	8	18.123 ‡	7	9	18.1	8	10	16.846
16.04	7	8	18.123 ‡	7	9	18.123 ‡			
16.5	7	8	18.123	7	9	18.6 ‡	8	10	17.342
17	8	9	18.963 ‡	8	10	17.837	9	11	16.8
17.5	8	9	18.963 ‡	8	10	18.333	9	11	17.2
18	8	9	18.963 ‡	8	10	18.828	9	11	17.7
18.14	8	9	18.963 ‡	8	10	18.963 ‡			
18.5	8	9	18.963	8	10	19.324 ‡	9	11	18.2
19	9	10	19.759 ‡	9	11	18.6	10	12	17.6
19.5	9	10	19.759 ‡	9	11	19.1	10	12	18.1
20	9	10	19.759 ‡	9	11	19.5	10	12	18.499
20.22	9	10	19.759 ‡	9	11	19.759 ‡			
20.5	9	10	19.759	9	11	20.0 ‡	10	12	18.9
21	10	11	20.517 ‡	10	12	19.4			

Table VIII. Case in which $(BW)_{lim} = (BW)_{if}$ and $r_m = \frac{(BW)_{if}}{N}$.

N	a = 0.8 $\frac{[\Delta\omega]}{r}$	a = 0.8 $\frac{[\Delta\omega]}{(BW)_{if}}$	a = 0.8 $\frac{(BW)_{disc}}{(BW)_{if}}$	a = 0.85 $\frac{[\Delta\omega]}{r}$	a = 0.85 $\frac{[\Delta\omega]}{(BW)_{if}}$	a = 0.85 $\frac{(BW)_{disc}}{(BW)_{if}}$
1	1.2	1.2	3.4	1.6	1.6	4.2
2	2.386	1.193	2.9	3.893	1.947	4.4
3	3.386	1.129	2.6	6.722	2.241	4.82
4	4.174	1.043	2.34	9.907	2.477	5.21
5	4.774	0.955	2.11	13.25	2.65	5.5
6	---	---	---	16.54	2.76	5.68
7	---	---	---	19.65	2.81	5.76
8	---	---	---	22.48	2.81	5.75
9	---	---	---	24.97	2.77	5.66
10	---	---	---	27.12	2.71	5.52

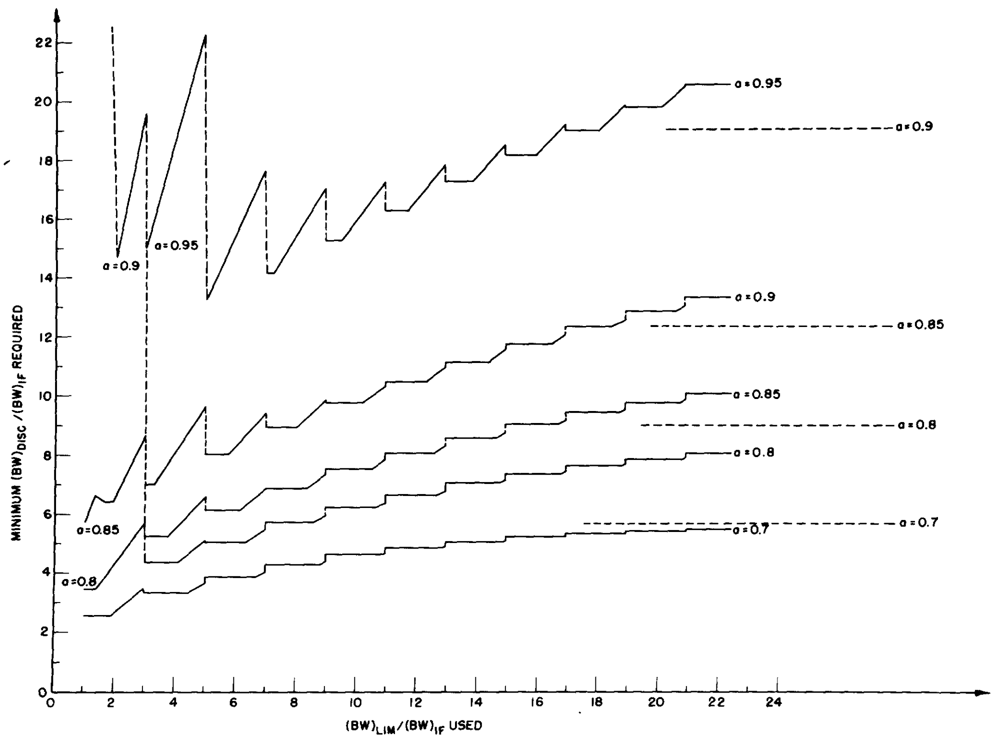

Fig. 18. Calculated values of the required minimum discriminator bandwidth as a function of the bandwidth of the ideal-limiter filter, when the discriminator is preceded by only one stage of ideal narrow-band limiting.

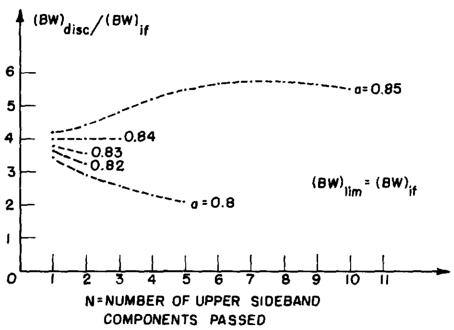

Fig. 19. Variation of the required minimum discriminator bandwidth with the number of upper sideband components passed by the ideal-limiter filter when the filter bandwidth equals the i-f bandwidth.

the minimum discriminator bandwidths that are required with the corresponding specified values of limiter bandwidth. The results, read from the daggered entries, have been plotted in Fig. 18.

Next, before embarking on a study of the plots in Fig. 18, let us consider the case in which the limiter bandwidth is just one i-f bandwidth, since this case deserves a special treatment. It is clear that this case applies only for $a \leq 0.863$. Here the situations in which only upper sideband components are passed with A_o will impose the greatest requirements in discriminator bandwidth; so, only such cases need be considered. Accordingly, in order for N upper sideband components to pass, the frequency difference r must have a maximum value of $r_m = (BW)_{if}/N$. Values of $[\Delta\omega]/(BW)_{if}$ showing how the effective frequency spike magnitude changes with N are presented in Table VIII. Also presented in Table VIII is a column showing which value of N requires the greatest bandwidth in the discriminator. The values of $(BW)_{disc}/(BW)_{if}$ for values of a between $a = 0.8$ and $a = 0.85$, plotted against N, are also shown in Fig. 19.

As for $a = 0.85$, Table VIII and Fig. 19 show that the required discriminator bandwidth rises with N, until it peaks, for $N = 7$, at the value of $5.76(BW)_{if}$, and then starts to decline. This ultimate decline is due to the fact that, beyond $N = 7$, the decrease in the maximum r with N wrests control of the requirement from the increase of $[\Delta\omega]$ with N. It is also seen, from Table VIII and Fig. 19, that, for all a's up to $a = 0.84$, the case in which only one upper sideband component is passed imposes the required value of the discriminator bandwidth.

The fact that, for all a's up to $a = 0.84$, the bandwidth requirement is set by the case in which only A_o and A_{-1} are passed, calls forth some interesting observations. The observation was first made by Granlund (1) that the table of spectral amplitudes shows a ratio, A_{-1}/A_o, which is always less than the corresponding a. In fact, for $a < 0.5$, the plot of A_{-1}/A_o versus a, shown in Fig. 20, shows that

$$\frac{A_{-1}(a)}{A_o(a)} \approx (1/2)\,a$$

and for other values of a, A_{-1}/A_o is always less than a, and approaches a only as $a \to 1$. This observation can also be directly confirmed from an analysis in terms of the expressions presented in section 1.2. Thus, if the limiter bandwidth passes only A_o and A_{-1} in the worst situation (i.e., the situation that dictates the required discriminator bandwidth), then, under the most critical condition, the resultant output will still be a superposition of two signals at the frequencies of the input signals, but then the ratio of weaker-to-stronger signal will be lower than before. This means that a reduction of the effective a has been achieved in the process. Consequently, if the process of limiting and filtering (with one i-f bandwidth) is repeated often enough, it should be possible to reduce the relative strength of the interfering signal to a negligible value and, consequently, to reduce the required discriminator bandwidth to the value of one intermediate-frequency bandwidth.

The condition for the success of this cascading scheme is seen to hinge upon the requirement that the configuration in which only A_0 and A_{-1} are passed must be the most serious one that can arise. From Fig. 19, we note that the configuration A_0, A_{-1} is the most critical one for all a's up to approximately a = 0.84. Also, according to the results of section 1.5, with a limiter bandwidth value of one i-f bandwidth, the configuration A_0, A_{-1} will remain the most critical one for all values of $a < 1$ as long as the frequency difference r is greater than one-half the i-f bandwidth. When r takes values that are equal to or less than one-half the i-f bandwidth, more than one upper sideband component will pass through. For a greater than 0.84, the argument breaks down for one of two reasons. Either the average frequency of the resultant signal at the output of the limiter filter will not always be equal to that of the stronger signal, or the most unfavorable configuration will involve more upper sideband components than just A_{-1}.

In accordance with the above scheme, Table IX shows how the required ratio $(BW)_{disc}/(BW)_{if}$, for a = 0.8, decreases with the number of cascaded limiter-filter stages. These numbers are also plotted in Fig. 21, where a similar plot for a = 0.7 is also shown.

Table IX

Number of limiter-filter stages	1	2	3	4	5	6	"Infinitely" wide-band limiter
$(BW)_{disc}/(BW)_{if}$ for a = 0.8	3.41	1.9	1.38	1.17	1.08	1.04	9

Returning to the plots of Fig. 18, we note that two important observations are clearly brought out. The first observation is that the minimum requirements on the discriminator bandwidth are always less than, but approach asymptotically, the values previously specified (1) by the formula

$$\frac{(BW)_{disc}}{(BW)_{if}} = \frac{1 + a}{1 - a} \tag{44}$$

It is, indeed, perfectly plausible that in the limit, as the limiter bandwidth becomes very large, the required minimum discriminator bandwidth should approach the asymptotes (shown dotted in Fig. 18) specified by Eq. 44. For, as the limiter bandwidth becomes very large, essentially all of the significant sideband components are passed, and the resultant signal at the output of the limiter filter approaches the amplitude-limited value of the resultant of the two signals delivered by the intermediate-frequency amplifier to the ideal limiter. Since the ideal-limiter action *per se* does not affect the instantaneous variations in the frequency of the resultant signal, the values specified by Eq. 44 become the limiting values that are approached as the limiter bandwidth

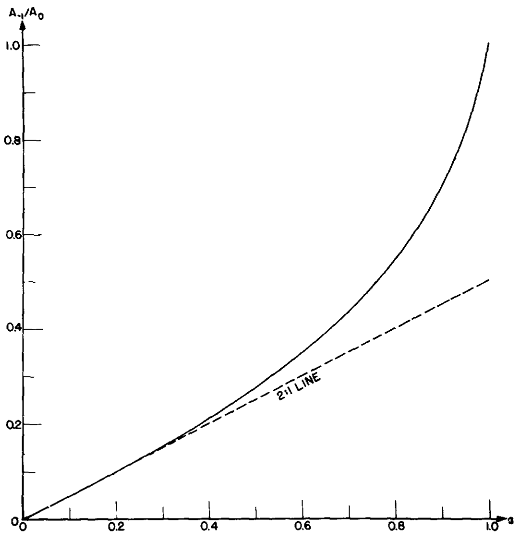

Fig. 20. Effect of the ideal amplitude-limiting process upon the amplitude, $A_{-1}(a)$, of the component that has the frequency of the weaker signal relative to the amplitude, $A_0(a)$, of the component that has the frequency of the stronger signal.

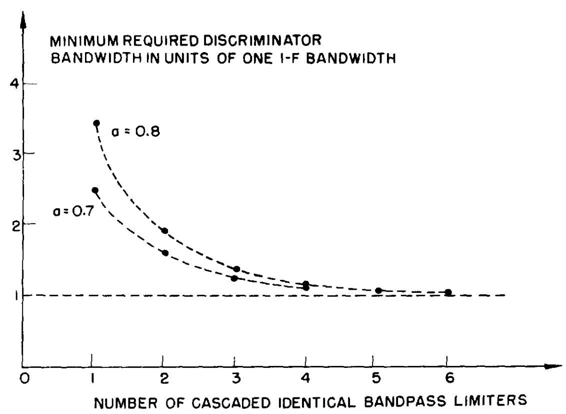

Fig. 21. Variation of the required minimum discriminator bandwidth with the number of cascaded ideal narrow-band limiters that precede the discriminator. The bandwidth of each limiter equals the i-f bandwidth.

becomes very large. Moreover, the plots of Fig. 18 make it abundantly clear that a limiter-filter bandwidth of $[(1+a)/(1-a)](BW)_{if}$ is by no means sufficient to pass all the spectral components that are necessary for a close reproduction of the instantaneous-frequency disturbance pattern shown in Fig. 2. The extent of the frequency spikes is far from being an approximate estimate of the extent of the significant spectrum. Consequently, with a limiter-filter bandwidth designed on the basis of the extent of the spikes alone, the conditions for a quasi-stationary analysis are not satisfied and it cannot be said that the discriminator bandwidth would have to be given by Eq. 44. This important conclusion will also be reached from a direct study of the conditions for the validity of a quasi-stationary approach which will appear in a future report. Indeed, about two thirds of $[(1+a)/(1-a)](BW)_{if}$ would suffice for $a > 0.7$ if only one limiter stage precedes the discriminator, and significantly less than 2/3 of this value if more limiters are used.

Associated with each of the broken curves of Fig. 18 is a smooth curve that may be passed through the values of $(BW)_{disc}/(BW)_{if}$ that are required at the odd integral values of $(BW)_{lim}/(BW)_{if}$. We shall call these smooth curves the "envelope" curves of the broken-line plots in Fig. 18, since the plots tend to be bounded by these smooth curves as the values of the limiter bandwidth grow large. Figure 22 shows the envelope curves superimposed upon the curves of Fig. 18.

We can readily show that the envelope curves are rising exponentials. The exponential character of the curves is demonstrated very clearly by the semilogarithmic plots of Fig. 23. In Fig. 23 the values of the deviation of each of the envelope curves from the corresponding asymptotic value for the curve are plotted on a logarithmic scale against the values of limiter bandwidth for which they occur, measured on a linear scale. The accuracy with which the plotted points fall on straight lines is a curious check on our calculations. This interesting coincidence enables us to develop a simple analytical expression relating close estimates (which are accurate only for odd integral values of $(BW)_{lim}/(BW)_{if}$) of the minimum discriminator bandwidth that is required to follow prescribed values of the limiter bandwidth.

The straight lines of Fig. 23 have equations of the form

$$\ln\left[\frac{1+a}{1-a} - y(a,x)\right] = -k(a)x + \ln B(a) \tag{45}$$

where

$$x \equiv \frac{(BW)_{lim}}{(BW)_{if}}$$

$y(a,x) \equiv$ envelope value of $(BW)_{disc}/(BW)_{if}$

$\quad = $ value of $(BW)_{disc}/(BW)_{if}$ at odd integral values of x

$-k(a) =$ slope of the straight line

$\ln B(a) =$ vertical intercept of the straight line

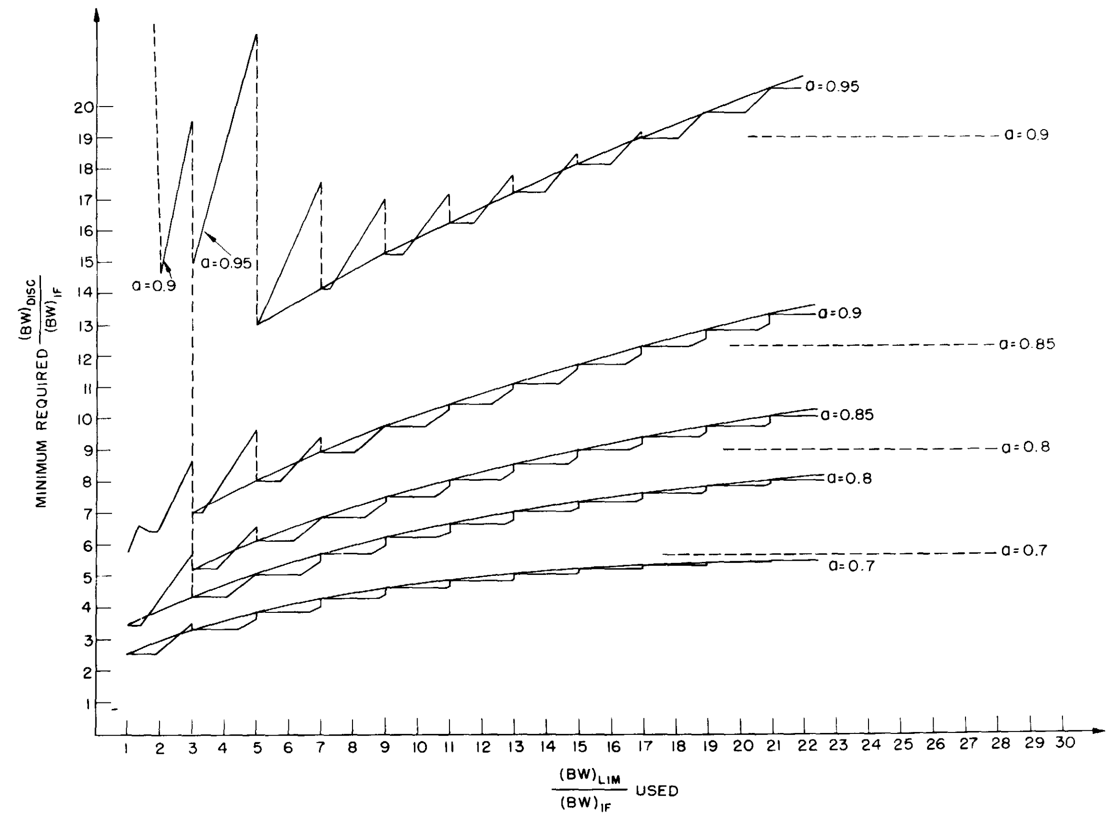

Fig. 22. Computed broken curves of Fig. 18 with their smooth "envelope" curves superimposed.

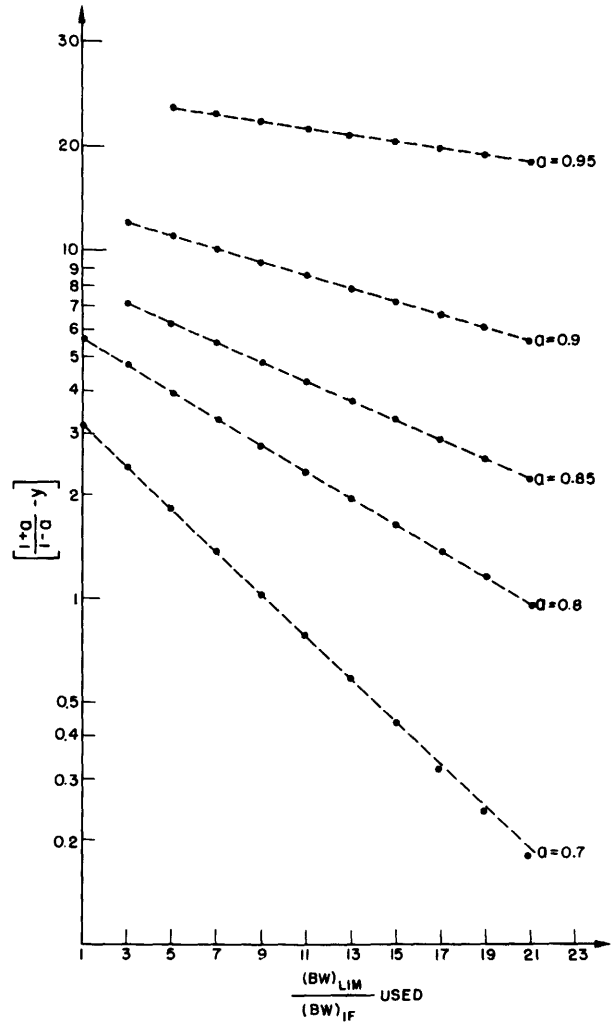

Fig. 23. Values of the deviation of each of the envelope curves of Fig. 22 from the corresponding asymptote of the curve, plotted on a logarithmic scale against the corresponding values of limiter bandwidth.

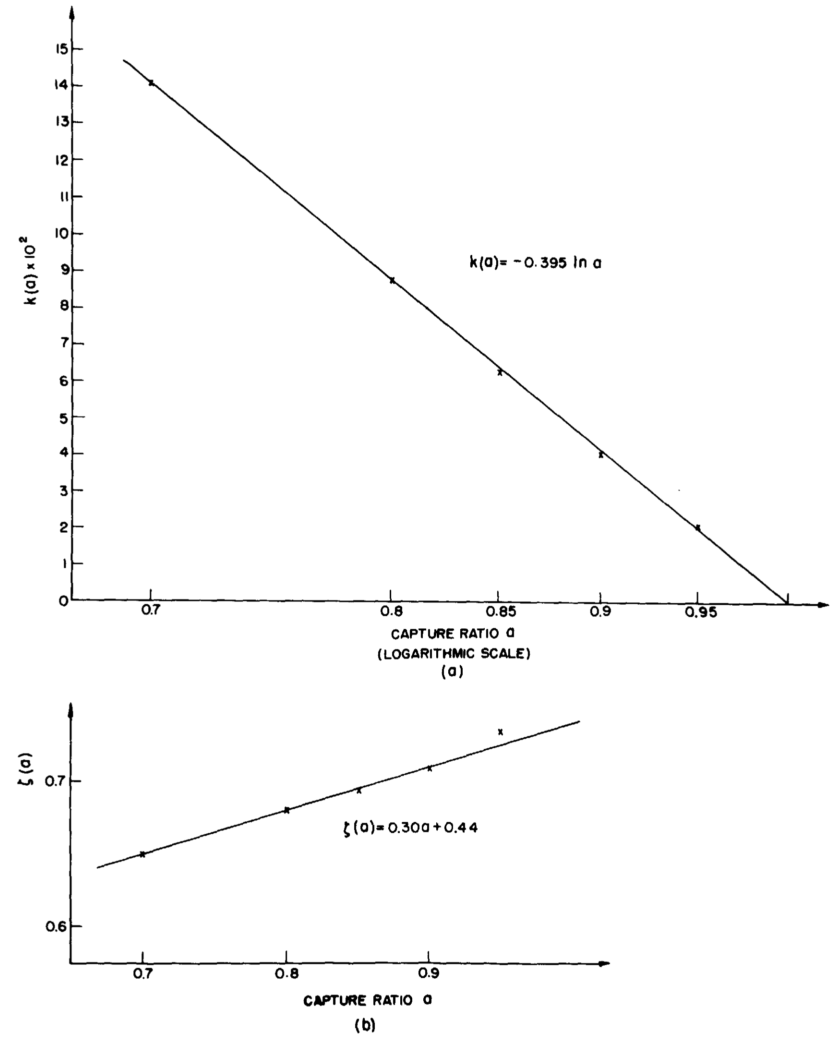

Fig. 24. Computed values of k(a) and ζ(a) of Table X.

Equation 45 can take the more convenient form

$$y(a, x) = \frac{1 + a}{1 - a} - B(a) \exp(-k(a)x)$$

$$= \frac{1 + a}{1 - a}\left[1 - \zeta(a) \exp(-k(a)x)\right] \qquad (46)$$

Equation 46 is the desired analytical expression for the envelope curves. Calculated values of the functions k(a) and

$$\zeta(a) \equiv \frac{1 - a}{1 + a} B(a)$$

are given in Table X and are plotted in Fig. 24a and b. For the values of a that are of interest, k(a) appears to satisfy the approximate expression

$$k(a) = -0.395 \ln a \qquad (47)$$

very closely (Table X and Fig. 24a). The values of $\zeta(a)$ (Table X and Fig. 24b) are based on k(a), as given by Eq. 47, with the reasonable assumption that the small deviations in the computed values of k(a) from the values given by Eq. 47 may be

Table X

a	k(a)	0.395 ln a	$\zeta(a)$	0.30a + 0.440
0.7	0.1401	0.1409	0.6500	0.650
0.8	0.0879	0.0881	0.6791	0.680
0.85	0.0630	0.0642	0.6925	0.695
0.9	0.0409	0.0416	0.7093	0.710
0.95	0.0210	0.0203	0.7345	0.725

attributed to small cumulative errors in the computations. Figure 24b and Table X show that the plotted values of $\zeta(a)$ fall rather closely on a straight line given by

$$\zeta(a) = 0.30a + 0.44 \qquad (48)$$

Equation 46 can be normalized into the form

$$\psi(a) \equiv y(a, x) \Big/ \left[\frac{1 + a}{1 - a}\right]$$

$$= 1 - \zeta(a) \exp(-k(a)x)$$

which is plotted in Fig. 25. The second term on the right-hand side of this equation gives the fractional amount by which the required minimum discriminator bandwidth has been reduced by passing the resultant two-path signal through an ideal limiter whose bandwidth is an odd integral multiple, x, of the bandwidth of the intermediate-frequency section. For each value of a covered by the plots, we observe that

$$\psi(a) \cong 2/3 \quad \text{at } x = \frac{1+a}{1-a}$$

This indicates that a limiter filter whose bandwidth is given by $[(1+a)/(1-a)](BW)_{if}$ will compress the extent of the largest frequency spikes by an amount that is sufficient to reduce the required minimum discriminator bandwidth to 2/3 of the value that would have been necessary had the limiter filter been able to reproduce the instantaneous-frequency spikes of the amplitude-limited resultant of the two signals.

In addition to the light it throws upon our results, Eq. 46 can be used to supplement the plots of Fig. 18 in the extrapolation of the correct values of y(a, x) at odd integral values of x which lie beyond the range covered by the plots, and approximations to y at other values of x. It could also be safely used for the same purpose if it is desired that a assume values between 0.7 and 0.95 that are not covered by the plots of Fig. 18. Moreover, the degree of approximation with which Eqs. 47 and 48 seem to describe k(a) and $\zeta(a)$ in the range $0.7 \leqslant a \leqslant 0.95$ seems to indicate that these expressions would also be useful for lower values of a.

Now it is recalled that if $[\Delta\omega]$ is the magnitude of the maximum instantaneous-frequency deviation of the resultant signal (at the output of the ideal-limiter filter) from the frequency of the stronger signal, then at the odd integral values of x, δ_m in Eq. 43 is unity and

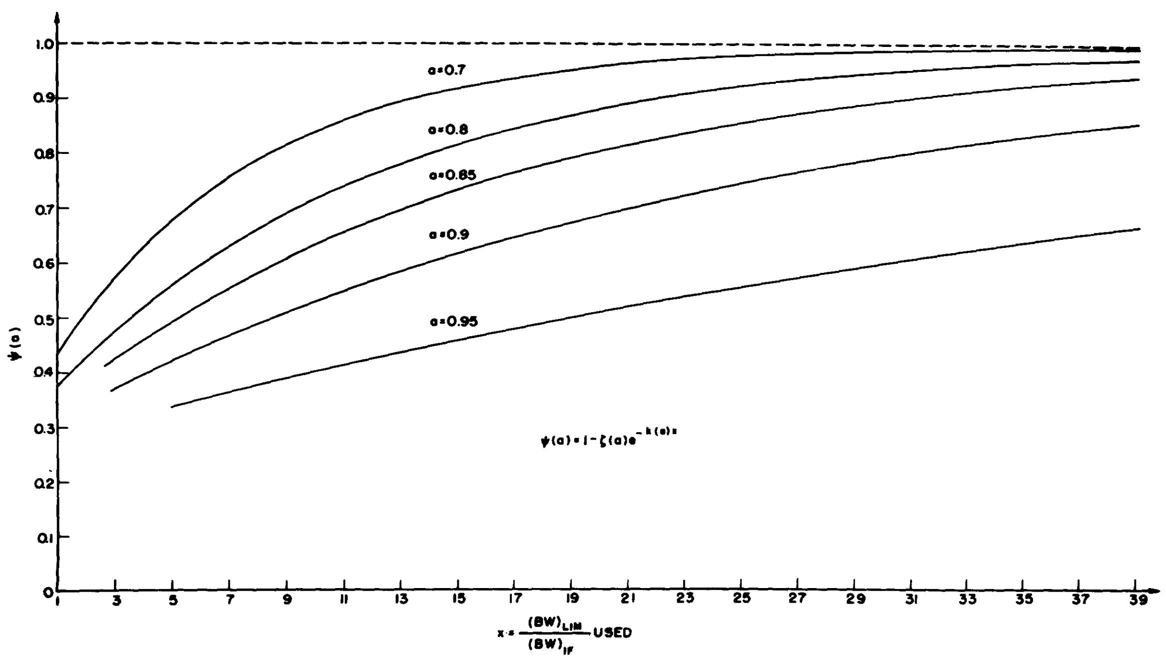

Fig. 25. Normalized plots of the "envelopes" of the broken curves of Fig. 18. The quantity $\psi(a)$ equals the normalized minimum discriminator bandwidths at the odd integral values of x.

$$y = \frac{(BW)_{disc}}{(BW)_{if}} = 2\,\frac{[\Delta\omega]}{(BW)_{if}} + 1 \tag{49}$$

Therefore, an expression for the magnitude of the frequency spikes that dictate the required minimum discriminator bandwidth at the odd integral values of x can be obtained by combining Eqs. 46 and 49. The result is

$$\frac{[\Delta\omega]}{(BW)_{if}} = \frac{1}{2}(y-1)$$

$$= \frac{a}{1-a} - \frac{1}{2}\left(\frac{1+a}{1-a}\right) \zeta(a)\,\exp(-k(a)x) \tag{50}$$

Further normalization yields

$$[\Delta\Omega] \equiv \frac{\Delta\omega}{(BW)_{if}} \cdot \frac{1-a}{a}$$

$$= 1 - \frac{1+a}{2a}\,\zeta(a)\,\exp(-k(a)x) \tag{51}$$

The second term on the right-hand side is the fractional amount by which the magnitude of the maximum deviation in the instantaneous frequency has been reduced ("damped") by passing the resultant two-path signal through an ideal limiter whose bandwidth is an odd integer multiple, x, of the bandwidth of the intermediate frequency.

The second important observation brought to light by the plots of Fig. 18 is that cascading alternate stages of limiting and filtering should reduce the requirement in discriminator bandwidth to smaller and smaller values. The calculations have considered the action of one stage of ideal limiting and filtering upon the resultant of two signals, but the results are indicative of the action of the same device when more than two sinusoids are present at its input. Clearly, if it were possible to do without a stage of amplitude limiting (e.g., when an amplitude-insensitive discriminator is used), the requirements in discriminator bandwidth would be dictated by the ratio (1+a)/(1-a), and the requirements would be the asymptotic values in the plots of Fig. 18. The reduction achieved by the action of the first stage of limiting and filtering upon the resultant of two sinusoids delivered to it by the intermediate-frequency amplifier might conceivably be duplicated (possibly, to a varying degree) by the action of the second stage of limiting and filtering upon the resultant of more than two sinusoids delivered to it, in turn, by the first stage, and so on. With enough cascaded stages, then, it should be possible to reduce the necessary discriminator bandwidth to that of the intermediate-frequency section. This important question will be pursued further in the next section.

Viewed in another way, the plots of Fig. 18 show that the effective magnitude of the instantaneous-frequency spike has been reduced (or the spike train has been effectively damped) by passing the resultant signal through the limiter-filter stage. This damping action could conceivably be duplicated (possibly, to a different extent) by further stages of bandpass limiting on the spike train associated with the resultant signal that appears at the output of the first stage. This gives justification for the cascading scheme.

We can now conclude that the function of a bandpass limiter in frequency-modulation receiver design may be viewed in a new light. For, in addition to eliminating interference and noise, which comes in as amplitude perturbations in the resultant signal, a bandpass limiter is effective in relaxing the bandwidth requirements on the frequency discriminator, in order to achieve rejection of frequency-modulation interference produced by signals that may approach the desired carrier in strength. Furthermore, a sufficiently long chain of such limiters, each in turn reducing the effective interference in the resultant signal delivered to it, would, theoretically, enable us to eliminate the interference completely, as the detailed discussion of this scheme in Section II will show.

II. THE EFFECT OF CASCADING NARROW-BAND LIMITERS

INTRODUCTION

We have established that a process of limiting followed by ideal filtering with a one i-f bandwidth will, if repeated a sufficient number of times, reduce the relative intensity of the interference to any desired degree, for all values of the initial ratio of weaker-to-stronger signal amplitude (at the input of the first limiter) that are less than 0.84. The theoretical demonstration of the success of this scheme hinges upon the fact that the configuration which is composed of the spectral components at the frequencies p and p + r rad/sec represents the most adverse condition of interference at the output of each limiter filter under the specified conditions. Therefore, the results of the spectral analysis at the output of the first limiter are directly applicable to the second and other limiters. However, for higher ratios of weaker-to-stronger signal amplitude at the input of the first limiter, the minimum permissible limiter bandwidths will accomodate unfavorable configurations that involve more than just two spectral components. For these situations, the results obtained from a study of the effect of ideal narrow-band limiting on the resultant of two sinusoids cannot be used as direct evidence of similar interference reduction when several sinusoids are present at the input of the limiter. Nevertheless, the plots of Fig. 18 demonstrate qualitative evidence that a second (third, and so on) narrow-band limiter would generally yield an additional reduction of the interference for arbitrary values of the ratio of weaker-to-stronger signal amplitude a delivered to the first limiter by the intermediate-frequency amplifier, provided the proper value of first-limiter bandwidth is used in the range of $a > 0.84$. Before any quantitative evidence can be produced in the range $a > 0.84$, an analysis of the spectrum that results from amplitude-limiting the resultant of more than two sinusoids must be carried out. This will be our starting point in the present discussion.

To bring the present task to sharper focus, some obstacles relating to the specification of the number and relative amplitude, frequency, and phase relationships of the sinusoids which will be superimposed at the input of the limiter must be overcome. The most general specifications would leave the amplitudes arbitrary, and the relative frequency and phase relationships random. But such an approach unduly complicates the problem by formulating it with unnecessary generality. Such specifications would be appropriate in the more general problem of multipath interference in which signals from more than two paths are accommodated simultaneously within the linear passband of the receiver. The problem of whether or not the effect of one stage of ideal narrow-band limiting on two-path interference will also be demonstrated under conditions of several-path interference (in which capture of one of these paths is possible), although it is worth while, is not really what we are after.

Basically, we are seeking to determine quantitatively what a scheme of cascaded narrow-band limiters will do to the interference arising from the simultaneous presence of only two carriers within the passband of the receiver. The most direct approach would, therefore, shift the interest from the investigation of the effect of the first narrow-band limiter upon the resultant of the two carriers to a study of the modifications that a second stage of narrow-band limiting produces upon the resultant signal delivered by the first stage. The sinusoidal components that make up the signal at the input of the second limiter will, therefore, be considered as a selection made by the first-limiter filter from the spectrum analyzed in section 1.2. This decision fixes the relative amplitudes, frequencies, and initial phases associated with the assumed sinusoidal components, and disposes of a major hurdle in the formulation of the problem.

To start with, we may reduce the number of possibilities drastically by considering only the configuration of components which presents the most serious capture problem associated with an assumed value of first-limiter filter bandwidth. This is readily done by drawing upon the results of section 1.6, wherein the most serious configurations associated with various values of limiter-filter bandwidth were determined. Thus the pursuit of our main objective involves another restriction upon the generality of the approach; we must, for definiteness, specify the value of the bandwidth for the first limiter, and then concentrate on the action of the second limiter upon the most troublesome configuration that can arise with this specified value of the bandwidth. It is recalled that the most serious configuration is the one that demands the largest value of permissible minimum discriminator bandwidth if it is fed directly into an amplitude-insensitive discriminator.

The only unfortunate part about restricting the choice of configuration is that the same computational effort must be expended as many times as different values are chosen for the first-limiter bandwidth. Furthermore, as the necessary number of sinusoids for any specific configuration is increased, the computational task increases quickly, and after the results of the simpler configurations have been determined, no additional information of fundamental significance about the importance of the cascading scheme will be gained from the consideration of the more complicated variations. Therefore, we find it feasible to confine our interest to the simplest situations that yield significant indications about the effectiveness of the cascading scheme in the suppression of interference.

After a Fourier analysis of the output of the second limiter in response to an input made up of a few interesting configurations, the questions investigated will relate to the requirements in the second, third, and so on, limiter bandwidths, and the bearing of the results upon the discriminator bandwidth requirements, the permissible maximum time constants in the grid circuit of a grid-bias limiter and in the output circuit of the discriminator, and the effect of narrow-band limiting upon the harmonic-component amplitudes in the structure of detected spike patterns that recur at an audible rate.

2.1 CAPTURE CONDITIONS AT THE OUTPUT OF THE SECOND LIMITER

Our first objective is best summarized in terms of the scheme shown in Fig. 26. The idealized i-f amplifier response accommodates two signal carriers whose properties have been described in section 1.2. The assumptions and notations previously

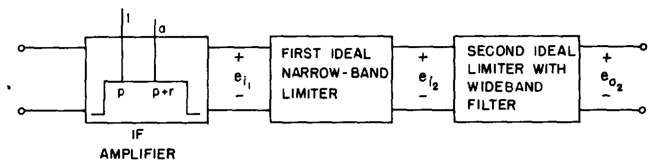

Fig. 26. Block diagram of the scheme whose output is Fourier analyzed under two-signal interference conditions at the input.

made are carried over unaltered. The i-f amplifier delivers the resultant of the two signals to the first limiter stage which has a specified bandwidth $W_{L_1} = [(BW)_{lim}/(BW)_{if}]_1$. This limiter delivers a resultant signal e_{i2}, the variation of whose composition and properties with W_{L_1} is now well understood. Our immediate task is to determine the spectral properties of e_{o2} when the worst composition of e_{i2} that corresponds to any assumed value of W_{L_1} is impressed at the input of the second limiter.

In Table XI various configurations which have been found to represent the most adverse capture conditions with the associated values of W_{L_1} are presented. Appropriately, only values of a in the range exceeding 0.84 are considered, since for values of $a < 0.84$ a workable cascading scheme has already been discussed quantitatively, in which each of the limiters had only one i-f bandwidth.

The simplest pertinent configuration arises with a value of first-limiter bandwidth whose use leads to the smallest required minimum discriminator bandwidth for $a = 0.9$, and the second smallest for $a = 0.85$. This value of first-limiter bandwidth is $W_{L_1} = 3$ and the associated worst configuration is $M = 1$, $N = 2$. It is unfortunate that this configuration does not also happen to represent the composition most unfavorable for capture associated with $W_{L_1} = 3$ for $a = 0.95$. The next simplest configuration that will apply for all three listed values of a is $M = 3$, $N = 4$, which can arise with $W_L = 7$. Only these two configurations will be treated here to establish quantitative evidence for the effectiveness of at least two cascading schemes (differing only in the bandwidth value of the first limiter) in the minimization and suppression of the interference. Other configurations arising with other assumed values for the first-limiter bandwidth may, if desired, be handled in a similar manner but with increased labor and no gain

Table XI

a	$W_L = \frac{(BW)_{lim}}{(BW)_{if}}$	Worst Configuration M	N	$\delta_m = \frac{r_{max}}{(BW)_{if}}$	Required Minimum $\frac{(BW)_{disc}}{(BW)_{if}}$
0.85	3	1	2	1	5.2790
	5	2	3	1	6.1430
	7	3	4	1	6.8784
0.9	3	1	2	1	6.9766
	5	2	3	1	8.0254
	7	3	4	1	8.9308
0.95	3	1	3	2/3	15.0
	5	2	4	3/4	13.292
	7	3	4	1	14.147

in additional fundamental information beyond that which the chosen configurations will lead to.

The linear superposition of the components corresponding to either of the two chosen configurations can be best illustrated by a phasor diagram similar to that of Fig. 6. For the configuration M = 1, N = 2, the resultant signal at the input of the second limiter is

$$e_{i2}(t) = \mathrm{Re}\left[e^{jpt} \sum_{n=-2}^{1} A_n e^{-jnrt} \right]$$

$$= \mathrm{Re}\left[A_o e^{jpt} \{1 + be^{j\phi} - ce^{-j\phi} - de^{j2\phi}\} \right] \tag{52}$$

where

$$\phi = rt, \quad b = |A_{-1}/A_o|, \quad c = |A_1/A_o|, \quad \text{and } d = |A_{-2}/A_o|$$

Thus, we can write

$$e_{i2}(t) = A(t) \cos(pt + \theta)$$

in which

$$A(t) = [R^2(\phi) + I^2(\phi)]^{1/2}, \qquad \theta = \tan^{-1} \frac{I(\phi)}{R(\phi)}$$

$$R(\phi) = 1 + (b-c) \cos \phi - d \cos 2\phi$$

$$I(\phi) = (b+c) \sin \phi - d \sin 2\phi$$

We associate (see sec. 1.2) sufficient selectivity with the second limiter to reject all harmonics of p and their associated sidebands, so that the signal at the output of this limiter (in the absence of narrow-band filtering) can be described by

$$e_{o2}(t) = \cos(pt + \theta) \tag{53}$$

where, again, the constant amplitude has been assumed to be unity for convenience. If we write

$$\cos\theta = \alpha_o + \sum_{n=1}^{\infty} \alpha_n \cos n\phi$$

$$\sin\theta = \sum_{n=1}^{\infty} \beta_n \sin n\phi$$

we can express $e_{o2}(t)$ in the form

$$e_{o2}(t) = \sum_{n=-\infty}^{\infty} B_n \cos(p - nr)t \tag{54}$$

where

$$B_o = \alpha_o$$

$$B_n = \frac{1}{2}\left(\alpha_n - \beta_n\right)$$

$$B_{-n} = \frac{1}{2}\left(\alpha_n + \beta_n\right)$$

We observe that, since

$$e_{o2}(0) = \sum_{n=-\infty}^{\infty} B_n = \sum_{n=0}^{\infty} \alpha_n = 1$$

we have a convenient check on the computation of the values of α_n.

It is readily appreciated from the experience with the much simpler two-signal problem that the task of deriving useful expressions for B_n and B_{-n} is rather hopeless. Moreover, such an attempt is not even justified, since we are chiefly interested in the values for a = 0.85 and a = 0.9, and, therefore, much less effort is involved in evaluating the B_n's by direct numerical analysis with the help of Fourier coefficient schedules. This evaluation has been carried out, and the results are presented in Table XII.

An examination of the spectral amplitudes, $B_{\pm n}$, reveals that they possess properties much like those exhibited by the $A_{\pm n}$'s in Section I. In particular, their signs alternate, beginning with B_o and B_{-1} positive, and B_1 negative. There is also a

Table XII

n	B_{-n} ($a = 0.85$)	B_n ($a = 0.85$)	B_{-n} ($a = 0.9$)	B_n ($a = 0.9$)
0		0.83628988		0.80822080
1	0.40959407	0.322199283	0.44654576	-0.32219420
2	-0.03194519	0.13354183	-0.05387307	0.15110632
3	0.02888303	-0.07886594	0.03744357	-0.09664754
4	-0.00704915	0.04463995	-0.01303143	0.06077806
5	0.005063703	-0.02714325	0.00854817	-0.04065850
6	-0.00210167	0.01659053	-0.00439994	0.02749133
7	0.00127244	-0.01034862	0.00275334	-0.01893679
8	-0.00065904	0.00650852	-0.00164511	0.01316734
9	0.00038567	0.00412995	0.00104331	-0.00923613
10	-0.00021817	0.00263663	-0.00066189	0.00651987
11	0.00012894	-0.00169202	0.00042899	-0.00462718
12	-0.00007581	0.00109045	-0.00027995	0.00329810
13	0.00004536	-0.00070531	0.00018446	-0.00235914
14	-0.00002737	0.00045745	-0.00012194	0.00169248
15	0.00001640	-0.00029769	0.00008022	-0.00121741
16	-0.00001009	0.00019385	-0.00005149	0.00087787
17	0.00000573	-0.00012688	0.00003103	-0.00063445
18	-0.00000352	0.00008271	-0.00001588	0.00045928
19	0.00000135	-0.00005454	0.00000394	-0.00033268
20	-0.00000047	0.00003539	0.00000662	0.00024094
21	-0.00000124	-0.00002364	-0.00001746	-0.00017448
22	0.00000205	0.00001506	0.00003002	0.00012634
23	-0.00000432	-0.00001028	-0.00004554	-0.00009130
24	0.00000309	0.00000309	0.00003266	0.00003266

$$\sum_{n=-24}^{24} B_n = 0.99999998 \qquad \sum_{n=-24}^{24} B_n = 0.99999998$$

Table XIII. Relative Spectral Amplitudes.

a	n	At Output of First Limiter $\lvert A_n/A_o \rvert$	At Output of Second Limiter $W_{L1} = 3$ $\lvert B_n/B_o \rvert$	$W_{L1} = 7$ $\lvert C_n/C_o \rvert$
0.85	-1	0.615582	0.489775	0.57199920
	-2	0.149476	0.038199	0.11063375
	-3	0.068389	0.034537	0.02494547
	-4	0.038120	0.008429	0.01190016
	1	0.361473	0.385025	0.37102900
	2	0.215906	0.159684	0.22470263
	3	0.147533	0.094305	0.15455640
0.9	-1	0.700212	0.552505	0.63646087
	-2	0.185342	0.0666564	0.13354826
	-3	0.091429	0.046464	0.03751897
	-4	0.054664	0.016124	0.00339219
	1	0.366596	0.398646	0.38228246
	2	0.226881	0.186962	0.24232382
	3	0.161961	0.119581	0.17520932
0.95	-1	0.811576		0.72493695
	-2	0.236194		0.17215592
	-3	0.126338		0.06440548
	-4	0.081347		0.01900334
	1	0.362808		0.38686609
	2	0.228495		0.25398435
	3	0.168012		0.19170545

qualitative correspondence between the distribution of relative amplitudes and the instantaneous-frequency pattern of the resultant signal (shown in Fig. 27). It is a simple matter to show that statements similar to those of theorems 1, 2, 3, and 4 apply directly. Finally, from Table XIII we find that not only have the (troublesome) upper sideband components shrunk in magnitude relative to the desired component B_0, but also the (helpful) lower sideband components have increased in relative amplitude. In the light of these facts, the insight gained in Section I leads us to expect significant reductions in the effectiveness of the interference from subjecting the new spectral distribution to the action of a narrow-band filter.

The first question that presents itself, at this point, is whether or not it is permissible to use a second-limiter filter bandwidth that is smaller than the bandwidth following the first limiter. To answer this question, we feel tempted to make use of the spectral amplitudes, $B_{\pm n}$, at the output of the second limiter in the same way that we used the $A_{\pm n}$ amplitudes delivered by the first, to determine the permissible minimum limiter bandwidths. But if we remember that the $B_{\pm n}$ only represent the spectral amplitudes at the output of the second limiter when the configuration $M = 1$, $N = 2$ is delivered by the first-limiter filter, the limitations on the usefulness of the $B_{\pm n}$ become immediately apparent. We now recall that the permissible minimum bandwidth after a limiter is equivalent to one i-f bandwidth only if, theoretically, the situation in which the desired component with the frequency of the stronger signal p, along with all the sideband components on the same side as the weaker signal relative to p, can be accommodated to the complete exclusion of all the helpful sideband components on the opposite side of the frequency p, and still retain an average frequency for the resultant signal that is equal to p. For such a situation to arise, the stronger signal in our analysis must lie infinitesimally to the right of the lower cutoff frequency, and r must be sufficiently small to allow all of the significant upper sideband components to pass. But if the two carriers delivered by the i-f amplifier take the necessary positions on the frequency scale for this limiting situation (and r is sufficiently small), it may be argued that the first limiter bandwidth of $3(BW)_{if}$ (or any other value greater than one $(BW)_{if}$, for that matter, since r can be assumed to be as small as is necessary) will be sufficiently wide to accommodate all of the significant sideband components on both sides of p, with the result that it will deliver to the second limiter, essentially, the amplitude-limited version of the resultant of the two input carriers without any significant instantaneous-phase and -frequency alterations. The spectrum at the output of the second limiter will then be described by Eq. 6 and the criterion for the permissibility of only one i-f bandwidth after the second limiter will be identical with that applying after the first limiter. The obvious conclusion can therefore be stated as the following theorem.

THEOREM 5. The minimum permissible limiter bandwidth is equal to one i-f bandwidth for all values of $a \leq 0.863$ delivered by the i-f amplifier, regardless of whether this limiter is the first, an intermediate or the last in a chain of limiters.

For a limiter which is preceded by other limiters, this conclusion holds provided the bandwidth of each of these other limiters is greater than or equal to one i-f bandwidth.

This theorem states that the minimum requirements in limiter bandwidth are unaffected by the action of any preceding narrow-band limiters upon the resultant signal passing through these limiters. The reason for this is, of course, the fact that, under the conditions of the situation which dictates the minimum bandwidth requirements, a finite nonzero extent in the passband below and above the cutoff frequencies of the i-f amplifier will, by reason of its being nonzero, admit a sufficiently smaller, but nonzero, r for which the effect of this preceding limiter filter upon the amplitude-limited resultant of the two input carriers will be insignificant. Although the argument is carried out for a situation in which the preceding limiter (or limiters) has a bandwidth greater than the i-f bandwidth, the use of one i-f bandwidth in a preceding stage obviously imposes an a priori restriction upon the usefulness of the combined cascade of limiters to capture ratios (at the input of the first limiter in the chain) of less than 0.863 (since this preceding stage will either be the first in the chain or will be preceded by another stage or stages of wider bandwidth). In other words, the chain is no stronger than its first weakest link.

The determination of the variation of the minimum permissible value of second-limiter bandwidth with a, for $a > 0.863$, involves laborious computation, and the result will vary with whichever bandwidth is used after the first limiter. Only when the first limiter has a bandwidth that will always pass the entire significant spectrum centered about p will the requirements in the bandwidth of the second limiter vary with a exactly as the first-limiter bandwidth did, because only then will the spectral amplitude distribution at the output of the second limiter be given by the $A_{\pm n}$'s. Since the extent of the significant (troublesome) $A_{-n}(a)$ components is greater the closer a is to unity, the narrow-band limiting effect upon the character of the resultant signal with any given value of first-limiter bandwidth will increase in significance with increase in a. This means that the effect of a narrow-band filter after the first limiter upon the minimum requirements in the bandwidth of the second limiter should be more noticeable in the range of a values that are closer to the maximum value of a which the first-limiter filter can handle successfully. This maximum value of a will mark the limit for the usefulness of the combination of the two limiters in cascade, since a loss of the desired average frequency will be introduced by the first limiter for larger values of a. Again, the failure of the first link in the chain marks the failure of the chain.

To illustrate these ideas, let us return to the original question of whether or not a bandwidth can be used after the second limiter which is narrower than the $3(BW)_{if}$ of the first limiter without jeopardizing the usefulness of the combination for all a up to 0.9807. It is recalled that the $a = 0.9807$ limit is set by the configuration $M = 1$, $N = 3$ upon the permissibility of $3(BW)_{if}$ after the first limiter. The narrow-band limiter action in the first stage will be in evidence at the output of the second stage as a general decrease in the amplitudes of the upper sideband components relative to the desired component (at p rad/sec) and an increase in the relative amplitudes of the (helpful) lower sideband components. Using a procedure similar to that illustrated in cases B and C of section 1.5, for $a > 0.863$, we proceed to determine up to what value of $a < 0.9807$ a certain number N_{max}, and M_{min} of upper and lower sideband components, specified as in section 1.6, can be accommodated within the second-limiter filter pass-band without upsetting the desired average frequency of the resultant. In the present instance, this determination involves determining the value of the bandwidth for the second limiter for which $M = M_{min}$, $N = N_{max}$ is a permissible limiting situation, and then determining the configuration of sideband components which the first-limiter filter must pass in order for $M = M_{min}$, $N = N_{max}$ to appear at the output of the second-limiter filter. The amplitude-limited resultant of the appropriate configuration delivered by the first limiter is then Fourier analyzed to determine the amplitudes of the $M = M_{min}$, $N = N_{max}$ sideband components that make up the limiting configuration of the second-limiter filter. From the results of the Fourier analysis we determine the maximum value of $\underline{a}$ for which the sum of the magnitudes of the M_{min} components and the desired component exceeds the sum of the magnitudes of the N_{max} upper sideband components. Up to this value of $\underline{a}$, the second-limiter bandwidth that will accommodate the specified $M = M_{min}$, $N = N_{max}$ as a limiting configuration, is the minimum permissible value of bandwidth. Since the narrow-band limiter action in the first stage will decrease the magnitudes of the upper sideband components and increase those of the lower sideband components relative to the magnitude of the desired component, the maximum value of $\underline{a}$ that results from the computation in terms of $M = M_{min}$, $N = N_{max}$ at the output of the second limiter will be higher than the corresponding value at the output of the first limiter. Even though the computational task is somewhat simplified by $M_{min} = 0$ for most of the practically important values of $a > 0.863$, the importance of the numerical results does not outweigh the labor involved.

One consequence of the effect of narrow-band limiting in the first stage upon the amplitudes of the upper sideband components at the output of the second limiter is that when the bandwidth of the second limiter is made equal to the i-f bandwidth, the configuration in which only the desired component at p and the component at $p + r$ ($r = (BW)_{if}$) are passed becomes the one that dictates the discriminator bandwidth requirement (i.e., the one that has the largest frequency-spike magnitude) not only for all $\underline{a}$'s up to 0.84, but also for $\underline{a}$'s that can be made to close the gap between 0.84 and 0.863. It may be argued that the first-limiter bandwidth need not exceed one i-f bandwidth in order for this to be achieved, but more than just two stages of narrow-band limiting may be needed to offset the importance of the upper sideband components, relative to the desired component, as $\underline{a}$ approaches 0.863.

With the second-limiter bandwidth taken equal to one i-f bandwidth, however, it is clear that when the first-limiter filter delivers its worst possible configuration, with $r = (BW)_{if}$, the corresponding configuration accommodated by the ideal filter following the second limiter will also represent the worst possible condition of interference for the over-all two-limiter scheme. Fortunately, the latter configuration happens to be the one in which only $B_0(a)$ at p rad/sec (corresponding to the stronger of the two carriers delivered by the i-f) and $B_1(a)$ at p + r rad/sec (corresponding to the weaker signal) are the only spectral components that are passed. Under the worst condition of interference, therefore, a scheme made up of two limiters in cascade, in which the first limiter has three times the i-f bandwidth and the second has only one i-f bandwidth, will deliver at its output only two sinusoids corresponding to the two input sinusoids with the ratio of weaker-to-stronger signal amplitude reduced from its input value of a to the value $B_{-1}(a)/B_0(a)$. Reference to Table XIII will show that for $a = 0.85$, this represents a reduction to approximately 0.49. Therefore, the indicated scheme will demonstrate the same effect upon the interference in the range of a between about 0.84 and 0.863 as one stage of ideal narrow-band limiting with only one i-f bandwidth did in the range $a < 0.84$; that is, under the worst condition of two-signal interference at the input (which arises with $r = (BW)_{if}$), the worst condition of interference at the output will also involve exactly two signals which will correspond to the input signals and will be separated in frequency by one i-f bandwidth, but the ratio of weaker-to-stronger signal amplitude at the output will be considerably smaller than that at the input. Starting with this output ratio of weaker-to-stronger signal amplitude (which will now be well within the range $a < 0.84$ in which one ideal narrow-band limiter with only one i-f bandwidth will be most effective in reducing the interference), we may retrace the argument concerning the possibility of reducing the interference under its worst condition to any desired low value by cascading the necessary number of ideal narrow-band limiters each of which has a bandwidth equal to that of the intermediate-frequency amplifier. We now summarize this result.

THEOREM 6. If a system of two or more cascaded ideal narrow-band limiters, in which the first limiter has a bandwidth a few times greater than (perhaps three times) or equal to one i-f bandwidth, and the others have just one i-f bandwidth each, is incorporated in an FM receiver, then the most adverse condition of two-signal interference will arise at both the input and the output of the scheme when the frequency difference $r = (BW)_{if}$. Under this condition of interference, the scheme will deliver at its output only two sinusoids, corresponding to the input carriers, with the ratio of weaker-to-stronger signal amplitude reduced from its input value of $a \leqslant 0.863$ to a value that can be made as small as desired by cascading the necessary number of narrow-band limiters.

Equivalently, this important theorem states that a scheme starting with an ideal narrow-band limiter that has a bandwidth three times that of the i-f bandwidth, followed by a sufficiently long chain of ideal narrow-band limiters each of one i-f bandwidth will reduce the necessary minimum discriminator bandwidth to essentially that of the i-f bandwidth for all input values of a less than about 0.863. In Table XIV we present the results of computations which show the speed with which the required minimum discriminator bandwidth decreases with the number of limiters used in this scheme when

Table XIV

Cascading Scheme: First-Limiter Bandwidth = $3(BW)_{if}$; Second and Later Limiters Have Bandwidths of $(BW) = (BW)_{if}$.

Number of Ideal Narrow-band Limiters	Required Minimum Discriminator Bandwidth
	a = 0.85
1	5.2790
2	2.922
3	1.74
4	1.325
5	1.15

Table XV

Cascading Scheme: Bandwidth of Each Ideal Narrow-band Limiter = $3(BW)_{if}$.

Capture Ratio	Frequency-Spike Magnitude			Required Minimum $(BW)_{disc}/(BW)_{if}$ after n Identical Narrow-band Limiters		
a	Output of i-f Section	Output of First Narrow-band Limiter	Output of Second Narrow-band Limiter	n=0	n=1	n=2
0.85	5.667r	2.1395r	1.1099r	12.333	5.2790	3.220
0.9	9r	2.9883r	1.3913r	19	6.9766	3.7826
		≈3r			≈7	

$3(BW)_{if}$ is used in the first stage. These results are also plotted in Fig. 28. Plots for a = 0.8 and a = 0.7 are also reproduced from Fig. 21 for comparison.

Let us consider next the situation in which the second limiter is given a bandwidth three times that of the i-f bandwidth. It is clear that with this value of bandwidth, the configuration that will dictate the required minimum discriminator bandwidth is again M = 1, N = 2, as it was for the first limiter. The present case is best illustrated by the plots of Fig. 27. In these plots, $\Omega_o(t)$ is the instantaneous-frequency variation of the resultant two-path signal delivered to the first limiter, over one period of the frequency difference r between the two carriers. If an amplitude-insensitive discriminator is used immediately following the i-f amplifier, the FM-to-AM conversion characteristic of the discriminator must be linear over the whole range of variation of

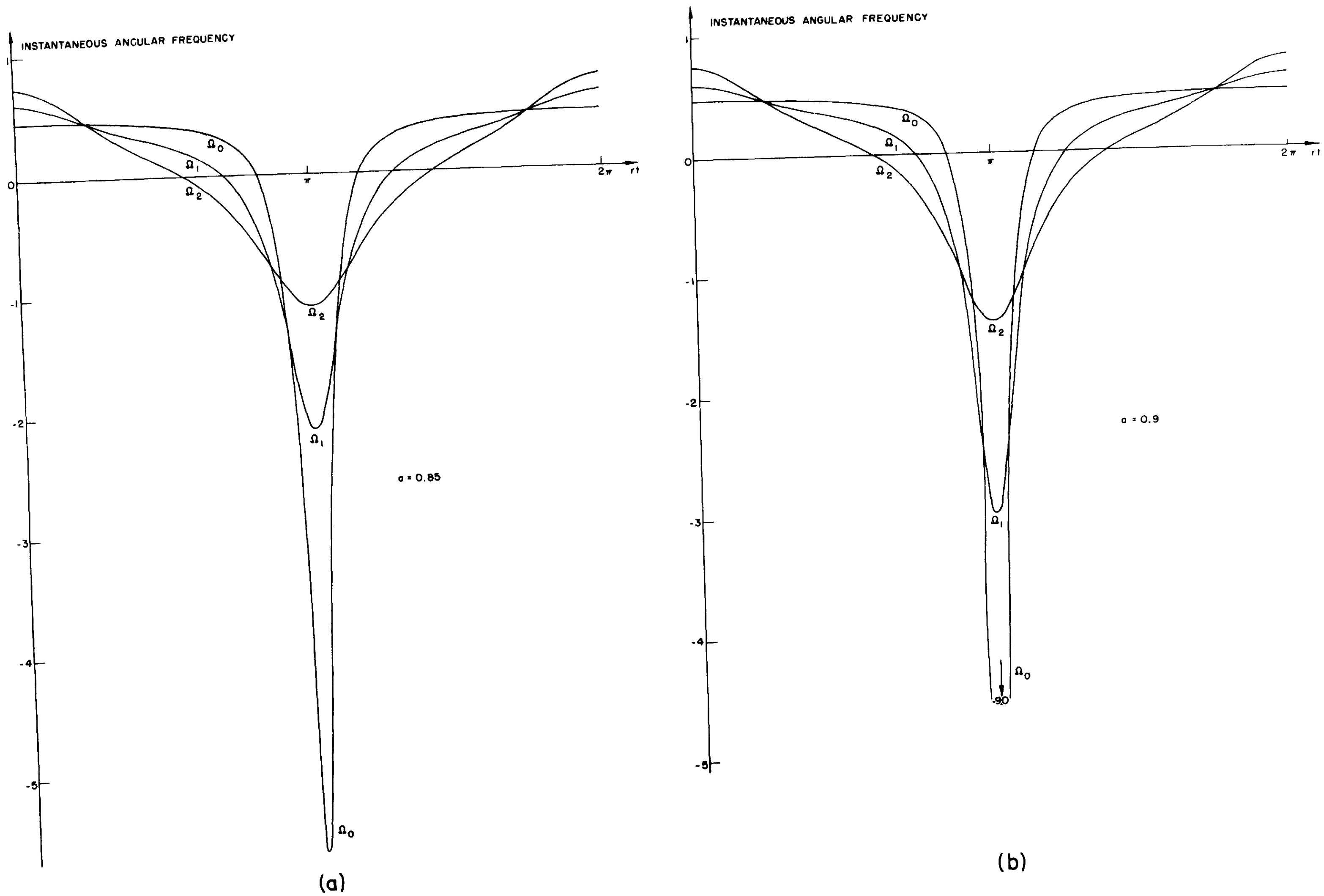

Fig. 27. Instantaneous deviations from the frequency of the stronger signal experienced by the frequency of the resultant signal. Ω_0, Ω_1, and Ω_2 illustrate the effect of no narrow-band limiting, one stage of narrow-band limiting, and two stages of narrow-band limiting.

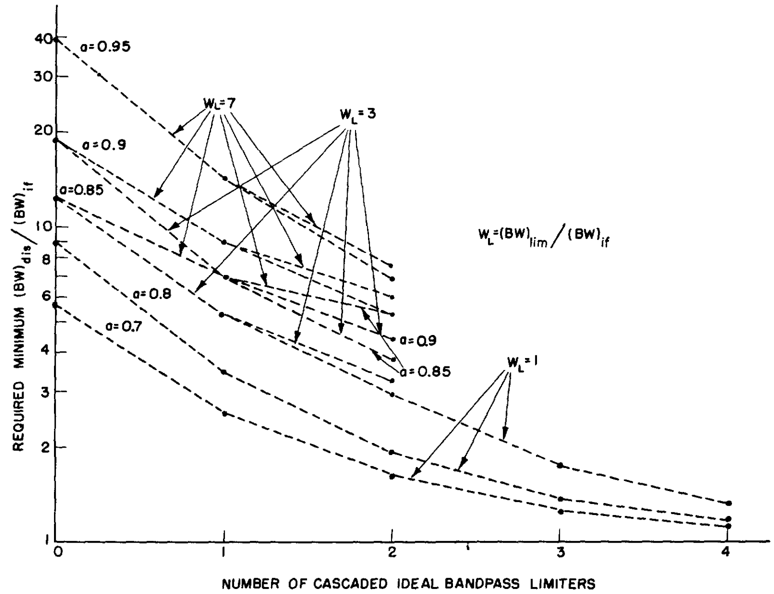

Fig. 28. Variation of the required minimum discriminator bandwidth with the number of cascaded narrow-band limiters.

$\Omega_0(t)$. However, if one stage of ideal narrow-band limiting is inserted between the i-f section and the amplitude-insensitive discriminator, the most serious variations in the instantaneous frequency of the resultant signal delivered to the discriminator will follow the curve denoted by $\Omega_1(t)$ when the ideal-limiter bandwidth is three times the i-f bandwidth. The effect of cascading two such identical stages of narrow-band limiting between the i-f section and the discriminator is that the most serious variations in the resultant signal frequency will now follow the curve marked $\Omega_2(t)$.

Table XV shows the values of required minimum discriminator bandwidth under the conditions of each of the plots of Fig. 27. These values are also plotted in Fig. 28. It is clear that the same kind of action indicated by these results will also be exhibited by further stages of ideal narrow-band limiting, in which each stage has a bandwidth of $3(BW)_{if}$, until, after a sufficient number of them has been cascaded, the required minimum discriminator bandwidth becomes essentially equal to that of the i-f bandwidth. The choice of $3(BW)_{if}$ for the bandwidth of each stage is inspired by the desire to investigate a scheme in which each limiter bandwidth meets the minimum requirement for all $\underline{a}$'s up to 0.9807. The choice of the permissible minimum value for this range is also in line with the basic aim of determining the greatest achievable reductions in the required discriminator bandwidth. For, with a scheme in which only two limiters are cascaded, the first of which has a bandwidth of $3(BW)_{if}$, it is clear that as the bandwidth of the second limiter is increased from its permissible minimum value to higher and higher values, the minimum discriminator bandwidth that is required after this second limiter increases from a small value toward a larger value which will be dictated by the worst configuration delivered by the first limiter in the absence of the second. The latter value is achieved when the second-limiter bandwidth becomes sufficiently large to accommodate all of the spectral components of significance in the structure of $e_{o2}(t)$.

Strictly speaking, the effectiveness of the cascading scheme in which each limiter has a bandwidth of $3(BW)_{if}$ has thus far been demonstrated only for values of a for which the configuration M = 1, N = 2 is the most critical one at the output of the first stage. With the help of Table VII this may be closely estimated to be the case for all a's up to about a = 0.91. Therefore, the results bear evidence that at least up to this value of a, a quantitative account of this scheme has been provided. Although for higher values of a the most adverse interference condition at the output of the first limiter does not correspond to $r = (BW)_{if}$ (which holds for the worst condition at the input), but to a smaller value of r, with M = 1, N = 3, up to 0.9807, we now have little doubt that the scheme will exhibit similar reductions in the over-all relative importance of the interference without the need of a separate computation starting with M = 1, N = 3 at the input of the second limiter, for a in the range $0.91 < a < 0.98$.

As a final illustration, let us associate with the first limiter a bandwidth seven times that of the i-f bandwidth. From Table XI it is evident that for this value of bandwidth, the worst configuration that will be delivered by the first limiter to the second is given by M = 3, N = 4 for all three listed values of a. Therefore, the present example will involve a = 0.95 in a direct computation, and will also reinforce the conclusions reached in the discussion above, and illustrate others. It is observed from Table XI that the worst condition of interference arises with $r = (BW)_{if}$ at the input, as well as at the output of the limiter.

With reference to Fig. 26, and with an analysis entirely analogous to the one carried out in deriving Eq. 54, it can be shown that when e_{i2} is the resultant of the configuration M = 3, N = 4, e_{o2} is expressible in the form

$$e_{o2}(t) = \sum_{n=-\infty}^{\infty} C_n \cos(p - nr)t \tag{55}$$

where the C_n's have been computed by numerical analysis. The values of the spectral amplitudes, $C_{\pm n}$, at the output of the second limiter are given in Table XVI for a = 0.85, 0.9, and 0.95. This table shows general properties of the spectral amplitudes that differ from those of $A_{\pm n}$ or $B_{\pm n}$ insofar as the new instantaneous-frequency pattern of the resultant signal at the input of the second limiter requires a slight reshuffling in the amplitudes of some of the components. This reshuffling can be shown not to affect the validity of statements concerning conditions at the output of the second limiter that are similar to those stated in theorems 1, 2, 3, and 4 of Section I. In Table XVI, as in Table XII, peculiarities of the method of computation have affected some of the signs, as well as the values, of the spectral amplitudes for values of $n > \sim 17$, so that in some instances they are quite unreliable. But this is not disturbing, because this range of

Table XVI

n	$a = 0.85$ C_{+n}	$a = 0.85$ C_{-n}	$a = 0.9$ C_n	$a = 0.9$ C_{-n}	$a = 0.95$ C_n	$a = 0.95$ C_{-n}
0	0.80017600		0.77188817		0.73523438	
1	-0.29688850	0.45770003	-0.29507931	0.49127662	-0.28443725	0.53299857
2	0.17980165	-0.08852647	0.18704689	-0.10308432	0.18673803	0.12657495
3	-0.12336723	0.01996077	-0.13524200	0.2896045	-0.14094844	0.04735312
4	0.05582540	0.00953726	0.06953726	0.00261839	0.08551518	-0.01397191
5	-0.03844722	0.01258463	-0.05113540	0.01696557	-0.06744533	0.02464101
6	0.02680961	-0.00589528	0.03810971	-0.00873571	0.05406953	-0.01477437
7	-0.01858683	0.00203819	-0.02833829	0.00390189	-0.04351496	0.00877105
8	0.01207127	-0.00014574	0.02013156	-0.00143672	0.03424925	-0.00537343
9	-0.00828326	0.00095336	-0.1494002	0.00197482	-0.02778427	0.00480514
10	0.00576595	-0.00063900	0.01121969	-0.00136341	0.02275956	-0.00351608
11	-0.00402425	0.00031986	-0.00845296	0.00082122	-0.01872353	0.00247015
12	0.00278912	-0.00013728	0.00635138	-0.00048925	0.01542139	-0.00173406
13	-0.00194982	0.00012896	-0.00480570	0.00039060	-0.01277079	0.00128370
14	0.00137033	-0.00009157	0.00365223	-0.00027679	0.01061500	-0.00088292
15	-0.00096552	0.00005484	-0.00278268	0.00017940	-0.00884549	0.00054239
16	0.00068065	-0.00003042	0.00212306	-0.00010854	0.00738420	-0.00025689
17	-0.00048101	0.00001987	-0.00162334	0.00006206	-0.00617525	0.00001288
18	0.00034066	-0.00001097	0.00124342	-0.00004154	0.00517067	0.00021461
19	-0.00024160	0.00000291	-0.00095349	-0.00001602	-0.00433220	-0.00043566
20	0.00017145	0.0000428	0.00073147	0.00005245	0.00362936	0.00065852
21	-0.00012169	-0.00001116	-0.00056099	-0.00009028	-0.00303767	-0.00089166
22	0.00008629	0.00001914	0.00042954	0.00013322	0.00253691	0.00114429
23	-0.00006096	-0.00002927	-0.00032760	-0.00018454	-0.00211029	-0.00142522
24	0.00004269	0.00004269	0.00024784	0.00024784	0.00174374	0.00174374

$\sum_{n=-24}^{24} C_{-n} = 0.99999999$ $\qquad \sum_{n=-24}^{24} C_{-n} = 1.00000001$ $\qquad \sum_{n=-24}^{24} C_{-n} = 1.00000001$

n values has been included mainly to improve the accuracy of the computation of the lower-order components that are of major significance in this study.

Table XIII shows that in the present instance, also, the amplitudes of the upper sideband components relative to the amplitude of the desired component, C_o, have been decreased, while those of the lower sideband components have been increased. We now accept this effect as characteristic of the action of the ideal narrow-band limiting process upon the resultant of two or more sinusoids which differ in frequency by an amount that is small compared with the frequency of either of them, but have such amplitudes and initial phases that with their various frequency differences harmonically related, the average frequency of the resultant over a period of the fundamental of the frequency differences is always equal to the frequency of the strongest of the component sinusoids. Those components which by virtue of their initial phases and frequency specifications tend to help keep the average frequency value at the frequency of the strongest component, will generally have their amplitudes increased relative to the amplitude of the strongest component, while those that tend to upset the average value will have their relative amplitudes diminished. At least, this is true when the various sinusoids exhibit the properties of the spectral components centered about the frequency p (or any of its harmonics, with the proper selectivity) in the structure of the amplitude-limited resultant of two sinusoids of different amplitudes but slightly different frequencies.

Suppose we now cascade two ideal narrow-band limiters, the first of which has seven times the i-f bandwidth. For the bandwidth of the second limiter we first choose $3(BW)_{if}$ and later we select $7(BW)_{if}$. In all cases, the worst condition of interference will arise at the input of the scheme, as well as at the output of each stage and at the output of the whole scheme, when $r = (BW)_{if}$. Therefore, when $W_{L_2} = 3$, the minimum discriminator bandwidth requirement will be dictated by the configuration M = 1, N = 2

Table XVII

Cascading Scheme: First-Limiter Bandwidth = $7(BW)_{if}$.

Capture Ratio	Required Minimum $(BW)_{disc}/(BW)_{if}$		
a	W_{L_2} (very large)	$W_{L_2} = 3$	$W_{L_2} = 7$
0.85	6.8784	4.3826	5.2681
0.9	8.9308	5.2002	5.8989
0.95	14.147	6.9461	7.5921

at the output of the second limiter. The values dictated when a = 0.85, 0.9 or 0.95 are presented in Table XVII and plotted in Fig. 28. Table XVII presents also the minimum values that are required when $W_{L_2} = 7$, in which case M = 3, N = 4 dictates the requirements. The decrease in required discriminator bandwidth that is brought about by each scheme suggests that if the first limiter with $7(BW)_{if}$ is followed by a chain of limiters in which every stage has $7(BW)_{if}$ or $3(BW)_{if}$, then a scheme capable of reducing the minimum discriminator bandwidth value to essentially that of the i-f bandwidth is at hand, provided the proper number of narrow-band limiters is used. It is also clear that the minimum requirement in $(BW)_{disc}$ will converge faster toward one $(BW)_{if}$ when the first limiter is followed by limiters with $3(BW)_{if}$ rather than $7(BW)_{if}$; that is to say, fewer stages each with $3(BW)_{if}$, after the first, would be needed to reduce the discriminator bandwidth requirement to a prescribed value than with higher values of limiter bandwidth. These conclusions can now be summarized.

THEOREM 7. Under conditions of two-signal interference, the minimum discriminator bandwidth requirement for all values of the ratio of weaker-to-stronger signal amplitude that are less than unity, delivered by the intermediate-frequency amplifier, can be reduced to a value that is as close to one i-f bandwidth as is desired by cascading the necessary number of ideal narrow-band limiters which have appropriately chosen bandwidths.

2.2 UPPER BOUNDS ON THE LIMITER AND DISCRIMINATOR TIME CONSTANTS

A basic requirement in receiver design to suppress multipath and cochannel disturbances is the use of the proper time constants in the limiter and discriminator circuits. The limiter circuit must be capable of following the sharp changes in the envelope of the resultant signal, which may recur at a rate that is equivalent (in cycles per second) to the i-f bandwidth. At the input of the first limiter, the ratio of maximum-to-minimum amplitude for a capture ratio a (ratio of peak value of interference to peak value of signal) is readily seen to be $(1+a)/(1-a)$. In the discriminator circuit, the output circuit across which the voltage level (varying with the instantaneous-frequency variations of the resultant signal impressed upon the discriminator) is taken must be capable of following the detected instantaneous-frequency pattern in order to avert the possibility of diagonal clipping and leave the variational control of the operation of the detecting diodes entirely in the hands of the amplitude of the hybrid signal (AM plus FM) delivered by the FM-to-AM converter section of the discriminator circuit. Here, also, the sharp changes in the detected voltage level may recur at a maximum rate of one i-f bandwidth in cycles/sec, and so an examination of the time-constant requirement in the output circuit is necessary.

We have demonstrated that the action of a stage of ideal bandpass-limiting upon the disturbances arising from multipath and cochannel interference results in a substantial reduction in the minimum bandwidth requirements (for interference rejection) in the FM-to-AM conversion characteristic of the discriminator and in certain reductions of limited significance in the limiter-bandwidth requirements. We have also demonstrated that the same effect will be observed with additional stages of bandpass limiters and, thus, that cascading a sufficient number of such limiters will successively reduce the minimum requirement in discriminator bandwidth to the bandwidth of the intermediate frequency. Since this effect may be looked upon as a reduction in the effective peak strength of the weaker signal relative to the stronger one, it is also reasonable to expect the successive reductions in the minimum required bandwidths to be accompanied by successive increases in the maximum permissible values of the time constants in the discriminator and limiter circuits, and successive decreases in the effective amplitudes of the audible harmonics in the structure of the detected instantaneous-frequency spike trains when the frequency difference between the two paths is audible.

In view of these disturbance-reducing characteristics, such a chain of bandpass limiters appears to be indispensable in any effective attempt to abate cochannel and multipath disturbances. It turns out, however, that the cascading of narrow-band limiters ahead of the discriminator is only one important means for achieving this result. Other equally interesting but more elaborate methods have resulted from new approaches to the solution of the problem.

A. Limiter time-constant requirements

Perhaps, the only important limiter circuit that presents a time-constant problem, at present, is the grid-bias pentode limiter circuit shown in Fig. 29. A discussion of the operation of this circuit has been presented by Arguimbau (2). It will suffice, for our purposes, to recall that the operation of this limiter depends upon the control, by the instantaneous amplitude of the input signal, of an automatic self-rectified grid bias, which in turn controls the conduction angle and the height of the plate-current pulses, increasing the height and decreasing the angle with increasing signal amplitude so that the net charge delivered by each pulse to the plate tank condenser is kept approximately constant. The condition that the dynamic self-bias must be exclusively controlled by the instantaneous amplitude of the input signal imposes necessary restrictions on the

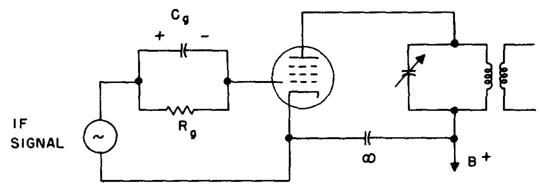

Fig. 29. Grid-circuit limiter for time-constant computation.

largest value of the RC product for the grid-leak-and-capacitor arrangement in the grid circuit – restrictions that are directly dictated by the reciprocal of the maximum ratio of the time derivative of instantaneous amplitude and the value of instantaneous amplitude.

The mechanism of the operation in the grid circuit of the grid-bias limiter lends itself to a treatment that is very much like the usual analysis of the simple diode peak detector. In our problem, it is recognized at the outset that if the time constant R_gC_g is too high, the grid-bias variations will only follow the slowest variations in the envelope value of the amplitude of the impressed signal; and if it is too low, the change in bias will not be great enough for effective smoothing of the amplitude variations. There are three time-constant considerations that must be taken into account, if this circuit is to be used in a receiver designed to handle some specified capture ratio a.

(a) The product R_gC_g must be sufficiently small to enable the circuit to follow the amplitude variations of a resultant two-path signal.

(b) C_g must be sufficiently large to by-pass R_g at radio frequencies and offer an impedance at these frequencies that is much lower than the input impedance of the tube when the grid potential is on the negative swing, in order for the intermediate-frequency voltage to appear effectively between grid and cathode of the tube. If ω_o is the i-f frequency, and C_{in} is the input capacity of the stage, the present requirements can be summarized as

$$C_gR_g \gg 1/\omega_o \quad \text{and } C_g \gg C_{in}$$

(c) R_g must be sufficiently large for the development of the necessary bias on the grid for effective smoothing of the envelope of the input wave. If the grid-to-cathode conduction resistance is r_g, then we want $R_g \gg r_g$.

For the determination of requirement a, we note that if A(t) denotes the instantaneous amplitude of the resultant signal impressed at the input to the grid circuit, then, assuming that grid-to-cathode conduction occurs only when the grid goes positive, and that the conduction resistance is negligible compared with R_g, we find that the grid current, averaged over one radiofrequency cycle, is given by

$$i_{av} = C_g \frac{dA(t)}{dt} + \frac{A(t)}{R_g} \geqslant 0 \qquad (56)$$

since, by assumption, current can flow only from the grid to the cathode and not in reverse. It is readily appreciated that the input envelope and the bias voltage will keep together, provided the grid draws current for a short interval during each radiofrequency cycle (which amounts to a process of sampling) and provided the fractional change in the envelope value during any one radiofrequency cycle is small. The latter proviso is comfortably met by the assumption $p \gg r$ (explicitly understood throughout this study), whereas the conduction (or sampling) condition is met only if the average value of grid current over one radiofrequency cycle is positive, as indicated in Eq. 56.

It is of interest to note that condition 56 can also be written in the form

$$-\frac{A(t)}{R_g C_g} \leqslant \frac{dA(t)}{dt} \equiv A'(t) \tag{57}$$

The quantity on the left-hand side is recognized as the negative of the magnitude of the rate at which the capacitor tends to discharge at the instant of time, t, when the amplitude of the input signal is given by $A(t)$ and the rate at which the amplitude is changing is $A'(t)$. Condition 57 states, therefore, that the magnitude of the rate at which the capacitor tends to discharge at any instant of time must always be greater than (or, at worst, equal to) the magnitude of the rate at which the amplitude is changing at that instant, in order for the self-rectified bias on the grid to follow the amplitude of the input signal. The problem could have been approached from this alternative, but entirely equivalent, point of view. Both views are needed for a thorough understanding of the situation. Both points of view must be satisfied simultaneously on a purely physical basis; and both imply exactly the same inequality. Thus, the condition on the RC time constant of the grid circuit can be written

$$R_g C_g \leqslant -\frac{A(t)}{A'(t)}$$

The quantity on the right-hand side will, of course, be positive, and it will vary with time. Therefore, if the most unfavorable situation is to be met, the condition should read

$$R_g C_g \leqslant \left[-\frac{A(t)}{A'(t)}\right]_{min} \tag{58}$$

Under two-signal interference conditions, we find (see Fig. 1) that

$$A(t) = (1 + 2a \cos rt + a^2)^{1/2}$$

which, upon substitution in Eq. 56, and after a straightforward simplification, yields

$$1 + a^2 + a\left(4 + R_g^2 C_g^2 r^2\right)^{1/2} \cos(rt + \psi) \geqslant 0 \tag{59}$$

where $\psi = \tan^{-1}(1/2)(R_g C_g r)$. The worst condition arises when $\cos(rt + \psi) = -1$; it will be met if

$$1 + a^2 \geqslant a\left(4 + R_g^2 C_g^2 r^2\right)^{1/2}$$

whence, we want

$$R_g C_g \leqslant \frac{1 - a^2}{ar} \tag{60}$$

The frequency difference r can have a maximum value, $r_{max} = 2\pi W_{if}$, where W_{if} denotes the intermediate-frequency bandwidth in cycles/sec. Therefore, under the worst conditions we must have

$$\tau_{\ell o} = W_{if} R_g C_g \leqslant \frac{1 - a^2}{2\pi a} \tag{61}$$

The maximum permissible value of $\tau_{\ell o}$, as given by the equality sign in Eq. 61, is shown in Fig. 30 plotted against a. The plotted values are also listed in Table XVIII. Multiplication by a scale factor appropriate to the value of W_{if} will convert the normalized values of time constant to microseconds.

For conditions at the input to the second-limiter stage, which immediately follows the filter of the first limiter, the problem is greatly simplified if the filter is assumed to have idealized amplitude and phase characteristics. Thus, if we consider the filter in the plate circuit of the first grid-bias pentode limiter to be an ideal filter, then the input signal to the next stage is described in terms of the worst configuration of side-band components that this filter will pass. The amplitude of the resultant of these components will not be constant, since, in general, the components will include only a finite number with significant amplitudes at the output of the first limiter. If the limiter performance is also assumed to be ideal, then the results of the spectral analysis of Section I are directly applicable. Thus, if the first-limiter filter is assumed to have a bandwidth equal to one i-f bandwidth (as is permissible for all $a \leqslant 0.863$), then we know (from section 1.6) that the configuration A_o, A_{-1} will be the worst possible spectrum for all a's up to approximately 0.84, and so, for all such cases, Eq. 60 is directly applicable with a replaced by A_{-1}/A_o. Since $(A_{-1}/A_o) < a$, it is recognized that the highest permissible value of the time constant $R_g C_g$ is larger at the input of the second limiter than it was at the input of the first — a decided improvement in the design conditions. The extent to which improvement has been achieved is readily seen from the curve for ${}_2\tau_{\ell 1}$ shown in Fig. 30. Conditions at the input of the third-limiter stage are similarly computed if the second stage is assumed to be identical with the first one, and so on. Figure 30 also shows a curve for ${}_3\tau_{\ell 1}$ which applies to the time constant at the input of the third limiter.

When the ideal filter following the first limiter has three times the i-f bandwidth, values of a up to 0.9807 may be considered. The critical configuration for this bandwidth is M = 1, N = 2 and the spectral analysis of section 2.1 (which is restricted to the values a = 0.85 and a = 0.9) is useful for a further study of conditions at the input of a

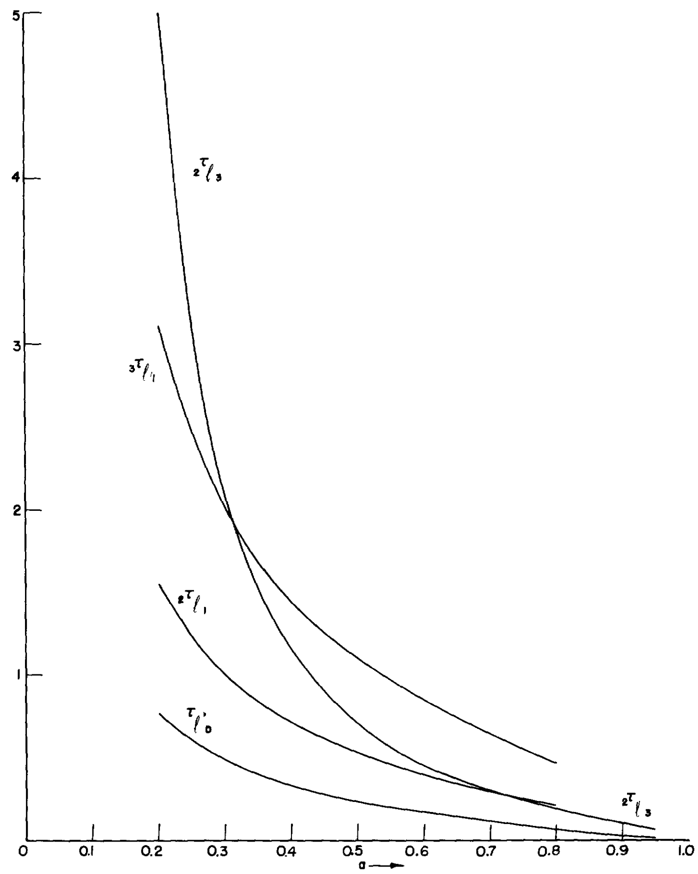

Fig. 30. Variation of the maximum permissible values of the grid-circuit time constant with the interference ratio a.

Table XVIII

Maximum Permissible Values of Normalized Time Constants:

$\tau_\ell = W_{if} R_g C_g$ in the Grid Circuit.

	First limiter	Second limiter $W_{L1} = 1$	Second limiter $W_{L1} = 3$	Third limiter $W_{L1} = W_{L2} = 1$
a	$\tau_{\ell o}$	${}_2\tau_{\ell 1}$	${}_2\tau_{\ell 3}$	${}_3\tau_{\ell 1}$
0.2	0.7639	1.5514	4.9694	3.115
0.3	0.4828	1.0001	2.1361	2.018
0.4	0.3342	0.7131	1.1548	1.451
0.5	0.2387	0.5308	0.7014	1.0936
0.6	0.1698	0.3994	0.4520	0.8386
0.7	0.1160	0.2950	0.2962	0.6380
0.8	0.0716	0.2042	0.1877	0.4652
0.85			0.1425	
0.9	0.0336		0.1004	
0.95			0.0576	
0.98	0.0064		0.0280	

Note. $W_L = (BW)_{lim}/(BW)_{if}$

third-limiter stage that may follow. Higher values of first-limiter bandwidth can be used, and the reasoning used before applies to the conditions at the input of the second limiter. However, in the general case, in which the configuration $M \neq 0$, $N > 1$ is to be handled, the step that corresponds to the transition from Eq. 59 to Eq. 60 is not obvious from inspection of the expression corresponding to Eq. 59, and it must be carried out by a process of minimization of the quantity $-A(t)/A'(t)$, as is indicated in formula 58.

With reference to Fig. 6, we recall that the square of the amplitude function of the resultant of the configuration M, N is given by

$$A^2(t) = \left[\sum_{n=-N}^{M} A_n \cos nrt\right]^2 + \left[\sum_{n=-N}^{M} A_n \sin nrt\right]^2$$

$$= \underset{M-N}{F}(rt) \tag{62}$$

with $A(t) \equiv F^{1/2}(rt)$, and $\phi = rt$, we can write

$$\frac{A(t)}{A'(t)} = (2/r)\,\frac{F(\phi)}{F'(\phi)} \tag{63}$$

and condition 58 becomes

$$T = rR_gC_g \leqslant \left[-2\,\frac{F(\phi)}{F'(\phi)}\right]_{min} \tag{64}$$

When the first-limiter bandwidth is three times the i-f bandwidth, the most unfavorable situation arises when $M = 1$, $N = 2$, and $r = 2\pi W_{if}$. In this case

$${}_1F_{-2}(\phi) = A_o^2\left[\sigma_1 + \alpha_1 \cos\phi + \beta_1 \cos 2\phi + \gamma_1 \cos 3\phi\right] \tag{65}$$

where σ_1, α_1, β_1, and γ_1 are combinations of sums and products of the amplitudes, $A_{\pm n}$. Thus, it is readily shown that when expression 65 is substituted in 64, and the derivative of the resultant right-hand member is set equal to zero, the value of ψ that gives a minimum is a root of a fifth-degree equation in $\cos\phi$. The root was determined graphically for various values of $\underline{a}$, and the results were used to determine the maximum permissible value of ${}_2\tau_{\ell 3} \equiv W_{if}R_gC_g$ (in the grid circuit of the second limiter when the first-limiter bandwidth is three times the i-f bandwidth) as a function of $\underline{a}$. The resultant maximum permissible values of ${}_2\tau_{\ell 3}$ are also plotted in Fig. 30, for comparison with the other normalized time-constant curves. The numerical values are also presented in Table XVIII.

With larger values of first-limiter bandwidth, the variations in the amplitude of the most troublesome signal delivered to the second limiter become less and less severe, and the upper bound on the permissible R_gC_g product at the input of the second stage

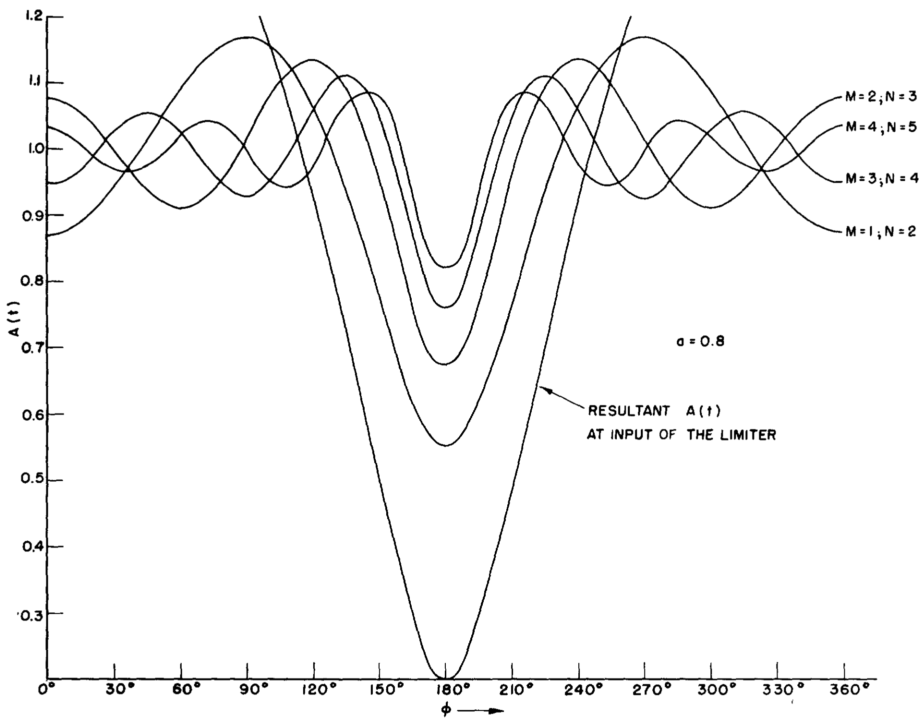

Fig. 31. Variations in the instantaneous amplitude of the resultant signal at the output of a narrow-band limiter compared with the corresponding variations at the input of the limiter

becomes higher and higher. This is also brought out by the plots of A(t) for various configurations shown in Fig. 31.

We may now conclude that, although the action of a narrow-band limiter upon the character of the resultant signal effects only a partial reduction in the severity of the amplitude variations of the signal (instead of a complete smoothing-out process), the partial abatement of the amplitude variation is, nevertheless, sufficient to show a significant increase in the upper bound on the permissible grid-circuit time constant. Eventually, with a sufficient number of cascaded narrow-band limiters, the resultant signal amplitude attains an essentially constant value, and the upper bound on the RC product in the later stages is sufficiently high to be of no importance in their design.

It may seem unnecessary to recall, before leaving this topic, that at least two other important considerations must be kept in mind in the choice of R_g and C_g values for the grid-bias pentode limiter.

B. Discriminator time-constant requirements.

The time-constant requirements of two commonly encountered discriminator low-frequency output circuits will now be discussed in the order of their simplicity.

The low-frequency output circuit of many discriminators can be reduced to the form shown in Fig. 32. The detected voltage, which is proportional to the instantaneous-

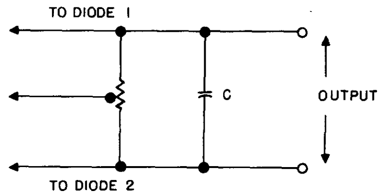

Fig. 32. Discriminator-output circuit for time-constant computation.

frequency variations of the signal at the discriminator input, appears across the equivalent RC combination. For some circuits, this RC combination is a reduced form of a slightly more elaborate connection; in others C is the capacitor across which the output voltage of the discriminator is taken, and R is the total equivalent resistance in parallel with C and composed mainly of a low equivalent output resistance that C sees when looking back into the rest of the detector circuit. In any case, the time constant or the equivalent combination that is shown, must be sufficiently low to enable the voltage across the capacitor C to follow the detected voltage. Failure of the voltage across the capacitor to follow the voltage variations dictated by the instantaneous frequency of the signal at the input of the discriminator, will cause the output low-frequency voltage to have an average value (over one frequency-difference cycle) which does not correspond with the value dictated by the frequency of the stronger signal at the input of the receiver. This obviously defeats our purpose, and the restrictions that must be imposed on the RC product to keep this loss of desired average-voltage level from arising will now be determined.

Let us first consider the situation in which the discriminator is either amplitude-insensitive, and hence is not preceded by any limiters, or amplitude-sensitive but preceded by an "infinitely" wideband ideal limiter. In either case, if the discriminator is assumed to have a linear over-all detection characteristic of unit slope, over the whole range of the instantaneous-frequency variations of the input signal, then, under the two-path interference conditions described in Section I, the voltage waveform that the output RC combination must handle is given by

$$e_1(t) = \frac{a \cos rt + a^2}{1 + 2a \cos rt + a^2} \, r \tag{66}$$

This waveform is superimposed upon a direct-voltage component that corresponds to the level dictated by the frequency p of the stronger of the two signals. If the average value of the output capacitor voltage (over a period of $2\pi/r$ sec) is to be maintained at the value dictated by the frequency p, the capacitor must, at every instant of time, tend to charge or discharge at a rate that is faster than (or, at worst, just as fast as) the rate at which the impressed waveform tends to change at that instant. Equivalently, the ratio of the instantaneous value of the capacitor voltage and the RC product must always exceed or, at worst, equal the time derivative of the impressed voltage evaluated at the same instant of time. The total capacitor voltage cannot, therefore, be allowed to go to zero at any time, except when it and the slope of the impressed waveform go to zero simultaneously. This will ensure sufficient rapidity of charging and discharging at all times to enable the capacitor voltage to remain in step with the dictates of the instantaneous frequency of the resultant input signal, and thus will ensure proper detector operation and lack of harmful diagonal clipping effects. Analogous to the similar problem of the limiter time constant, the restriction to be imposed on the RC product is, from condition 58,

$$RC \leqslant \left[-\frac{e(t)}{e'(t)} \right]_{min} \qquad e'(t) \equiv \frac{d}{dt}\, e(t) \tag{67}$$

A direct attempt to substitute from Eq. 66 in Eq. 67, however, meets with frustration, since $e_1(t)$ in Eq. 66 goes to zero when $rt = \cos^{-1}(-a)$, while $e'(t)$ is not zero at that point. In order to avoid this difficulty, the voltage variations described by Eq. 66 must be superimposed upon a steady capacitor voltage that is greater than $ar/(1-a)$. In other words, the expression that must be used for e(t) in condition 67 must be

$$e(t) = E_o + ar \frac{\cos rt + a}{1 + 2a \cos rt + a^2} \tag{68}$$

where

$$E_o > \frac{ar}{1 - a} \tag{69}$$

Since E_o is the voltage level that corresponds to the frequency p and, since p is likely to lie on either side of the center frequency of the i-f passband, it is evident that condition 67 will define a nonzero upper bound on the RC product only if the discriminator output, as seen across the RC combination of Fig. 32, is not balanced to give zero voltage at the center frequency. Furthermore, values of E_o that do not satisfy condition 69 will cause $e(t)/e'(t)$ to have a zero minimum magnitude; hence they will require that RC = 0. The equality sign is excluded in condition 69 for the same reason.

If the expression for e(t), given by Eq. 68, is used, we can show that the ratio

$$-\frac{e(t)}{e'(t)} = \frac{[E_o(1 + a^2) + a^2 r + a(2E_o + r) \cos \phi][1 + 2a \cos \phi + a^2]}{ar^2(1 - a^2) \sin \phi}$$

has a minimum magnitude at the negative real root of

$$\cos^3 \phi - \left[\frac{5}{2} + \frac{E_o(1 + a^4) + a^4 r}{2a^2(2E_o + r)}\right] \cos \phi - \left[\frac{E_o(1 + a^2) + a^2 r}{a(2E_o + r)} + \frac{1 + a^2}{2a}\right] = 0$$

which has a magnitude smaller than unity. An analytical expression for the resultant upper bound on the RC product is available, but it is too cumbersome to be useful.

When the discriminator is preceded by one stage of narrow-band limiting, the formulas of the preceding computation are directly applicable to the computation of the maximum permissible discriminator RC product when the limiter bandwidth is equal to one i-f bandwidth. Here, since the most unfavorable conditions at the output of the limiter filter arise when only A_o and A_{-1} pass and $r = 2\pi W_{if}$, it is only necessary to replace a in the formulas by (A_{-1}/A_o). We recall that the conditions of the present situation apply only in the range $a < 0.84$. A similar extension of the computation to the situation in which more than two limiters, each of which has a bandwidth of one $(BW)_{if}$, precede the discriminator is fairly evident.

It is of interest to observe that with a discriminator whose output across the RC combination which is shown in Fig. 32 is balanced, the upper bound on the RC product will be nonzero as long as the two signals lie symmetrically on opposite sides of the center frequency. For this situation, the upper bound can be shown to be specified by

$$rCR \leqslant \frac{1 - a^2}{2a} \tag{70}$$

which is one-half the upper bound on the grid-circuit time constant of the limiter. Frequently, the capture ratio of a receiver is measured by simulating an interference situation in which the two signals lie at opposite ends of the i-f passband. It is clear that, under this condition of measurement, a discriminator whose output is balanced about the intermediate frequency when it is observed across the RC combination of Fig. 32 will appear to meet the test if its RC product satisfies condition 70. Evidently, under the more general (but milder) interference conditions, it will fail.

In order to illustrate the manner in which the cascading of narrow-band limiters raises the upper bounds on the output-circuit time-constant requirements of the discriminator, we shall choose a discriminator whose characteristic goes through zero at a frequency that corresponds to a cutoff frequency of the i-f amplifier. In view of condition 69, this situation is mainly of academic interest (unless the interference is effectively suppressed before it reaches the discriminator). The weaker of the two carriers will be assumed to fall at the frequency of balance. In the absence of narrow-band limiting, the voltage waveform impressed across the output capacitor is

$$e(t) = r \, \frac{1 + a \cos rt}{1 + 2a \cos rt + a^2}$$

Consequently, the maximum permissible values of $\tau_{do} = W_{if}RC$ are those plotted in Fig. 33. If one narrow-band limiter whose bandwidth equals the i-f bandwidth precedes the discriminator, the upper bounds are specified by the plotted values of ${}_1\tau_{d1}$. The curve marked "${}_2\tau_{d1}$" applies when two limiters, each of one i-f bandwidth, precede the discriminator.

Finally, consider the situation in which the discriminator is preceded by a limiter of bandwidth three times that of the i-f bandwidth. Here, the instantaneous-frequency pattern that accompanies the configuration $M = 1$, $N = 2$, with $r = (BW)_{if}$, dictates the critical requirements. Simple as this configuration may seem, the

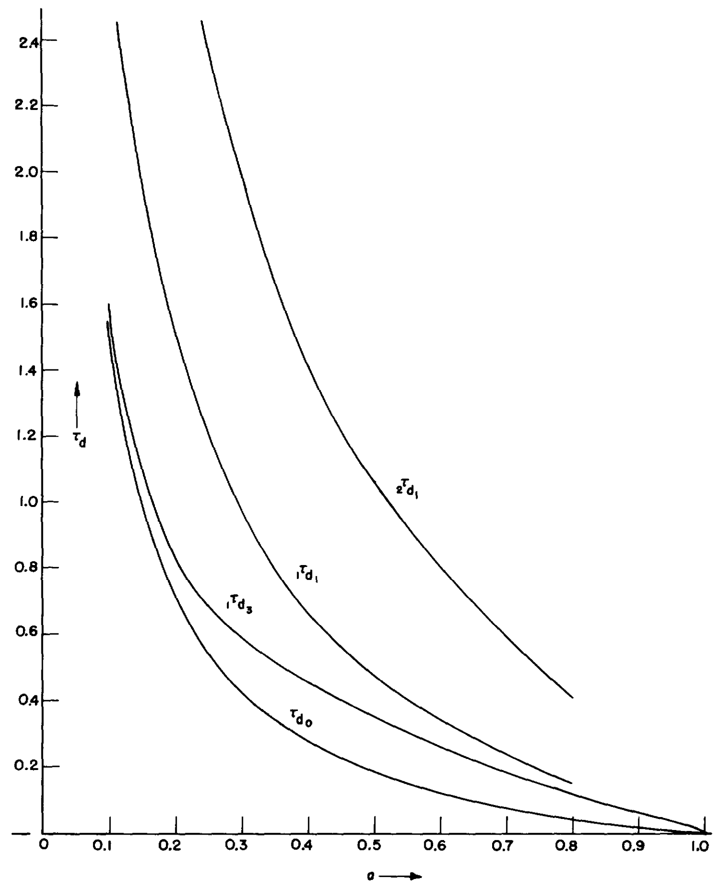

Fig. 33. Curves showing effect of cascading scheme upon the upper bounds on the time constant of the circuit of Fig. 32, under the conditions of the illustrative example.

computation in this case is extremely tedious. The value of $\phi \equiv rt$ that leads to the minimum value of the right-hand member of condition 67 turns out to be a root of the equation

$$\sum_{n=0}^{11} \lambda_n \cos n\phi = 0$$

in which the coefficients, λ_n, are extremely involved. Naturally, the desired root was determined graphically. The maximum permissible values of $_1\tau_{d3}$ are plotted in Fig. 33. The vertical scale in Fig. 33 can be calibrated in microseconds by multiplying by a scale factor appropriate to the value of W_{if} that is used.

It is of interest to observe that comparison of the curves for $_1\tau_{d3}$ and τ_{do} reveals that the percentage by which the upper bound on τ_d has been raised by the action of one limiter of three times the i-f bandwidth is greatest for values of a in the vicinity of a = 1, and decreases rapidly with decreasing values of a. This can be explained by the fact that a limiter bandwidth of $3(BW)_{if}$ approaches more closely the order of

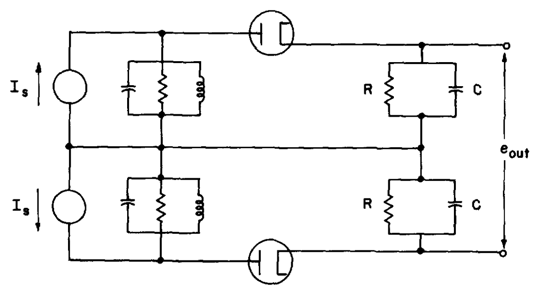

Fig. 34. Common form of balanced discriminator.

bandwidth needed to reproduce the input two-signal frequency-spike pattern, with increasingly less distortion as the value of a is decreased. In terms of the spectrum, the limiting configuration M = 1, N = 2 (i.e., the most unfavorable configuration for $(BW)_{lim} = 3(BW)_{if}$) forms an increasingly important percentage of the significant spectral components at the output of the limiter as the value of a is made smaller. In fact, for a = 0.1, the lowest value considered in the plot, this configuration includes essentially all of the components of significant amplitude (see Table I), and so the resultant instantaneous-frequency pattern (which dictates the curve ${}_1\tau_{d3}$) is a slightly distorted version of the original input two-signal pattern (which dictates the curve τ_{do}).

The plots of Fig. 33 show clearly how the upper bound on the permissible time constant in the output circuit of the discriminator is raised when the discriminator is preceded by a process of narrow-band limiting. The effect produced by one narrow-band limiter will, clearly, be displayed by additional stages. After a sufficient number of stages has been cascaded, the upper-bound on the maximum permissible time-constant values becomes sufficiently high, so that it exercises no important restraint on the design of the output circuit of the discriminator.

We shall next illustrate the computation of the maximum permissible time constant for the common type of discriminator circuit shown in Fig. 34. Alternative forms of this circuit (particularly, the one in which the FM-to-AM conversion is achieved through a double-tuned transformer, in which the top of the primary is connected to the center tap of the secondary) can be manipulated into the form of Fig. 34, and are, therefore, included in this treatment. The output voltage, e_{out}, is a superposition of the envelope of one tank-circuit response upon that of the other tank circuit with reversed polarity.

Assuming that the high-Q tank circuits have the same damping factor α, let one circuit be tuned to a frequency that is $b\alpha$ rad/sec above the center frequency of operation ω_o, and the other to $b\alpha$ rad/sec below ω_o. Also, let the instantaneous frequency of the excitation be deviated by $\alpha\, x(t)$ rad/sec from the center frequency of operation. Then, if the conditions for a quasi-stationary analysis of the tuned-circuit responses are satisfied, and if the RC combination of each peak detector is able to follow the envelope of the corresponding tank-circuit response, the output voltage can be normalized into the form

$$e_{out} = [1 + (x+b)^2]^{-1/2} - [1 + (x-b)^2]^{-1/2} \tag{71}$$

For optimum linearity (2), b must be 1.225. Indeed, with this value of b,

$$e_{out} = k\,x(t) \quad \text{for } -0.6 < x < 0.6$$

The extent of this almost perfect linearity is, therefore, 1.2α rad/sec centered about $x = 0$, while the peak-to-peak separation of the discriminator characteristic is 2.45α rad/sec.

Under conditions of two-signal interference, the instantaneous frequency of the amplitude-limited resultant of the two signals is

$$\omega_i(t) = \omega_{ave} - \frac{1}{2} r \, \frac{1 - a^2}{1 + 2a \cos rt + a^2}$$

where $\omega_{ave} = p + \frac{1}{2} r$, the average of the two signal frequencies. Therefore,

$$x(t) = x_{ave} - \frac{r}{2\alpha} \cdot \frac{1 - a^2}{1 + 2a \cos rt + a^2} \tag{72}$$

where $x_{ave} = \frac{1}{\alpha}(\omega_{ave} - \omega_o)$. When the two signals are symmetrically disposed with respect to the center frequency, $x_{ave} = 0$, and

$$x(t) = -\frac{r}{2\alpha} \, \frac{1 - a^2}{1 + 2a \cos rt + a^2} \tag{73}$$

If r is given its maximum value of $(BW)_{if}$, Eq. 73 will represent the most troublesome instantaneous-frequency deviations from ω_o, measured in units of α rad/sec.

Results that will be published in a later report indicate that if

$$\frac{r}{2\alpha} = \beta \, \frac{1 - a}{1 + a} \tag{74}$$

then the conditions for quasi-stationary analysis, which justify the use of Eq. 71, are satisfied if β is bounded by the values plotted in Fig. 35. Substitution in Eq. 73 leads to

$$x(t) = \frac{-\beta(1-a)^2}{1 + 2a \cos rt + a^2} \tag{75}$$

The upper bound on the RC product can now be found by requiring that the RC combination follow the voltage waveform given by

$$e(t) = \left\{1 + [x(t) + 1.225]^2\right\}^{-1/2} \tag{76}$$

x(t) being given by Eq. 75. In Fig. 36 we present the results of a computation in which

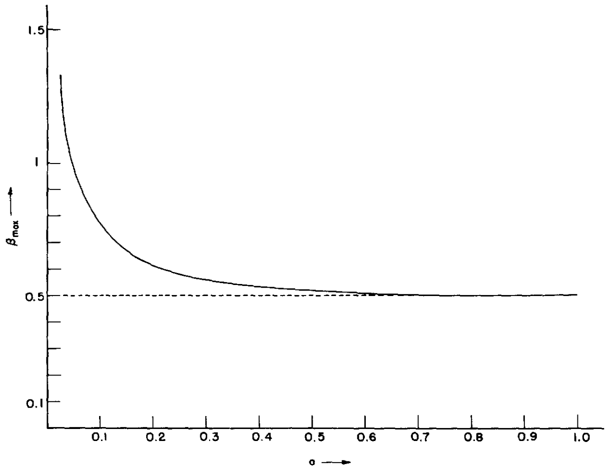

Fig. 35. The upper bounds on the value of β in Eq. 74.

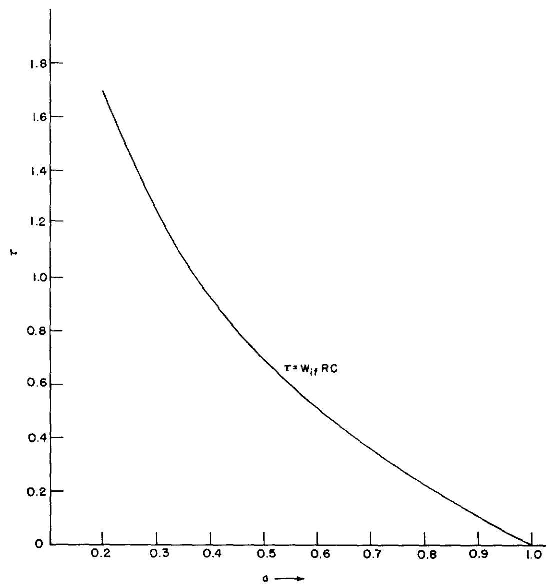

Fig. 36. Maximum permissible values of $\tau = W_{if}RC$ for each half of the balanced discriminator.

use was made of the values of $\beta = \beta_{max}$ that are plotted in Fig. 35. For all values of $a \leq 0.84$, the plot of Fig. 36 can be used in conjunction with the plot of Fig. 20 to determine the maximum permissible values of $\tau = W_{if}RC$ when any specified number of narrow-band limiters, each of one i-f bandwidth, are cascaded ahead of the discriminator. Multiplication by a scale factor appropriate to the value of W_{if} that is used will convert the dimensionless vertical scale in Fig. 36 into microseconds.

2.3 HARMONIC STRUCTURE OF DETECTED DISTURBANCE

In general, after the instantaneous-frequency variations are properly translated into instantaneous-voltage variations (about the direct-voltage level dictated by the frequency of the stronger signal), the voltage variations are modified by the action of the de-emphasis and the audio filters that follow the discriminator circuit. If the frequency difference between the two input carriers lies beyond the range of audibility (as it will much of the time in long-distance communication, but less frequently in communication over shorter distances), the Fourier components of the recurrent spike train will all be filtered out by the two low-pass filters. This will, effectively, be the end of the disturbance caused by the presence of the weaker signal within the i-f passband. However, if the frequency difference between the two paths is audible, the component with the fundamental frequency of recurrence, plus a number of harmonics, depending upon the position of this frequency in the audible spectrum, will pass through the low-pass filters, and will, therefore, disturb the output signal.

Note that two factors play more or less obvious roles in minimizing the importance of the disturbance that leaks through: the action of the two low-pass filters in rejecting most of the harmonic components, and the fact that the magnitude of the interference spikes (hence the amplitude of each constituent Fourier component of the detected spike train) is directly proportional to the value of the frequency difference. Consequently, even though more undesired harmonics of the frequency difference are likely to get through the low-pass filters as this frequency difference assumes lower and lower values in the audio range, the amplitudes of the passed components will also be lower and lower.

A third factor that tends to minimize the importance of the audible disturbance is brought about by the effect of narrow-band limiting upon the shape and magnitude of the spike trains. This distortion of the extraneous modulation by the narrow-band filter after the limiter generally produces the result that the Fourier components of the modified spike train have smaller amplitudes than their counterparts in the structure of the undistorted waveform. The effect becomes increasingly more pronounced with increasing values of the frequency difference r and the ratio of weaker-to-stronger signal a. But the lower the value of the beat frequency r or the ratio a, the less significant is this effect. The reason for the dependence of this effect upon the magnitude of the frequency difference r is readily appreciated from the fact that as r gets lower, the number of significant sideband components accommodated within the limiter-filter passband increases. With a given constant value of a, the instantaneous-frequency variations of the resultant signal tend to become an increasingly less distorted copy of the instantaneous-frequency pattern of the amplitude-limited resultant of the two signals delivered by the i-f amplifier. When r is well within the audio range, the effect will tend to be negligible, especially with the lower values of a, even when the smallest permissible value of one $(BW)_{if}$ is used after the limiter. The instantaneous-frequency variations introduced by the presence of the weaker signal will then tend to be indistinguishable (hence less separable) in their characteristics (pertaining to spike magnitude, or maximum deviation, repetition rate, and maximum time rate of change during a cycle of the fundamental) from the variations that represent the desired message modulation. When r is held constant at some audible value, the distortion of the spike train by the narrow-band filtering after the limiter will have an increasingly noticeable effect with increasing values of a; therefore the effect of this distortion upon the amplitudes of the harmonic components in the structure of the detected spikes (which will be illustrated presently), becomes increasingly significant, and vice versa.

Consider the situation in which the two input carriers accommodated within the i-f passband have constant amplitudes and frequencies, and are impressed directly upon an amplitude-insensitive discriminator (or a discriminator which is preceded by an ideal limiter that passes essentially all of the significant spectrum centered about the frequency of the stronger signal). The detection of the instantaneous-frequency variations of the resultant signal results, at the output of the discriminator, in a voltage proportional to

$$\begin{aligned} f(t) &= r\,\frac{a^2 + a\cos rt}{1 + 2a\cos rt + a^2} \\ &= -r \sum_{n=1}^{\infty} (-a)^n \cos nrt \end{aligned} \tag{77}$$

Equation 77 shows that the amplitude of each harmonic varies directly with r. As a function of a, the amplitude of the fundamental component is directly proportional to a, and the amplitude of the n^{th} harmonic component relative to the amplitude of the fundamental component is given by a^{n-1}. Plots of the relative amplitudes of the various harmonics, as compared with the amplitude of the fundamental, are shown in Fig. 37a, b, and c, in which each is marked "a^n-curve."

When an ideal limiter of some specified bandwidth $W_L = (BW)_{lim}/(BW)_{if}$ is inserted in the path of the resultant signal before it gets to the discriminator, we have found that the action of the narrow-band limiting process damps out the fast and large excursions of the instantaneous frequency of the resultant signal that goes through. The effect of the resultant modifications in the waveform of the instantaneous-frequency variations upon the amplitudes of the harmonic components in the structure of this waveform is

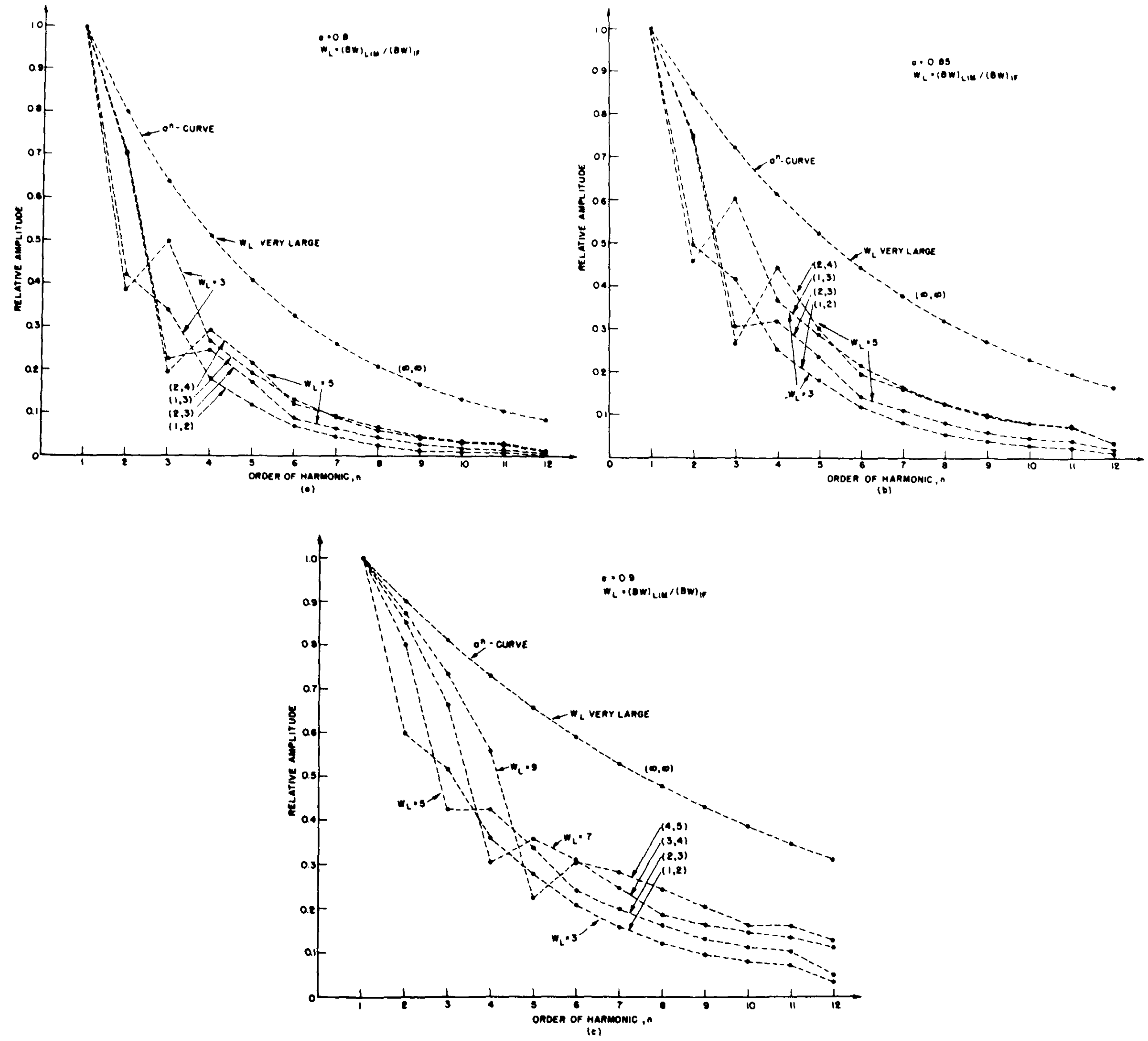

Fig. 37. Relative amplitudes of the Fourier components in the structure of the detected instantaneous-frequency disturbance.

best studied quantitatively by direct Fourier analysis of a few typical, informative cases. The general character of this effect may also be anticipated on the basis of the concept of the equivalent interference ratio. This concept recognizes the importance of the repetition frequency r and the spike magnitude of the instantaneous-frequency spike train of the resultant signal at the output of a narrow-band limiter in providing a basis for comparing the capture conditions at that point with conditions elsewhere in the receiver. On this basis, the effect of narrow-band limiting upon the capture conditions is equivalent to a reduction in the equivalent interference ratio. Consequently, this reduction should generally result in a reduction of the relative amplitudes of the various harmonics in the structure of the instantaneous-frequency waveform. This effect is illustrated by the plots of Fig. 37a, b, and c. The examples chosen for illustration correspond to resultant signals composed of a number of lower sideband components M, and a number of upper-sideband components N, these numbers being indicated in parenthesis in the order (M, N). Each configuration (M, N) is associated with the proper set of points by an arrow. The various configurations that are indicated give rise to most troublesome resultants when they are associated with the limiter bandwidths W_L, whose values are indicated in the plots. When the indicated values of W_L are used, the values of r for which these configurations can arise may not be audible, depending upon the i-f bandwidth that is used. If smaller values of W_L are used, such as unity for $a = 0.8$ or $a = 0.85$, some of the indicated configurations may arise with audible values of r. In any case, these configurations were chosen only because of the convenience of illustration.

The computations leading to the plots of Fig. 37a, b, and c have shown that the absolute values of the harmonic components are generally decreased by the narrow-band limiting effect below the corresponding values given by a^n in the absence of narrow-band limiting. Figure 37 illustrates the decrease in the relative amplitudes of the harmonics, as compared with the corresponding fundamental component. These results illustrate the effect of one stage of narrow-band limiting — the first one. A second or later stage will usually have a less troublesome signal to cope with at its input than the first or earlier intermediate stages. Nevertheless, these later stages will exhibit the same effect as long as the instantaneous-frequency variations caused by the interference differ from the kind of expected message modulation by virtue of characteristics that enable a narrow-band filter to distinguish them.

III. CONCLUDING REMARKS

In this report the interest has been centered on the effect of a process of ideal amplitude limiting followed by ideal filtering (or frequency limiting) upon the instantaneous frequency of the resultant of two carriers that differ in frequency as well as in amplitude. The main objective has been to determine the necessary changes in the basic design requirements, in order to secure proper operation under the most adverse interference conditions, with the express purpose of enhancing the capture of the stronger signal. Our philosophy has not been guided merely by a desire to specify the requirements that would lead to the realization of an ideal frequency demodulator which would be insensitive to amplitude changes and would meet the bandwidth requirements dictated by the extraneous instantaneous-frequency variations caused by the presence of the interference.

We have also recognized certain fundamental features in the nature of serious interference, as it is usually constituted, and have chosen to take advantage of these features in preparing the receiving circuits, so that by proper design the disturbance will be minimized or completely eliminated. This attitude has led us to a basic change in the approach to the question of limiter bandwidth requirement, and has brought to light a new philosophy concerning the limiter's share of the task of interference suppression in FM reception. We now ask the limiter stage to do more than just eliminate undesirable changes in the amplitude – we also require that it contribute to the abatement of the FM disturbances wrought by the presence of the weaker signal, by decreasing the range and intensity of the extraneous instantaneous-frequency variations of the resultant signal through a process of instantaneous-frequency limiting or, perhaps, quasi-limiting. Evidently, the undesirable frequency changes must differ in the degree of their extent, rate of change, and rate of recurrence – any one of these or all of them combined – from the changes that the usual message causes in the instantaneous frequency of the carrier, in order that the narrow-band-limiting effect will not also distort the desired message modulation. Granted these differences, the limiter proper (in this new task) prepares the resultant signal for the (frequency-limiting) treatment by eliminating the amplitude changes. This spreads out the significant spectrum, and the sluggish (narrow-band) filter, immediately following, performs the (frequency-limiting) operation on the instantaneous frequency of the resultant signal by refusing to follow the more drastic frequency variations or, equivalently, by eliminating portions of the spread-out spectrum that owe their existence to the interfering signal and which are only necessary for the undistorted reproduction of the instantaneous-frequency spike pattern of the amplitude-limited resultant of the two input sinusoids. The minimum requirement of one i-f bandwidth for the limiter filter is calculated to meet the prerequisites of undistorted reproduction of the expected message modulation.

This situation is analogous in philosophy to the well-known procedures for eliminating certain types of impulsive AM noise and interference. A clipper is introduced in the path of the input signal, whose threshold exceeds the maximum value of instantaneous amplitude that is expected with the desired message amplitude modulation, in order to leave the desired message modulation unaffected. The impulsive interference should, therefore, exceed the clipping threshold, and the duration of the individual impulses should be short compared with the period of a desired modulating frequency for the "noise-silencing" scheme to introduce noticeable improvement in the reception. Here again we capitalize fundamental differences between the resultant waveforms in the presence or absence of the interference. In other words, the modulation introduced by the interference must differ in a manner that can be distinguished by the circuitry in the path of the signal from the changes introduced by the desired message modulation to gain abatement of the interference through special arrangement or design that does not significantly affect the desired message.

Generally speaking, interference can be described as any extraneous modification of the instantaneous variations of the message-modulated parameter of the carrier wave. The suppression of interference may be tackled in the radiofrequency or in the low-frequency sections of the receiver or in both. Wherever the extraneous effects of interference are to be minimized or eliminated, it is important to realize that what we may call the "fundamental principle of interference rejection" forms the basis of any effective interference-suppressing scheme. Interference can be suppressed if its disturbance is in the form of modifications in the instantaneous variations of the message-bearing parameter of the carrier wave that are fundamentally distinguishable from the variations that an expected message modulation would inflict upon this parameter. If the extraneous variations cannot be distinguished from the variations caused by the message modulation, then the interference cannot be suppressed. A successful scheme for interference suppression would have to be capable of discriminating against the characteristic features of the disturbance which are not normally expected in proper message modulation, without affecting significantly the message modulation itself. Recognition and appreciation of these facts helps us in accounting for the pronounced capture possibilities of an FM system, and the inherent vulnerability of an AM system, which is particularly manifest with cochannel disturbances.

The most important distinguishable feature of an FM disturbance is the highest rate of variation in the instantaneous frequency of the resultant signal which is caused by the interference. This rate of instantaneous-frequency variation combines in one package the highest frequency deviation, as well as the repetition rate of this deviation. Under conditions of high-level interference, this rate is sufficiently higher than the highest rate of variation that can be expected in the message modulation so that it is possible to insert filters, at appropriate places in the signal path, that would be too sluggish to follow the disturbance, without noticeably distorting the message modulation. The appropriate places for these filters in the high-frequency sections of the receiver are not in the linear stages, because in these stages the desired carrier and the interference combine linearly and their resultant spectrum is fully accommodated within the i-f passband. The concentration of the spectrum of this resultant signal within the i-f passband makes it impossible to separate the two signals, or to improve the predominance of the stronger one, purely by linear filtering. But a process of ideal amplitude limiting will spread over several i-f bandwidths the significant spectrum that is necessary for the reproduction of the original frequency disturbance of the interference. The extent of this range increases with an increase in the gravity of the interference which, in turn, results from a decrease in the amplitude difference and an increase in the frequency difference between the two signals. Since the instantaneous frequency of the desired signal will always place it within the extent of one i-f bandwidth, we recognize immediately that we can filter after the limiting process to exclude sizeable portions of the interference spectrum, without affecting the message-bearing spectrum.

One stage of limiting and filtering, however, will still retain at its output a spectrum with a significant amount of the interference in the form of components that could not be rejected without impairing some important phase of the operation. Therefore, subjecting the resultant of this retained spectrum to a process of ideal limiting will spread out the spectrum of the retained disturbance, again, over a sufficiently wider frequency range to enable additional filtering to be effective. This cycle of spreading out of the interference spectrum, followed by rejecting the outer portions of the significant spread-out spectrum, may be repeated until so little is left of the disturbance that the significant spectrum of the amplitude-limited resultant signal at the end of the chain becomes essentially confined within the limits of the permissible minimum passband. At this point, the maximum rate at which the disturbance will vary the instantaneous frequency of the resultant signal becomes comparable, and almost indistinguishable, from the variations that may arise with the expected message modulation. In other words, the cascading scheme will continue to sap the energy in the spectrum of the disturbance until the remaining energy gives rise to a spectrum that is not significantly distinguishable in extent from the spectrum that can arise with the usual message modulation. Phrased differently, the cascading of limiters followed by sluggish filters will remain effective in the abatement of the disturbance until the variations in the instantaneous frequency caused by this disturbance begin to resemble the variations that the message modulation may be expected to cause. Beyond this point, continuation of this scheme is not profitable.

Thus a properly designed narrow-band limiter in the path of the resultant of two signals that differ in strength by an arbitrarily small amount will modify the character of this resultant signal in such a way that the effective disturbance caused by the presence of the weaker signal is reduced and the capture of the stronger signal is enhanced. The amount of improvement in the favorable conditions for capture of the stronger signal, per stage of narrow-band limiting, is predictable in accordance with the techniques and results of this report. The degree of improvement achieved per stage is greatest under the most adverse interference conditions. In general, these conditions prevail when the two carriers are farthest apart in frequency, while their individual frequency modulations are slow. As the frequency difference between the two carriers decreases, the intensity of the disturbance will decrease, and so will the degree of improvement in the capture conditions that are achievable with each stage of narrow-band limiting. When the frequency difference decreases to a value r_{min}, which is specifiable as a small fraction of the limiter-filter bandwidth, the extraneous modulation caused by the interference becomes sufficiently slow for the filter to follow it through quasi-stationary states, and the disturbance will pass through unabated. The closer the amplitude interference ratio approaches unity, the smaller will be the value of r_{min} that marks the limit of noticeable improvement in the capture. An analytic expression for r_{min} has been derived, and will be presented in a later report.

Acknowledgment

The author is grateful to Professor Jerome B. Wiesner for his keen interest in, and critical supervision of, this research; to Professor Ernst A. Guillemin for his enthusiastic encouragement in many informal discussions of the work; and to Professor Peter Elias for his helpful criticism.

References

1. J. Granlund, Interference in frequency-modulation reception, Technical Report 42, Research Laboratory of Electronics, M.I.T., Jan. 20, 1949.
2. L. B. Arguimbau, Vacuum Tube Circuits (John Wiley and Sons, Inc., New York, 1948).
3. F. L. H. M. Stumpers, Theory of frequency-modulation noise, Proc. IRE 36, 1081 (Sept. 1948).
4. H. von Ziepel, Encyklopädie der mathematischen Wissenschaften, Bd. 6, 2, 557-665 (Teubner, Leipzig, 1905-1923).
5. D. Bierens de Haan, Nouvelles Tables D'intégrales Définies (G. E. Stechert and Co., New York, 1939), Table 67, No. 2.
6. H. Hancock, Theory of Elliptic Functions, Vol. I (John Wiley and Sons, Inc., New York, 1910), p. 181.
7. W. Grobner and N. Hofreiter, Integraltafel, Zweiter Teil, Bestimmte Integrale (Springer Verlag, Wien and Innsbruck, 1950) p. 1, No. 1.
8. Ibid., p. 331, Nos. 93 and 94.
9. H. T. Davis, Tables of the Higher Mathematical Functions, Vol. I (The Principia Press, Bloomington, Indiana, 1933).
10. E. A. Guillemin, The mathematics of Circuit Analysis (John Wiley and Sons, Inc., New York, 1949).