#3

THEORY OF LOW-DISTORTION TRANSMISSION OF FM SIGNALS THROUGH LINEAR SYSTEMS

ELIE J. BAGHDADY

LOAN COPY only

TECHNICAL REPORT 332

JULY 30, 1957

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 332

July 30, 1957

THEORY OF LOW-DISTORTION TRANSMISSION OF FM SIGNALS THROUGH LINEAR SYSTEMS

Elie J. Baghdady

Abstract

The convergence properties of the Carson and Fry and of the van der Pol-Stumpers expansions for the complex amplitude of the steady-state response of a filter to an FM excitation are discussed. The theory of quasi-stationary analysis and of FM-to-AM conversion with low distortion is presented. The use of Taylor's formula leads to error estimates and to a simple condition that specifies an upper bound on the error incurred in restricting the solution to the quasi-stationary term (the instantaneous-frequency method). A sluggishness ratio and an index of stiffness are defined for filters whose system functions have poles only in the left half-plane. Sluggishness ratios and indices are given for various filters of wide practical interest. The results indicate that filter bandwidths must be prescribed on the basis of the fastest rate at which the instantaneous frequency of the excitation is swept. Applications to sinusoidal modulation and to two-path interference that emphasize the limitations of specifying filter bandwidths on the sole basis of frequency deviation are offered. Applications to the reproduction of FM video waveforms are discussed. The discussion concludes with an analysis of harmonic and intermodulation distortion in the quasi-stationary response.

I. INTRODUCTION

The present study has been motivated by at least two important questions. First, the conditions for satisfactory reproduction of a frequency modulation in the response of a bandpass filter do not appear to have been sharply defined in general terms in published work. The specification of filter bandwidths in frequency-modulation systems has so far been largely based on the extent of the significant spectrum of the FM wave without sufficient regard to the type of filter used. Moreover, some misconceptions appear to have grown with regard to the significance of a high modulation index (or deviation ratio) and its usefulness in the classification of FM waves into those whose significant spectra are confined essentially within the bounds of the maximum deviation from the center frequency and those whose significant spectra are not so confined.

The second question concerns the definition of the conditions for the validity of using the concept of instantaneous frequency as a tool in the solution of problems on the response of a linear system to variable-frequency excitations. The problem of designing discriminator filters for low-distortion conversion of FM to AM is but one important instance in which these conditions are of fundamental importance. More generally, in the solution of forced-response problems in modulation theory, the need for sharply defining the conditions for reasoning on an instantaneous-frequency (or instantaneous-amplitude for AM) basis also arises when the filter is viewed as a system which, by virtue of its energy-storage elements, will exercise an inertia or sluggishness that sets a limit on the kind of frequency (or amplitude) changes that will be reproduced in its response. A widely used alternative approach views the filter as a selector of certain frequency bands. This spectral approach is commonly referred to as the "Fourier method." First, the spectrum of the impressed signal is determined. Then, the spectrum of the steady-state response of the filter is obtained by multiplying each input spectral component by the value that the system function of the filter assumes at the frequency of the input component. This method is conceptually simple and straightforward. It is often the final resort in checking the validity and accuracy of results obtained by other methods. Unfortunately, when the number of significant spectral components is large, the computation becomes extremely laborious, and, in perspective, the significance of the results may be lost in a maze of complex numbers. In this connection, it is often convenient to idealize the filter amplitude and phase characteristics in an effort to simplify the computation and facilitate the reasoning — particularly in studying the effect of the various portions of the impressed spectrum upon the instantaneous frequency and amplitude of the resultant steady-state response (1,2). Despite the fact that the idealization of the filter characteristics (which constitutes an acceptable deviation from reality) can be quite helpful, the analysis often remains laborious.

It is the main object of the present study to explore the conditions that a general, linear, system function must satisfy in order to ensure an acceptable reproduction of a specified frequency modulation (when this function pertains to a selective filter) or a proper FM-to-AM conversion of this modulation (when the system function characterizes a discriminator circuit). It is fairly evident that these conditions must involve joint restrictions on the character of the modulating function, as well as on the system function. A criterion is presented for the specification of filter bandwidths which emphasizes the essential characteristics of the filter and of the signal, and takes these characteristics into account jointly. In adducing this criterion, we recognize that a fundamental requirement for low-distortion transmission of an FM wave through a bandpass filter, and for proper conversion of the FM into AM by a filter, is that the filter follow the frequency-modulated excitation in a quasi-stationary manner. This means that we must be able to analyze the filter response by viewing the FM excitation essentially in terms of the resultant of all of its constituent spectral components, and visualizing it on the plots of the filter characteristics (versus frequency) as a sinusoid whose instantaneous position on the frequency scale varies in accordance with the dictates of the modulating wave. To be sure, this condition is not sufficient for either of the two aforementioned purposes. For faithful FM-to-AM conversion we must require in addition that the swept portion of the amplitude characteristic be linear. For low-distortion transmission of the frequency modulation we must, in addition, require that the phase characteristic of the filter be linear.

The conditions for quasi-stationary analysis have been studied by Carson and Fry and by van der Pol and Stumpers among others. Our purpose here is to present a new analysis which leads to a more complete as well as more concise account of these conditions and emphasizes their bearing upon the choice of filter bandwidths in frequency-modulation systems.

II. TWO IMPORTANT EXPANSIONS AND THEIR CONVERGENCE PROPERTIES

Consider a linear (realizable and stable) system which is characterized by its response, $h(t)$, to a unit-impulse excitation. The time function, $h(t)$, has a bounded envelope made up of decaying exponentials. Impulses in $h(t)$ are possible only at $t = 0$.

Let this system be excited by a current described by

$$i(t) = e^{j[\omega_c t + \theta(t)]} \tag{1}$$

where $\theta(t)$ is some arbitrary function of time. If the forced-response voltage, after the initial transients have died out, is denoted by

$$e(t) = E(t)e^{j[\omega_c t + \theta(t)]} \tag{2}$$

we have, from a well-known superposition integral,

$$e(t) = \int_0^\infty h(\tau)\, i(t - \tau)\, d\tau \tag{3}$$

Substitution from Eqs. (1) and (2) yields

$$E(t) = \int_0^\infty h(\tau)e^{-j[\omega_c \tau + \theta(t) - \theta(t - \tau)]}\, d\tau \tag{4}$$

Two useful expansions of the integral on the right-hand side are possible. These expansions depend upon expressing the integrand in the form

$$h(\tau)e^{-j\omega\tau} \cdot g(t, \tau)$$

where, in generating the first expansion, $\omega = \omega_c$, while for the second expansion

$$\omega = \omega_i(t) = \frac{d}{dt}\,[\omega_c t + \theta(t)]$$

The first expansion brings out discriminator action (3); the second is more appropriate for adducing the conditions for the validity of a quasi-stationary analysis (4,5,6). The derivations of these expansions, which we present in Table I, follow parallel steps and bring out their convergence properties. The functions $B_n(t)$ and $C_n(t)$ are defined in

Table I

Development of the Expansions for the Complex Amplitude, E(t), of the Steady-State Filter Response to a Variable-Frequency Excitation, $\exp j[\omega_c t + \theta(t)]$.

Carson and Fry Expansion	van der Pol-Stumpers Expansion

1. $$E(t) = \int_0^\infty h(\tau) e^{-j[\omega_c t + \theta(t) - \theta(t-\tau)]} d\tau = \int_0^\infty h(\tau) e^{-j\omega\tau} \cdot g(t, \tau)\, d\tau$$

h(t) = impulse response of linear passive network

Carson and Fry Expansion	van der Pol-Stumpers Expansion
$\omega \equiv \omega_c$ and $g(t, \tau) = e^{-j\theta(t)} \cdot e^{j\theta(t-\tau)}$	$\omega \equiv \omega_i = \omega_c + d\theta/dt$, and $g(t, \tau) = e^{j[\theta(t-\tau) - \theta(t) + \tau\theta'(t)]}$

2. Assume that $\theta(t)$ has finite derivatives of all orders for all values of t. Thus,

$$\theta(t-\tau) = \sum_{n=0}^{\infty} \frac{\theta^{(n)}(t)}{n!} (-\tau)^n$$ will converge uniformly and absolutely for all values of t and τ.

Note that $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$ has an infinite range of uniform and absolute convergence.

3. Therefore,

Carson and Fry Expansion	van der Pol-Stumpers Expansion
$g(t, \tau) = \sum_{n=0}^{\infty} \frac{1}{n!} B_n(t)(-\tau)^n$	$g(t, \tau) = \sum_{n=0}^{\infty} \frac{1}{n!} C_n(t)(-\tau)^n$

converges uniformly for all values of t and τ. $B_n(t)$ and $C_n(t)$ are given in Table II.

4. Substitute from 3 in 1, interchange summation and integration, and obtain

Carson and Fry Expansion	van der Pol-Stumpers Expansion
$E(t) = \sum_{n=0}^{\infty} \frac{1}{n!} B_n(t) \int_0^\infty (-\tau)^n h(\tau) e^{-j\omega_c \tau} d\tau$	$E(t) = \sum_{n=0}^{\infty} \frac{1}{n!} C_n(t) \int_0^\infty (-\tau)^n h(\tau) e^{-j\omega_i \tau} d\tau$

5. If we write

$$Z(j\omega) = \int_0^\infty h(\tau) e^{-j\omega\tau} d\tau, \text{ then } Z^{(n)}(j\omega) = \frac{d^n Z(j\omega)}{d(j\omega)^n} = \int_0^\infty (-\tau)^n h(\tau) e^{-j\omega\tau} d\tau$$

6. Substitute from 5 in 4 to obtain

Carson and Fry Expansion	van der Pol-Stumpers Expansion
$E(t) = \sum_{n=0}^{\infty} \frac{1}{n!} B_n(t)\, Z^{(n)}(j\omega_c)$	$E(t) = \sum_{n=0}^{\infty} \frac{1}{n!} C_n(t)\, Z^{(n)}(j\omega_i)$

For assumed properties of $\theta(t)$ and for a Z(s) that is analytic in right-half of the s-plane and on $j\omega$-axis, these expansions converge uniformly for all values of t.

Table II

The Functions $B_n(t)$ and $C_n(t)$ of Table I

n	$B_n(t) \equiv e^{-j\theta(t)} \dfrac{d^n}{dt^n} e^{j\theta(t)}$	$\dfrac{1}{n!} C_n(t)$
0	1	1
1	$j\, d\theta/dt$	0
2	$-\left(\dfrac{d\theta}{dt}\right)^2 + j\,\dfrac{d^2\theta}{dt^2}$	$j\,\dfrac{\theta''(t)}{2!}$
3	----	$j\,\dfrac{\theta'''(t)}{3!}$
4	$B_{n+1} = \left[j\dfrac{d\theta}{dt} + \dfrac{d}{dt}\right] B_n$	$j\,\dfrac{\theta^{iv}}{4!} + \dfrac{1}{2!}\left[j\,\dfrac{\theta''}{2!}\right]^2$
5		$j\,\dfrac{\theta^{v}}{5!} + \left[j\,\dfrac{\theta''}{2!}\right]\left[j\,\dfrac{\theta'''}{3!}\right]$
6		$j\,\dfrac{\theta^{vi}}{6!} + \left[j\,\dfrac{\theta''}{2!}\right]\left[j\,\dfrac{\theta^{iv}}{4!}\right] + \dfrac{1}{2!}\left[j\,\dfrac{\theta'''}{3!}\right]^2 + \dfrac{1}{3!}\left[j\dfrac{\theta''}{2!}\right]^3$
7		$j\dfrac{\theta^{vii}}{7!} + \left[j\dfrac{\theta''}{2!}\right]\left[j\,\dfrac{\theta^{v}}{5!}\right] + \left[j\,\dfrac{\theta'''}{3!}\right]\left[j\,\dfrac{\theta^{iv}}{4!}\right] + \dfrac{1}{2!}\left[j\dfrac{\theta''}{2!}\right]^2\left[j\,\dfrac{\theta'''}{3!}\right]$
8		$j\,\dfrac{\theta^{viii}}{8!} + \dfrac{1}{2!}\left[j\,\dfrac{\theta''}{2!}\right]^2\left[j\,\dfrac{\theta^{iv}}{4!}\right] + \dfrac{1}{2!}\left[j\,\dfrac{\theta'''}{3!}\right]^2\left[j\,\dfrac{\theta''}{2!}\right]$
		$+\left[j\dfrac{\theta''}{2!}\right]\left[j\,\dfrac{\theta^{vi}}{6!}\right] + \left[j\,\dfrac{\theta'''}{3!}\right]\left[j\,\dfrac{\theta^{v}}{5!}\right] + \dfrac{1}{2!}\left[j\,\dfrac{\theta^{iv}}{4!}\right]^2 + \dfrac{1}{4!}\left[j\,\dfrac{\theta''}{2!}\right]^4$

$\dfrac{1}{n!} C_n(t)$ is the sum of all possible product terms

$$\prod_{p,q,m} \left[j\,\frac{\theta^{(p)}}{p!}\right] \cdot \frac{1}{m!}\left[j\,\frac{\theta^{(q)}}{q!}\right]^m \quad \text{that satisfy the requirement}$$

$$\sum_p p + \sum_{m,q} mq = n,$$

where m = 0, 1, 2, 3, ... , and p and q are positive integers (different from unity) that are expressed in Roman numerals to indicate p and q differentiations with respect to the independent variable, t.

Table II. The convergence of these expansions is not affected by the detailed properties of $Z(j\omega)$ or its derivatives, as long as this system function has associated with it an impulse-response function $h(t)$ which is bounded and composed of decaying exponentials and damped oscillations. These properties pertain to all linear, passive networks whose system functions have no poles on the $j\omega$-axis. With reference to Table I, poles of $Z(s)$ on the $j\omega$-axis must be excluded, because their presence gives rise to sustained oscillation in $h(t)$ for which the convergence of the integrals in steps 4 and 5 is defined only by limit processes that lead to singularity functions (impulses and n-tuplets).

In addition to ensuring the proper convergence of the integral that defines $Z(j\omega)$, the boundedness of $h(t)$ is required so that when every term of the uniformly convergent series of step 3 is multiplied by $h(\tau)e^{-j\omega\tau}$, the resulting series is also uniformly convergent, and, hence, integrable term by term to yield another uniformly convergent series.

As for the convergence of the power-series expansion in step 3, we note that such an expansion will converge over an infinite range if $\theta(t)$ itself possesses a power expansion whose interval of convergence is not bounded. This may be justified as follows. Let us write

$$e^{j\theta(t-\tau)} = \sum_{n=0}^{\infty} \frac{1}{n!} \left[j\theta(t-\tau) \right]^n$$

which, for a bounded $\theta(t)$, is uniformly and absolutely convergent, with an infinite radius of convergence in terms of $\theta(t)$, as well as of t or τ. If

$$\theta(t-\tau) = \sum_{k=0}^{\infty} \frac{1}{k!} \theta^{(k)}(t)(-\tau)^k$$

is uniformly and absolutely convergent over an infinite range, the same will also be true of

$$e^{j\theta(t-\tau)} = \sum_{n=0}^{\infty} \frac{1}{n!} \left[j \sum_{k=0}^{\infty} \frac{1}{k!} \theta^{(k)}(t)(-\tau)^k \right]^n$$

This conclusion follows from a theorem in the theory of power series (7). From the uniqueness property of a power-series expansion, it is evident that if the right-hand side of this last equation is expanded and rearranged in ascending powers of τ, the result will be identical with the series for $\exp[j\theta(t-\tau)]$ that is used in step 3.

The power-series expansion of a sinusoid converges uniformly and absolutely for all finite values of its argument. If $\theta(t)$ is composed of a finite sum of sinusoidal terms, then its power-series expansion may be considered as evolving from the addition of the separate power-series expansions of its sinusoidal components. Since each of the component power expansions converges uniformly and absolutely over an essentially infinite range, the power series that results from the combination of the finite number of terms of like powers will also converge uniformly and absolutely over an infinite range. Consequently, if $\theta(t)$ possesses a finite spectrum, or if the spectrum of $\theta(t)$ can be considered finite without introducing a significant error, then $\theta(t)$, as well as all of its derivatives, is continuous everywhere, and a power-series expansion for $\theta(t)$ will converge uniformly and absolutely over an infinite range. All physical, periodic, modulating functions fall into this category, with obvious exceptions — for example, the functions associated with modulated waves whose envelopes contribute to their zero crossings. The property of Fourier analyzability for $\theta(t)$ is not enough to ensure the desired convergence properties of power-series expansions of $\theta(t)$.

An important property of the van der Pol-Stumpers expansion was demonstrated by Stumpers. Stumpers assumed a periodic frequency modulation, s(rt), of fundamental repetition frequency r rad/sec, and derived an expansion for the particular integral of the linear differential equation with constant coefficients which relates the excitation and response in a linear, lumped, finite, and passive system. The complex amplitude of this response can be expressed as the series

$$E(rt) = \sum_{n=0}^{\infty} G_n(rt)\, r^n \tag{5}$$

which is asymptotic for $r \rightarrow 0$. The functions $G_n(rt)$ for the earlier terms are presented in Table III. This expansion is readily obtainable from the expansion in Table I by substituting

$$\theta^{(n)}(t) = r^{n-1} s^{(n-1)}(rt)$$

and then grouping the result in terms of ascending powers of r. We may therefore conclude that when the frequency modulation of the excitation is periodic, the complex amplitude of the forced response can be expanded into an asymptotic series about $r = 0$, where r is the fundamental angular frequency of the modulation.

It is now recalled that if the remainder after the sum of the first $n + 1$ terms of an expansion in powers of r is denoted by $R_n(r)$, then this expansion is defined as being asymptotic for $r \rightarrow 0$ if, for a fixed value of n,

$$\lim_{r\rightarrow 0} r^n R_n(r) = 0$$

whereas, for a fixed value of r,

$$\lim_{n\rightarrow\infty} \left| r^n R_n(r) \right| \rightarrow \infty$$

Table III

The Coefficients, $G_n(rt)$, in Stumpers' Asymptotic Series

n	$G_n(\phi)$, $\phi \equiv rt$
0	$Z(j\omega_i)$
1	$j\frac{1}{2!}\, s'(\phi)\, Z^{(2)}(j\omega_i)$
2	$j\frac{1}{3!}\, s''(\phi)\, Z^{(3)}(j\omega_i) + \frac{1}{2!}\left(j\,\frac{s'(\phi)}{2!}\right)^2 Z^{(4)}(j\omega_i)$
3	$j\,\frac{s'''(\phi)}{4!}\, Z^{(4)}(j\omega_i) + \left(j\,\frac{s'(\phi)}{2!}\right)\left(j\,\frac{s''(\phi)}{3!}\right) Z^{(5)}(j\omega_i) + \frac{1}{3!}\left(j\,\frac{s'(\phi)}{2!}\right)^3 Z^{(6)}(j\omega_i)$
4	$j\,\frac{s^{iv}(\phi)}{5!}\, Z^{(5)}(j\omega_i) + \left[\left(j\,\frac{s'(\phi)}{2!}\right)\left(j\,\frac{s'''(\phi)}{4!}\right) + \frac{1}{2!}\left(j\,\frac{s''(\phi)}{3!}\right)^2\right] Z^{(6)}(j\omega_i)$
	$+\frac{1}{2!}\left(j\,\frac{s'(\phi)}{2!}\right)^2\left(j\,\frac{s''(\phi)}{3!}\right) Z^{(7)}(j\omega_i) + \frac{1}{4!}\left(j\,\frac{s'(\phi)}{2!}\right)^4 Z^{(8)}(j\omega_i)$
	$G_n(\phi) = \sum$ coefficients of r^n in the van der Pol-Stumpers expansion after substitution of $\theta^{(\nu)}(t) = r^{\nu-1} s^{(\nu-1)}(\phi)$

Such expansions have the peculiar property that their terms diminish in magnitude with increasing n, until a minimum is reached, after which the terms increase with n beyond limits. Thus the rearrangement of the terms in the manner indicated in Eq. (5) seems to have destroyed the convergence of the series of Table I. This signifies that the van der Pol-Stumpers expansion in Table I is not absolutely convergent, although it is uniformly convergent, as is clear from its derivation.

The indicated behavior of the terms in an asymptotic expansion does not impair the usefulness of such expansions. Indeed, asymptotic expansions possess some remarkably useful properties (8). The following facts are of interest. First, the error incurred in approximating the expanded function by the sum of only the first n terms of the series is less than the first rejected term. The best approximation is obtained, therefore, when n is so chosen that the $(n + 1)^{th}$ term is the smallest term in the expansion. Thus, while a convergent series can be used to approximate the expanded function within an arbitrarily small tolerance, an asymptotic series cannot. But it is not uncommon for an asymptotic series to yield a better numerical approximation with a given number of terms than a convergent series.

III. THEORY OF FM-TO-AM CONVERSION

The FM discriminator can be considered, in this study, as a network that is designed to convert the instantaneous-frequency variations of an amplitude-limited wave into instantaneous-amplitude variations. These can subsequently be detected by appropriate AM detection techniques.

The form of the excitation function assumed in Eq. (1) pertains to an amplitude-limited frequency-modulated carrier. The phase-modulating function $\theta(t)$ can be considered as carrying the specifications of the frequency modulation, with the frequency ω_c representing the average value of frequency about which the frequency changes are executed. We shall now use Taylor's formula to determine the condition for distortionless FM-to-AM conversion and to prescribe estimates for the error incurred in approximating the complex amplitude of the steady-state response of a network (to a frequency-modulated excitation) by a finite number of terms in the Carson and Fry expansion.

With reference to Table I, we assume that $\theta(t)$ and its first (n - 1) derivatives are continuous for all t, and that $\theta^{(n)}(t)$ exists for all finite t, and write

$$f(t - \tau) \equiv e^{j\theta(t - \tau)} = \sum_{\nu=0}^{n-1} \frac{1}{\nu!} f^{(\nu)}(t)(-\tau)^{\nu} + R_{nf}$$

where

$$R_{nf} = \frac{1}{n!} f^{(n)}(t - \eta\tau)(-\tau)^{n}, \quad 0 < \eta < 1$$

Evidently, $\left| f^{(n)}(t - \eta\tau) \right| \leq \left| f^{(n)}(t) \right|_{max}$. The auxiliary function $g(t, \tau)$ can therefore be expressed in the form

$$g(t, \tau) = e^{-j\theta(t)} \sum_{\nu=0}^{n-1} \frac{1}{\nu!} f^{(\nu)}(t)(-\tau)^{\nu} + R_{ng} \tag{6}$$

where $\left| R_{ng} \right| = \left| R_{nf} \right|$.

Substitution from Eq. (6) in step 1, Table I, followed by an interchange of summation and integration and the use of the relations defined in step 5, Table I, leads to

$$E(t) = e^{-j\theta(t)} \sum_{\nu=0}^{n-1} \frac{1}{\nu!} f^{(\nu)}(t)\, Z^{(\nu)}(j\omega_c) + R_{nE}(t)$$

whence

$$E(t) = \sum_{\nu=0}^{n-1} \frac{1}{\nu!} B_\nu(t) Z^{(\nu)}(j\omega_c) + R_{nE}(t) \tag{7}$$

where

$$R_{nE}(t) = \frac{1}{n!} e^{-j\theta(t)} \int_0^\infty f^{(n)}(t - \eta\tau)(-\tau)^n h(\tau) e^{-j\omega_c \tau} d\tau$$

and

$$R_{nE}(t) \le \frac{1}{n!} \left| B_n(t) \right|_{max} Z^{(n)}(j\omega_c) \tag{8}$$

The magnitude of the quantity on the right-hand side of this inequality therefore represents an upper bound on the magnitude of the error incurred in using only the first n terms in the Carson and Fry expansion to represent E(t). This error is thus bounded by the magnitude of the first rejected term.

For a network that is intended for FM-to-AM conversion with little distortion, we choose n = 2, and write

$$E(t) = Z(j\omega_c) + jZ'(j\omega_c) \frac{d\theta}{dt} + R_{2E}(t) \tag{9}$$

where

$$\left| R_{2E}(t) \right| \le \frac{1}{2} \left| \sqrt{\left(\frac{d\theta}{dt}\right)^4 + \left(\frac{d^2\theta}{dt^2}\right)^2} \right|_{max} \cdot \left| \frac{d^2 Z(j\omega_c)}{d(j\omega_c)^2} \right|$$

Faithful detection by an ideal envelope detector of unit detection efficiency yields

$$\left| E(t) \right| = \left| Z(j\omega_c) \right| \cdot \left| 1 + j \frac{Z'(j\omega_c)}{Z(j\omega_c)} \cdot \frac{d\theta}{dt} + \frac{R_{2E}(t)}{Z(j\omega_c)} \right|$$

For good sensitivity and low distortion we require that $jZ'(j\omega_c)/Z(j\omega_c)$ be real (as for a pure reactance), that $\left| Z'(j\omega) \right|$ be high, and that

$$\epsilon = \frac{1}{2} \left| \sqrt{\left(\frac{d\theta}{dt}\right)^4 + \left(\frac{d^2\theta}{dt^2}\right)^2} \right|_{max} \cdot \left| \frac{Z''(j\omega_c)}{Z(j\omega_c)} \right| \ll 1 \qquad (10)$$

As an illustration, let

$$\theta(t) = \frac{\Delta\Omega}{\omega_m} \sin \omega_m t = \delta \sin \omega_m t$$

and let the network be a high-Q parallel-resonant circuit whose half-bandwidth between half-power points is given by α rad/sec. If x_c represents the deviation of ω_c from the resonant frequency in units of α, then

$$\epsilon = \begin{cases} \left(\dfrac{\Delta\Omega}{\alpha}\right)^2 \left(1 + x_c^2\right)^{-1} < \left(\dfrac{\Delta\Omega}{\alpha}\right)^2 & \text{for } \delta > 1 \\ \dfrac{\omega_m \Delta\Omega}{\alpha^2} \left(1 + x_c^2\right)^{-1} < \left(\dfrac{\omega_m}{\alpha}\right)\left(\dfrac{\Delta\Omega}{\alpha}\right) & \text{for } \delta < 1 \end{cases}$$

It is interesting to digress a little into the problem of FM disturbances caused by co-channel interference. Thus, let the excitation i(t) be the amplitude-limited resultant of two or more sinusoids of the type that appears at the output of an idealized narrow-band limiter when the i-f amplifier delivers two carriers differing in amplitude and in frequency, the frequency of the stronger signal being $\omega_c = p$ rad/sec. As before, let Z be the system function of the FM-to-AM conversion network (or discriminator). If the envelope $|E(t)|$ of the response of the network is faithfully detected by an ideal, linear, amplitude detector of unit detection efficiency, the first term, Z(jp), in Eq. (9), represents the direct-voltage level dictated by the frequency, p, of the stronger signal when this frequency is considered as essentially constant or, perhaps, as slowly varying. The second term, $jZ'(jp)\, d\theta/dt$, is the first-order disturbance caused by the interference, and so on.

The system function Z must evidently be some varying function of frequency for FM-to-AM conversion to be possible. The distortion caused by the interference (as well as the distortion that a nonlinear variation of $Z(j\omega)$ with $j\omega$ will introduce into the desired direct-voltage level variation) will be minimized if $Z(j\omega)$ is a linear function of $j\omega$ (as for a pure inductance) at least over the expected range of frequency variations. The detected output will then consist only of the slowly varying direct-voltage level, Z(jp), dictated by the frequency p, and the interference term, $jZ'(jp)\, d\theta/dt$, without any distortion of the waveshape of $d\theta/dt$.

Two questions immediately arise: the first concerns the desirability of this faithful translation of the waveshape of $d\theta/dt$ from a frequency variation to an amplitude variation; the second concerns the explicit conditions on $Z(j\omega)$, when both its magnitude and phase vary with frequency, that will ensure a reasonable approximation to this faithful translation. The first question is quickly answered, if we recall that the average value of $d\theta/dt$ over one cycle of the frequency difference between the two carriers delivered by the i-f amplifier is exactly zero, if the resultant of these two carriers is delivered to the discriminator through a limiter of proper bandwidth (1,2). Therefore, if $d\theta/dt$ is faithfully reproduced and envelope-detected, the average value of

$$Z(jp) + jZ'(jp)\, d\theta/dt$$

over one frequency-difference cycle, is exactly $Z(jp)$, the direct-voltage level dictated by the frequency of the stronger signal. To ensure this faithful reproduction of $d\theta/dt$, the discriminator network must be capable of following the excitation of Eq. (1) through quasi-stationary states — that is, at every value of the instantaneous frequency, $p + d\theta/dt$, we should be able to approximate the instantaneous amplitude of the response, on a steady-state basis, by computing the value of $|Z(j\omega)|$ evaluated at $\omega = p + d\theta/dt$, and multiplying it by the (constant) amplitude of the impressed excitation. The conditions that a general system function, Z, must satisfy in order for this quasi-stationary analysis to be applicable will be explored in Section IV, and will be found to agree with those implied in condition (10).

It is reasonable to anticipate that the FM-to-AM conversion characteristic should generally be characterized by second and higher derivatives, whose peak values grow smaller with an increase in the order of the derivative. The amount of permissible deviation from linearity can then be restricted by requiring that

$$\left| Z''(j\omega)/Z'(j\omega) \right|_{max} \ll 1$$

since the negligibility of this ratio would certainly ensure the predominance of the first derivative over the higher derivatives. The implications set by the desirability of a small value of $|Z''/Z'|_{max}$ on the character of Z may be clarified by setting

$$Z(j\omega) = A(\omega)\, e^{j\phi(\omega)}$$

and evaluating the indicated ratio. This leads to

$$\frac{Z''(j\omega)}{Z'(j\omega)} = 2\phi'(\omega) - \frac{\dfrac{A''(\omega)}{A(\omega)} + \phi'^2(\omega) + j\phi''(\omega)}{\phi'(\omega) - j\,\dfrac{A'(\omega)}{A(\omega)}}$$

Evidently, this ratio can be decreased significantly if $A(\omega)$ and $\phi(\omega)$ are assumed to vary linearly with frequency. Thus, if we set

$$A(\omega) = m(\omega - \omega_o) \quad , \quad \text{and } \phi(\omega) = t_d(\omega - \omega_o)$$

we find that

$$\left|\frac{Z''(j\omega)}{Z'(j\omega)}\right|_{max} = 2\, t_d = 2\phi'(\omega)$$

this value occurring at $\omega = \omega_o$. We may, therefore, conclude that the ratio $|Z''/Z'|$, for any discriminator circuit, will be minimized if $A(\omega)$ and $\phi(\omega)$ are linear functions of frequency.

IV. THE CONDITION FOR QUASI-STATIONARY ANALYSIS

The van der Pol-Stumpers expansion states that if $\theta(t)$ and $Z(j\omega)$ combined have properties that make the second and later terms negligible compared with the first, then the steady-state response of the filter will be given, essentially, by

$$e(t) = Z(j\omega_i(t))\, e^{j\int_0^t \omega_i(t)\, dt} \tag{11}$$

This is the same, formally, as the expression for the steady-state response to a constant-frequency excitation. Under these conditions, the filter is said to follow the excitation through quasi-stationary states. This means that, at any instant of time t_1, we may approximate the complex-envelope value of the steady-state response by evaluating the sinusoidal steady-state value of the system function at the value of instantaneous frequency $\omega_i(t_1)$, as though this value were maintained for a sufficiently long time to allow a build-up of the response to the sinusoidal steady-state value. This approach is usually referred to as the "instantaneous-frequency method" of evaluating the steady-state response, and the condition for its validity can be considered as the condition for the applicability of the instantaneous frequency concept as a tool in evaluating the steady-state response of a filter.

The degree to which the filter is able to present a quasi-stationary response is a measure of the faithfulness of the reproduction of the instantaneous-frequency variations of the excitation in the response. When the filter response follows the excitation through quasi-stationary states, we shall show later on that the instantaneous-frequency variations of the response will constitute essentially a delayed replica of those of the excitation if the filter phase characteristic is linear at $\omega = \omega_c$. A filter will be called "wideband" relative to a prescribed frequency-modulated excitation, if the filter's action is sufficiently rapid to allow its response to follow the instantaneous-frequency variations of the excitation through quasi-stationary states. The filter will be called "narrow-band" if it is too sluggish to follow the instantaneous-frequency variations of the excitation. This terminology is inspired by the extent of the significant spectrum of the excitation relative to the usable passband of the filter.

Equation (11) clearly shows that, under conditions of quasi-stationary response, the amplitude characteristic of the filter describes the envelope of the response. If $|Z(j\omega_i)|$ varies with ω_i over the range of instantaneous-frequency variations, then discriminator action, or FM-to-AM conversion, results. For proper, undistorted conversion, the time variations in the resultant envelope should be related linearly to the time variations of the instantaneous frequency. This requirement can, evidently, be met only if $|Z(j\omega_i)|$ is a linear function of ω_i over the entire range of the desired undistorted conversion.

The condition for quasi-stationary analysis will now be adduced by using Taylor's formula to expand the pertinent auxiliary function, $g(t,\tau)$, into the number of terms that is appropriate for the desired analysis. The remainder will again offer a basis for estimating the error incurred in the quasi-stationary approximation.

Thus, assume that $\theta(t)$ and $\theta'(t)$ are continuous for all t, and that $\theta''(t)$ exists for all finite t. Then

$$\theta(t - \tau) = \theta(t) - \tau\theta'(t) + R_{2\theta}$$

where

$$R_{2\theta} = \frac{1}{2}\theta''(t - \eta\tau)(-\tau)^2 \quad , \quad 0 < \eta < 1$$

and

$$\theta''(t - \eta\tau) \leq \left|\theta''(t)\right|_{max}$$

Therefore, with reference to the right-hand column in Table I,

$$g(t,\tau) = e^{jR_{2\theta}} = 1 + R_{1g} \tag{12}$$

where

$$R_{1g} = jR_{2\theta}\, e^{j\mu R_{2\theta}} \quad , \quad 0 < \mu < 1$$

and

$$\left| R_{1g} \right| = \left| R_{2\theta} \right|$$

Substitution from Eq. (12) in step 1, Table I, followed by an interchange of summation and integration and the use of the relations defined in step 5, Table I, leads to

$$E(t) = Z(j\omega_i) + R_{1E}(t) \tag{13}$$

where

$$R_{1E}(t) = \int_0^\infty R_{1g}h(\tau)\, e^{-j\omega_i\tau}\, d\tau$$

and

$$\left| R_{1E}(t) \right| \leq \frac{1}{2} \left| \theta''(t) \right|_{max} \left| Z^{(2)}(j\omega_i) \right| \tag{14}$$

The magnitude of the quantity on the right-hand side of this inequality therefore represents an upper bound on the magnitude of the error incurred in using only the first term in the van der Pol - Stumpers expansion to represent E(t). A similar procedure couldbe used to define an upper bound on the error incurred in using only the first n terms in the van der Pol - Stumpers expansion, but the deduction of the final estimate gets involved. A quick guide to such an estimate is provided by the asymptotic nature of the expansion when it is expressed in the form of Eq. (5) for the case of a periodic frequency modulation.

For a transmission network whose response is intended to carry an adequate reproduction of the angular modulation of the excitation, we can therefore write

$$E(t) = Z(j\omega_i)$$

with the requirement that the maximum relative error, ϵ, in this approximation be negligible compared with unity. From Eqs. (13) and (14), this means that

$$\epsilon_m = \frac{1}{2} \left| \frac{d^2\theta}{dt^2} \right|_{max} \cdot \left| \frac{Z''(j\omega_i)}{Z(j\omega_i)} \right|_{max} \ll 1 \tag{15}$$

It is important to note that the validity of this analysis requires only that $\theta(t)$ and its first derivative be continuous for all t and that $\theta''(t)$ exist for all t. No conditions on the continuity of $\theta''(t)$ and the existence of the higher derivatives of $\theta(t)$ are necessary. However, Z(s) can have only left half-plane poles, as before.

When the modulation is periodic, the asymptotic character of expansion 5 enables us to state that E(t) is closely approximated by $Z(j\omega_i)$, provided that $\theta(t)$ and Z(s) satisfy the indicated restrictions, and

$$\epsilon = \frac{1}{2} \left| \theta''(t) \right| \cdot \left| \frac{Z''(j\omega_i)}{Z(j\omega_i)} \right| \ll 1 \tag{16}$$

at all instants of time. Although this is generally a milder condition than 15, we recall that expansion 5 requires the existence of all derivatives of $\theta(t)$ for all t, while condition 15 is generally true under much milder restrictions.

Clearly, if the maximum possible value of the quantity on the left-hand side in (16) is negligible compared with unity, condition (16) will be satisfied at all times. With a given FM excitation and a given filter, the value of ϵ will vary with the position of the unmodulated carrier frequency, ω_c, relative to the center of the filter passband. It is not difficult to conceive of practical situations in which $\omega_i(t)$ is such that $\left| d/dt\ \omega_i(t) \right|$ assumes its maximum value at precisely those values of t that make $\omega_i(t)$ equal to the frequency at which $\left| Z''(j\omega)/Z(j\omega) \right|$ is maximum (note absence of subscripts i). In applications where this situation is likely to arise, conditions 15 and 16 can be taken in the form

$$\epsilon_{max} = \frac{1}{2} \left| \theta''(t) \right|_{max} \cdot \left| \frac{Z''(j\omega)}{Z(j\omega)} \right|_{max} \ll 1 \tag{17}$$

where ϵ_{max} represents the absolute maximum value of ϵ.

V. APPLICATIONS

The Sluggishness Ratio of a Filter

For a specified phase-modulation function $\theta(t)$, the conditions on a filter characterized by $Z(j\omega)$ that will ensure quasi-stationary response are expressible as restrictions on the ratio $|Z''/Z|$. Evidently, the quantity $|Z''/Z|_{max}$ is a sluggishness parameter that is characteristic of the filter. Its evaluation for some filters of wide practical interest follows.

Consider, first, the high-frequency model of a parallel resonant circuit, shown in Fig. 1. The impedance of this circuit, with assumptions of high Q and small deviations from the center frequency, can be expressed in the form

$$Z(j\Omega) = \frac{R}{1 + j\dfrac{\Omega}{\alpha}}$$

where $\Omega = \omega - \omega_0$ is the deviation of the frequency ω of the excitation from the resonance frequency ω_0, and $\alpha = 1/(2\,RC) = (BW)/2$ is the damping factor of the circuit. It is readily shown that the ratio $|Z''/Z|$ attains its maximum value at the center frequency ω_0, where

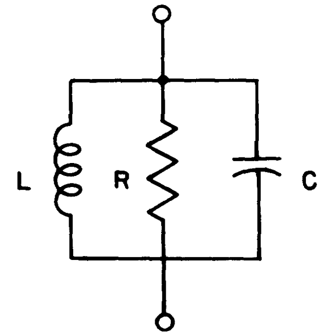

Fig. 1. High-frequency model of a parallel-tuned circuit.

$$\left|\frac{Z''(j\Omega)}{Z(j\Omega)}\right|_{max} = 8/(BW)^2$$

and (BW) is the bandwidth (in rad/sec) between half-power points. The maximum value of the ratio $|Z''/Z|$ for a parallel-resonant circuit is actually attained at a frequency slightly below ω_0, but the discrepancy is small for high Q values.

Consider, next, the general n-pole Butterworth bandpass filter. Such a filter is usually characterized by a pole pattern in which the poles fall on a semicircle whose center lies on the $j\omega$-axis and whose radius equals one-half the over-all bandwidth of the filter between half-power points. The exact positions of the poles of an n^{th}-order Butterworth filter are at the locations of the $2n^{th}$ roots of $(-1)^{n+1}$ that lie in the left half-plane. It can be readily shown that the transfer system function of such a filter can be normalized into the form

$$Z_{bn}(j\Omega) = \frac{1}{F\left(j\,\dfrac{\Omega}{\alpha}\right)}$$

where α is half of the over-all bandwidth, in rad/sec. Here

$$\left|\frac{Z''_{bn}}{Z_{bn}}\right| = \frac{1}{\alpha^2} \cdot \left|\frac{-F(jx)F''(jx) + 2F'^2(jx)}{1 + x^{2n}}\right| \equiv \frac{1}{\alpha^2} f_n(x)$$

whence

$$\left|\frac{Z''_{bn}}{Z_{bn}}\right|_{max} = \frac{1}{\alpha^2} [f_n(x)]_{max} \tag{18}$$

where $x = \Omega/\alpha$. Since $\alpha = \frac{1}{2}(BW)$, expression (18) is of the form

$$\left|\frac{Z''_{bn}}{Z_{bn}}\right|_{max} = k_n/(BW)^2 \tag{19}$$

where k_n is a constant. We propose to call this constant the index of stiffness or sluggishness of the filter. The quantity $|Z''/Z|$ may be called the stiffness or sluggishness ratio of the filter described by $Z(j\omega)$. Plots of the sluggishness ratio for the Butterworth bandpass filters of orders up to the sixth are presented in Fig. 2. In Fig. 3 the corresponding indices of stiffness of these filters are plotted against n, the order of the filter.

The maximum value of the ratio $|Z''/Z|$ can be expressed in the form of Eq. (19) for any filter described by $Z(j\omega)$. A summary of results pertaining to various filters is presented in Table IV. The sluggishness ratios for n cascaded amplifiers that employ synchronously tuned high-Q parallel-resonant circuits, or identical second-order Butterworth filters are plotted in Fig. 4.

From the appearance of the curves in Figs. 2 and 4 it is evident that the form (17) of condition (16) may appear more conservative than it need be for many applications in which $|Z''/Z|$ attains its maximum outside the range of desired instantaneous-frequency values. For such applications, we may prefer to use condition (16) in the form

$$\frac{1}{2}\left|\,|\theta''(t)|\,|Z''/Z|\,\right|_{max} = \epsilon \ll 1 \tag{20}$$

in which case plots similar to those of Figs. 2 and 4 prove quite helpful. However,

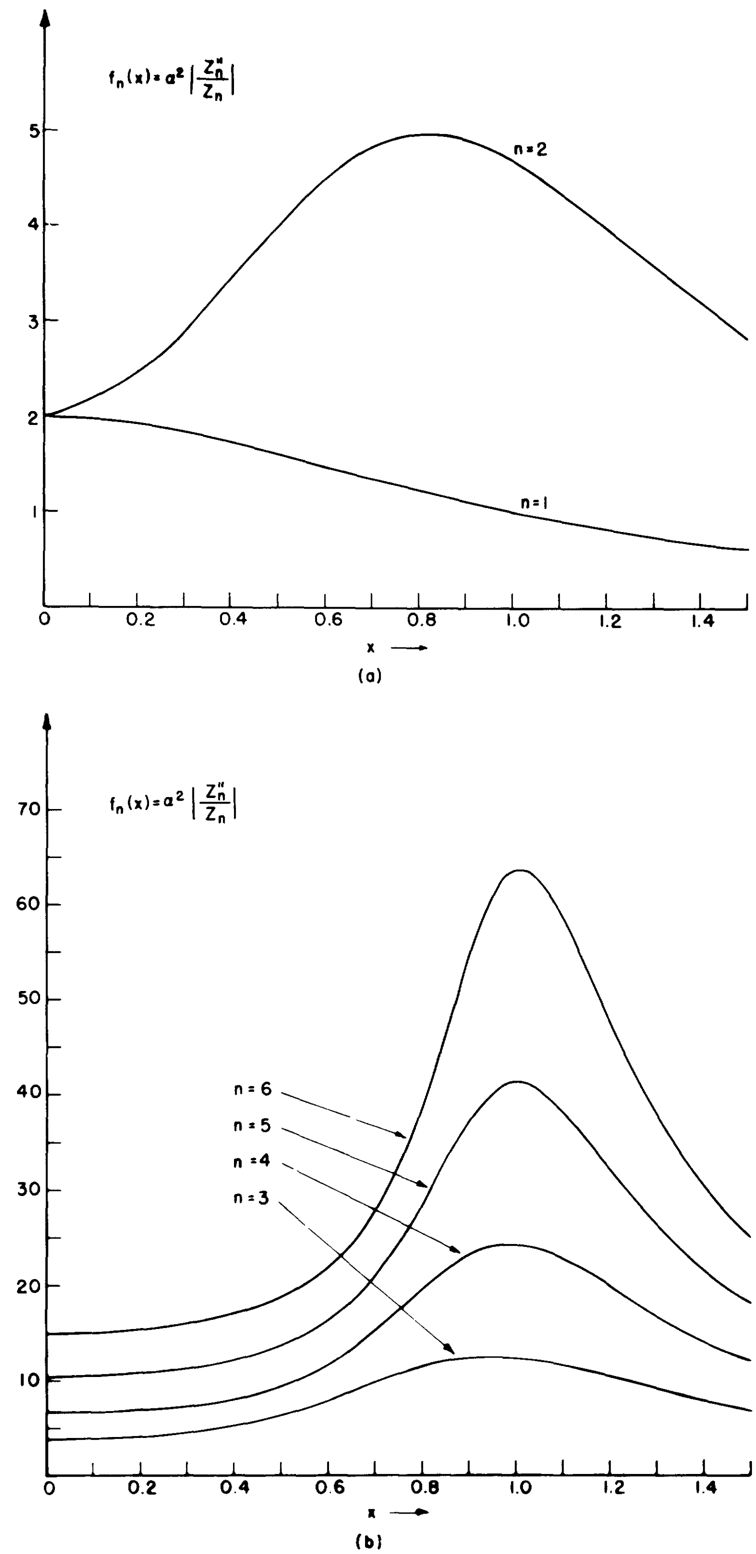

Fig. 2. Normalized sluggishness ratio plots for n^{th}-order Butterworth filters. The sluggishness ratio is $|Z''/Z|$; α is one-half the over-all bandwidth (in rad/sec) between half-power points; x measures the deviation from the center frequency in units of one α.

when the frequency of the excitation can be expected to sweep the frequencies at which $| Z''(j\omega)/Z(j\omega) |$ has its maxima, the use of condition (17) ensures that the error will always be smaller than that which we may be willing to tolerate. The filter will then respond in a quasi-static manner, regardless of which part of the response characteristics is swept by the variable instantaneous frequency of the excitation. With a single-tuned circuit, the validity of the quasi-stationary solution will meet its stiffest test when the instantaneous frequency of the excitation equals the resonant frequency of the filter at the instant in the modulation cycle when $\theta''(t)$ assumes its maximum value. With the higher-order Butterworth filters, the stiffest test is near the edges of the passband.

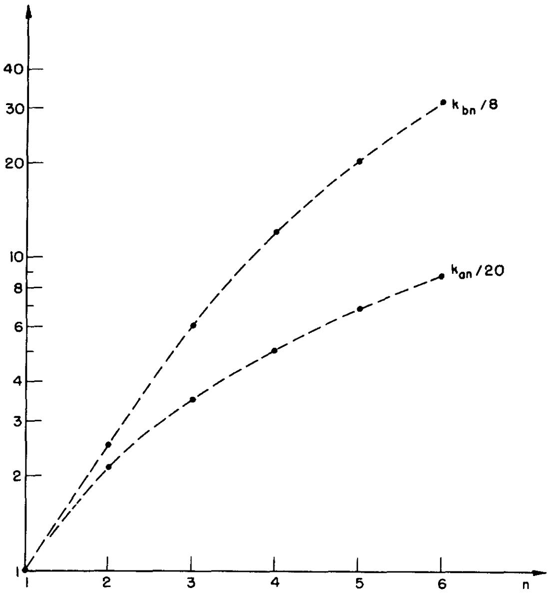

Fig. 3. Variation of the index of stiffness k_n with n: k_{bn} is for n^{th}-order Butterworth filter, k_{an} is for n identical cascaded second-order Butterworth bandpass amplifiers.

Application to Sinusoidal Modulation

We shall next consider some interesting examples of the phase-modulation function $\theta(t)$. The simplest example is

$$\theta(t) = \delta \sin \omega_m t \tag{21}$$

wherein δ may be interpreted as either a phase deviation independent of ω_m or as the ratio of frequency deviation, $\Delta\omega$, to the modulating frequency ω_m. Thus

$$| \theta''(t) |_{max} = \omega^2_m \delta = \omega_m \cdot \Delta\omega$$

Condition (17) therefore reduces to

$$\frac{1}{2} k \cdot \frac{\omega_m}{(BW)} \cdot \frac{\Delta\omega}{(BW)} = \epsilon \ll 1 \tag{22}$$

which specifies the condition that must be satisfied jointly by the over-all filter bandwidth (BW), the index of stiffness k, the modulating frequency ω_m, and the frequency deviation $\Delta\omega$, in order for the filter described by k to follow the variable-frequency

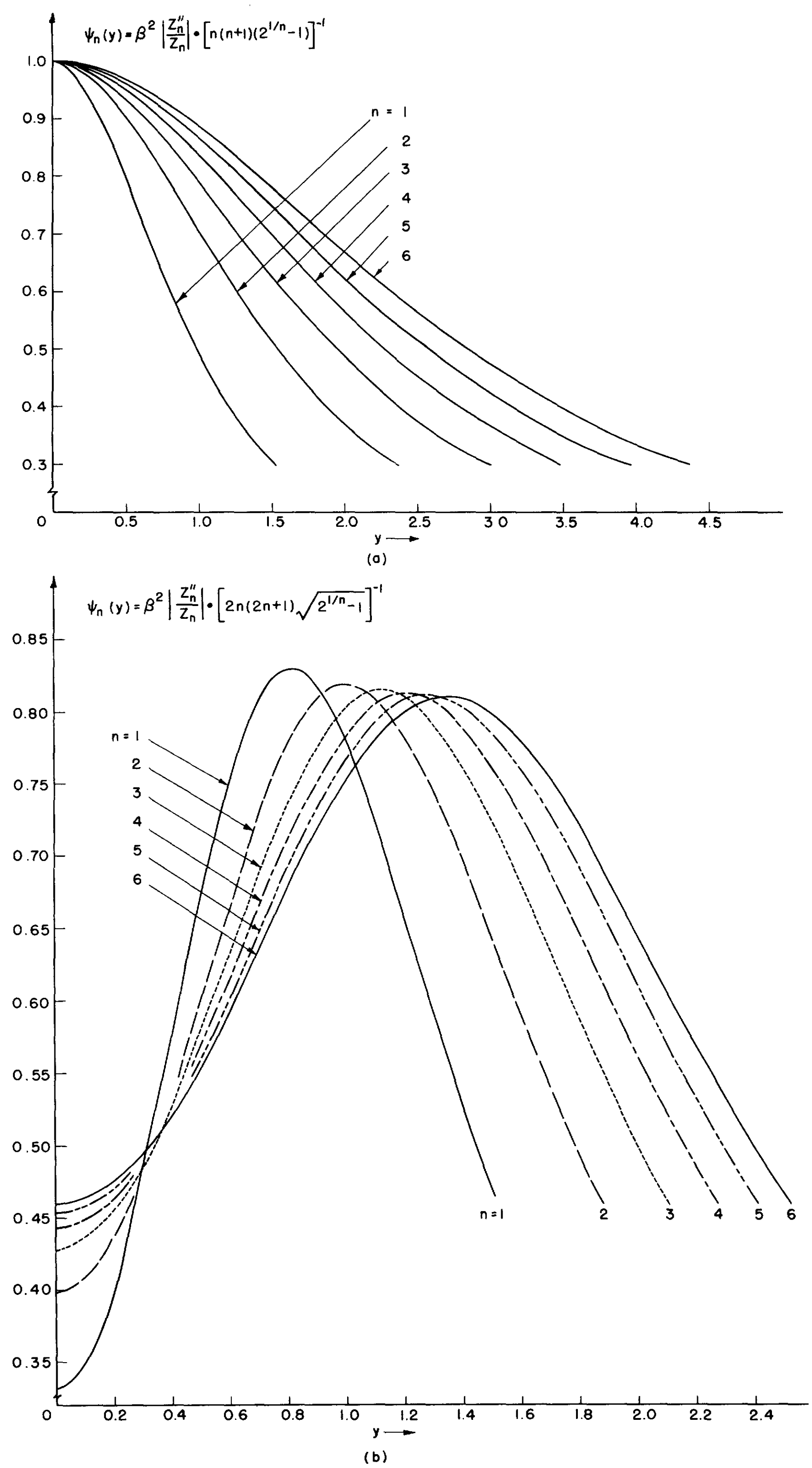

Fig. 4. Normalized sluggishness ratio plots for n identical cascaded tuned amplifiers. The sluggishness ratio is $|Z''/Z|$; β is one-half the over-all bandwidth (in rad/sec) between half-power points; y measures the deviation from the center frequency in units of one β. Each stage has: (a) a high-Q single-tuned circuit; (b) a second-order Butterworth bandpass filter.

Table IV

Indices of Stiffness of Various Filters

A. Butterworth Bandpass Filters

$\Omega \equiv \omega - \omega_0$ = frequency deviation from center frequency of the filter

$\alpha \equiv \frac{1}{2}(BW)$ = one-half over-all bandwidth (in rad/sec) between half-power points

Filter	$Z(j\Omega)$	$(BW)^2 \, \lvert Z''/Z \rvert_{max}$	Value of Ω at which $\lvert Z''/Z \rvert$ has a maximum
High-Q parallel-resonant circuit	$\dfrac{1}{1 + j\dfrac{\Omega}{\alpha}}$	8	$\Omega_m = 0$
Second-order Butterworth (critically coupled, or flat-staggered high-Q pair)	$\dfrac{1}{1 - \left(\dfrac{\Omega}{\alpha}\right)^2 + j\sqrt{2}\left(\dfrac{\Omega}{\alpha}\right)}$	$\dfrac{72}{\sqrt{13}} \approx 20$	$\Omega_m = \pm 0.815\alpha$
Third-order Butterworth	$\dfrac{1}{1 - 2\left(\dfrac{\Omega}{\alpha}\right)^2 + j\left[2\left(\dfrac{\Omega}{\alpha}\right) - \left(\dfrac{\Omega}{\alpha}\right)^3\right]}$	49.3	$\Omega_m = \pm 0.954\alpha$

B. Cascaded Tuned Amplifiers

n identical stages, α = half bandwidth of each stage

Single-stage filter	Over-all $Z(j\Omega)$	Over-all index of stiffness	Ω_m	β = half over-all bandwidth
Single-tuned, high-Q, parallel-resonant circuit	$\dfrac{1}{\left(1 + j\dfrac{\Omega}{\alpha}\right)^n}$	$4n(n+1)(2^{1/n} - 1)$ $\approx 2.78(n+1)$ for $n > 2$	0	$\alpha\sqrt{2^{1/n} - 1}$
Second-order Butterworth	$\dfrac{1}{\left[1 - \left(\dfrac{\Omega}{\alpha}\right)^2 + j\sqrt{2}\left(\dfrac{\Omega}{\alpha}\right)\right]^n}$	For $n > 2$ $k_{an} \approx 5.40\sqrt{n}(2n + 1)$	For $n > 2$ $\Omega_m \approx \pm 0.79\alpha$	$\alpha\sqrt[4]{2^{1/n} - 1}$

excitation through quasi-stationary states.

If the filter is a single-tuned circuit, k = 8 and Eq. 22 becomes

$$\frac{\omega_m}{(BW)/2} \cdot \frac{\Delta\omega}{(BW)/2} \ll 1 \tag{23}$$

which states that the product of the modulation frequency and the maximum frequency deviation, when each is measured in units of one-half the filter bandwidth, must be negligible compared with unity, in order for a computation of the steady-state response of the tuned circuit on the instantaneous-frequency basis to closely approximate the true response.

It is of interest to examine the accuracy of the generally accepted rule of thumb whereby the significant bandwidth occupied by the spectrum of a frequency-modulated sinusoid is estimated as being given by twice the sum of the maximum frequency deviation and the fundamental modulation frequency. It is common practice to use this rule in specifying the bandwidths for use in filters that will intercept this modulated wave, in order to achieve an acceptable compromise between the desire to minimize the distortion of the modulation by the filter and the desire to minimize the bandwidth offered to background noise.

Let the filter used have an index of stiffness k. If, in accordance with the above rule of thumb, we choose $(BW) = 2(\Delta\omega + \omega_m)$, then Eq. 22 gives

$$\epsilon = \frac{k}{8} \cdot \frac{\delta}{(1+\delta)^2} = \frac{k}{8} f(\delta)$$

where δ = modulation index = $\Delta\omega/\omega_m$. For $\delta < 1/10$, $f(\delta) \approx \delta$, and for $\delta > 10$, $f(\delta) \approx 1/\delta$. The quantity ϵ is the upper bound for the magnitude of the fractional error involved in assuming that the filter response will follow the excitation through quasi-stationary states. When $\delta = 1$, ϵ takes on its maximum value $k/32$. For a single-tuned circuit, $k = 8$ and $\epsilon_{max} = 1/4$, which does not represent a negligible error. The maximum value of ϵ for all values of δ can be kept below 1/10 by modifying the rule in the form

$$(BW) = 2[\Delta\omega + \nu\omega_m]$$

where $\nu \geq k/3.2$.

We may conclude that this rule of thumb involves errors that lie within tolerable limits only for very large ($\delta > 5k/4$) or very small ($\delta < 4/5k$) values of the modulation index, δ . In these cases, we can say that the bandwidth is of the order of twice the deviation or twice the modulation frequency, whichever is predominantly larger. Intolerable errors, however, are introduced in the range $5k/4 > \delta > 4/5k$.

In the case of a general phase-modulating function $\theta(t)$ which satisfies the necessary requirements stated in connection with condition (15), and if the filter is intended for low-distortion transmission, we may use the formula

$$(BW) = K\sqrt{|\theta''(t)|_{max}}\ , \quad K = \sqrt{\frac{k}{2\epsilon}},\ \epsilon < 1/10 \tag{24}$$

in choosing the bandwidth of a filter whose index of stiffness is k. The choice must thus be made, generally, on the basis of the maximum rate at which the instantaneous frequency is swept and not on the basis of the maximum amount by which it is deviated. The type of desired filter, and often the portion of the filter characteristic that will be swept, must also be considered.

In the light of these ideas, it is also appropriate to point out that it is not generally true that for high modulation indices (ratios of maximum frequency deviation to modulation frequency) the significant spectrum of an FM wave is contained within a bandwidth of nearly twice the frequency deviation. Although this is true of a few simple modulations (for example, a sinusoid) our next example will illustrate how the practice of estimating filter bandwidths on this basis can lead to serious errors. Condition (17) suggests that a safer basis for classifying frequency-modulations for the purpose of filter-response analyses is provided by $|\theta''(t)|_{max}$ as compared with the speed of response of the system.

Application to the Two-Path Interference Problem

The second $\theta(t)$ time function that we shall consider is of special interest in the multipath interference problem (1,2). Consider the situation in which two unmodulated carriers, differing slightly in frequency and amplitude, fall within the passband of the intermediate-frequency amplifier. Let the two signals be of relative constant amplitudes 1 and a, where $a < 1$, and let their frequencies be p and $p + r$ rad/sec, the former being that of the stronger signal. Consider the resultant signal to be passed through an ideal limiter that is followed by a filter of bandwidth $(BW)_{lim}$ rad/sec, and index of stiffness k. Assume that, as a result of a certain selectivity that is associated with the limiter circuit proper, this limiter delivers only the spectral components centered about the frequency p of the stronger signal, while the harmonics of p, with $r \ll p$, and their associated sidebands are completely rejected or negligible. Thus, with $A(t) \cos [pt + \theta(t)]$ at its input, the limiter will deliver to the filter the excitation

$$i(t) = \cos [pt + \theta(t)]$$

provided that A(t) contributes none of the zero crossings. It is readily shown (1) that here $\theta(t)$ is given by

$$\theta(t) = \tan^{-1} \frac{a \sin rt}{1 + a \cos rt} \tag{25}$$

which is the instantaneous deviation (from the phase of the stronger of the two signals passed by the i-f amplifier) that the resultant signal sustains as a result of the presence of the weaker signal. The problem is to determine the bandwidth condition on a filter of specified index of stiffness, k, under which the response of the filter can be computed on a quasi-static basis within the bounds of a specified small error. The answer to this question has great bearing on the theory of high-level interference suppression in FM receivers (9).

As we have pointed out, the fractional error involved in the assumption that the complex amplitude E(t) of the filter response is given by $Z(j\omega_i)$ is bounded by the value of ϵ, as defined by Eq. (17). The desired bandwidth condition on the filter can, therefore, be expressed in terms of the upper bound, ϵ, on the error incurred, by means of Eq. (24). For the modulation $\theta(t)$, introduced by the interference and defined by Eq. (25), we find that

$$\left|\theta''(t)\right|_{max} = ar^2 (1 - a^2) W^2(a) \tag{26}$$

where

$$W(a) = \frac{\left[(2/a^2) \left\{(1 + a^2)(1 + 34a^2 + a^4)^{1/2} - (1 + 10a^2 + a^4)\right\}\right]^{1/4}}{3(1 + a^2) - (1 + 34a^2 + a^4)^{1/2}}$$

Let us set $(BW) = (BW)_{lim}$ to emphasize that this is the bandwidth of the filter that follows the limiter. If we replace r by its maximum possible value of one i-f bandwidth, $(BW)_{if}$ (assuming this bandwidth to be well defined), we can write

$$(BW)_{lim} = KB_{lim} \frac{1 + a}{1 - a} (BW)_{if} \tag{27}$$

where

$$K = \left(\frac{k}{2\epsilon}\right)^{1/2}$$

and

$$B_{lim} = \frac{[a(1 - a^2)]^{1/2}}{\left(\frac{1 + a}{1 - a}\right)} W(a) \tag{28}$$

The function $B_{lim}(a)$ is plotted in Fig. 5. Except for a scale factor, this function represents the value of $(BW)_{lim}/(BW)_{if}$ normalized with respect to the ratio $(1 + a)/(1 - a)$. The plot of Fig. 5 shows that B_{lim} rises steeply from zero for $a = 0$, to 0.3 for $a = 0.15$, and then tends to rise rather slowly thereafter, reaching a limiting value of 0.403 for $a = 1$.

In terms of $B_{lim}(a)$, Eq. (26) takes the form

$$\left| \theta''(t) \right|_{max} = B_{lim}^2 (a) \left[\frac{1 + a}{1 - a} r \right]^2 \tag{29}$$

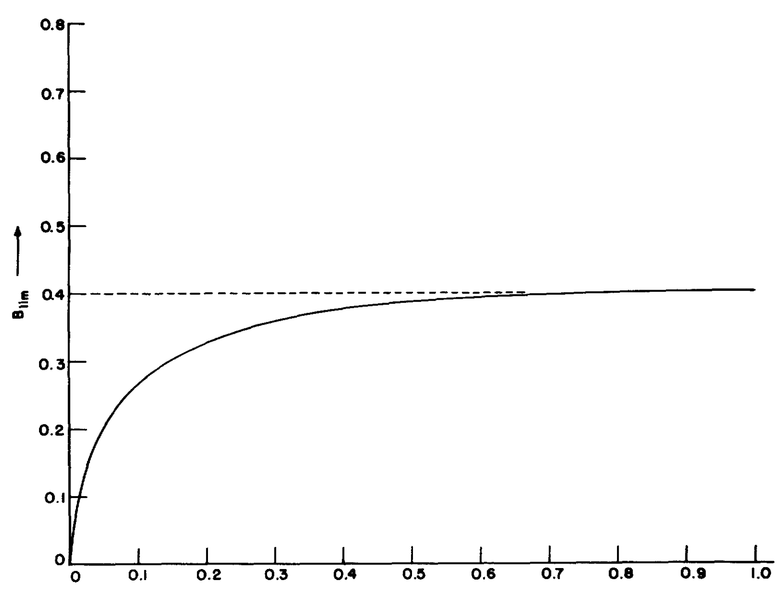

Fig. 5. Plot of the auxiliary function $B_{lim}(a)$ defined in Eq. 28.

The important results of this analysis are given by Eqs. (27) and (29). Equation (29) shows that the FM disturbance, $\theta'(t)$, caused by the interference of the weaker signal, is clearly distinguishable from the usual message modulation on the basis of the maximum slope in its waveform. For if we take the ratio of the absolute maximum slope of the disturbance waveform to the absolute maximum slope of the sinusoidal FM modulation of Eq. (21) with $r = (BW)_{if}$ and $\Delta\omega = 1/2\,(BW)_{if}$, the result is

$$\rho = \left|\theta''_N\right|_{max} \Big/ \left|\theta''_S\right|_{max} = 2B^2_{lim}\left(\frac{1+a}{1-a}\right)^2 \frac{(BW)_{if}}{\omega_m} \tag{30}$$

With a wideband FM system, $\delta = (BW)_{if}/2\omega_m$ is large. For example, if $\omega_m/2\pi = 15$ kc and $(BW)_{if}/2\pi = 150$ kc, then for interference ratios, $\underline{a}$, in the range $0.4 < a < 1$, Eq. (30) becomes

$$\rho = 3.2\left(\frac{1+a}{1-a}\right)^2$$

This means that bandpass filters that are too sluggish to reproduce the sharp disturbance could be introduced in the limiter circuit without distorting the message modulation. In this way, the disturbance can be significantly reduced by a sluggish filter of proper bandwidth (9).

Equation (27) offers an important illustration of the fact that the range of the maximum instantaneous-frequency deviation is not always a reliable basis for estimating the filter bandwidth that is necessary for adequate reproduction of the instantaneous-frequency waveform — even when the modulation index is extremely high. This is readily seen if we note that the maximum range that can be covered by the instantaneous-frequency excursions of the amplitude-limited resultant of the two superimposed sinusoids (if the i-f amplifier bandwidth is sharply defined by one $(BW)_{if}$) is given by (1)

$$(BW) = \frac{1+a}{1-a}\,(BW)_{if} \tag{31}$$

This value of (BW) is only $1/KB_{lim}$ of the value stipulated in Eq. (27). Thus, if the filter is a simple, parallel-resonant circuit, and if the upper bound, ϵ, on the magnitude of the error incurred in retaining only the first term in expansion (5) is so chosen that $\epsilon < 1/10$, then the product $KB_{lim} > 2.5$ for all $a > 0.4$. For a two-pole Butterworth filter, the product $KB_{lim} > 4$ for all $a > 0.4$, if $\epsilon < 1/10$. In the range $a > 0.4$, B_{lim} is approximately constant and equal to 0.4. Consequently, in this range, $KB_{lim} = 1$ for $\epsilon = 16/25$ with a parallel-resonant circuit, and for $\epsilon = 8/5$ with a two-pole Butterworth circuit. These figures indicate that excessive errors will be incurred in assuming that a value of limiter-filter bandwidth given by Eq. (31) is sufficient for the determination of the steady-state filter response on the instantaneous-frequency basis. A filter of this bandwidth is well within the narrow-band classification with respect to an excitation whose phase modulation is defined in Eq. (25).

These conclusions check with the results of a computation based on the Fourier approach that utilizes a filter with idealized amplitude and phase characteristics (1,9). These results reveal that the necessary bandwidth for enclosing the significant spectrum of the amplitude-limited resultant of the two sinusoids passed by the i-f amplifier is of the order of five times the value given by Eq. (31).

As for the values of r for which a filter of bandwidth $(BW)_{lim}$ and index of stiffness k will not distort the FM disturbance noticeably, we may say that the filter is sluggish, or narrow-band, for all values of the frequency difference, r, that satisfy the condition

$$r > r_{min} = \frac{1}{KB_{lim}} \cdot \frac{1-a}{1+a} (BW)_{lim} \tag{32}$$

The preceding results are of great importance in the theory of certain methods for improving the capture performance of FM receivers (9,10).

An interesting significance may be attached to the product KB_{lim}. The instantaneous frequency of the amplitude-limited resultant of the two assumed sinusoids may deviate from the center frequency of the limiter filter by a maximum amount given by

$$\Omega_{max} = \frac{1}{2} \cdot \frac{1+a}{1-a} (BW)_{if}$$

assuming good alignment and a sharply defined i-f passband. In units of $1/2\ (BW)_{lim}$, this absolute maximum frequency deviation from the center of the limiter-filter passband is given by

$$x_{max} = \frac{\Omega_{max}}{(BW)_{lim}/2} = \frac{1}{KB_{lim}}$$

For a simple, parallel-resonant circuit, the requirement that KB_{lim} be greater than 5/2 for $\epsilon < 1/10$ implies that the resultant amplitude-limited signal impressed at the input of the filter should be accommodated over the whole extent of its instantaneous-frequency excursions within an essentially flat portion of the amplitude characteristic, and within an essentially linear portion of the phase characteristic. This follows from the fact that the stipulation $KB_{lim} > 5/2$ implies the restriction $x_{max} < 2/5$. At this value of x, the amplitude characteristic will be less than 8 per cent below its center-frequency value, and the phase characteristic will deviate from linearity by less than approximately

5 per cent. Of course, the filter will then be capable of following the excitation through quasi-stationary states, whether the instantaneous-frequency variations sweep the peak or the sides of the amplitude characteristic. With a simple resonant circuit, the validity of the quasi-stationary solution will meet its stiffest test when the instantaneous frequency of the excitation equals the resonant frequency of the filter at the instant in the modulation cycle when $\theta''(t)$ assumes its maximum value.

With a two-pole Butterworth filter, if the stipulation $KB_{lim} > 4$ is imposed to ensure an acceptable approximation to a quasi-static response, then, at $x = x_{max} = 1/4$, the amplitude characteristic is less than 0.2 per cent below its center-frequency value, and the phase characteristic deviates from linearity by less than 2 per cent. Again, the filter will respond in a quasi-static manner, regardless of which part of the response characteristics is swept by the variable instantaneous frequency of the excitation. The validity of the quasi-stationary solution meets its stiffest test when the instantaneous-frequency deviation from the center frequency of the filter is $\pm 0.815(BW)/2$ at the instant when $\theta''(t)$ achieves its maximum value.

FM Video Applications

We shall consider three additional forms for the phase-modulating function $\theta(t)$, the choice of which will be based upon the properties of the function

$$f(\phi) = \frac{\sin A\phi}{\sin \phi} \tag{33}$$

where A is an integer. For infinitesimally small ϕ, the numerator and the denominator are closely approximated by their arguments. Therefore, $f(0) = A$. A similar argument in terms of $\phi = \pi - \delta$, where δ is an infinitesimal quantity, shows that $f(\pi) = - A \cos A\pi$. Therefore, if A is even, $f(\pi) = - A$, and if it is odd, $f(\pi) = A$. A sketch of $f(\phi)$ for A even is shown in Fig. 6. The humps at $\phi = 0$ and at $\phi = \pi$ are A integral units high, and $2\pi/A$ radians wide at the base.

We immediately recognize that $f(\phi)$ with A even is quite useful because its integral, as illustrated in Fig. 6b, is a reasonable analytical approximate to a square wave, and $f(\phi)$ approximates a series of recurrent pulses of alternating polarities. The parameter A controls the height of each pulse directly, and the width inversely. For the integral of $f(\phi)$, the parameter A is the maximum slope, and is, therefore, inversely proportional to the rise time of the step.

The properties of $f(\phi)$ for A odd are very much like those for A even, with obvious differences. For A odd, the integral of $f(\phi)$ is a rising staircase, and $f(\phi)$ itself is a series of recurrent pulses of the same polarity.

Now, let us set $\phi = rt$, and consider the case in which

$$\theta(t) = \int_0^{\phi} \frac{\sin A\phi}{\sin \phi}\, d\phi \tag{34}$$

where A is an arbitrary integer. This case represents either square-wave or staircase phase modulation. The instantaneous-frequency modulating function is a train of recurrent pulse-like humps of the same or alternately different polarity. The pulse trains resemble a series of short pulses that have been distorted by the action of a low-pass filter. For this modulating function, condition (17) takes the form

$$k \left[\frac{Ar/2}{(BW)}\right]^2 \ll 1$$

If we note that Ar is the instantaneous-frequency pulse amplitude, we can interpret this condition as stating that the product of k times the square of one-half the pulse height, measured in units of one filter bandwidth (between the half-power points), must be negligible compared with unity in order for the filter to follow the excitation through quasi-stationary states.

As a second example, let

$$\theta(t) = \frac{\sin A\phi}{\sin \phi} \tag{35}$$

where, again, $\phi = rt$. As before, a factor of unity has been arbitrarily chosen as a multiplier, although for purposes of proper scaling any other constant multiplier may be chosen. This form of the phase-modulating function, with A an integer that may be even or odd, represents a type of pulse phase modulation. The maximum phase deviation is given by A, within a normalization factor implied in Eq. (35); the maximum frequency deviation is given by $\Omega_{max} = 0.436A^2r$, for large A. The condition for quasi-stationary response is

$$\frac{k}{6A}\left[\frac{\Omega_{max}}{(BW)/2}\right]^2 \ll 1$$

In terms of the phase-modulation pulse amplitude A, and the pulse repetition frequency $r/2\pi$, the condition is

$$\frac{k}{6}A^3\left[\frac{r}{(BW)}\right]^2 \ll 1$$

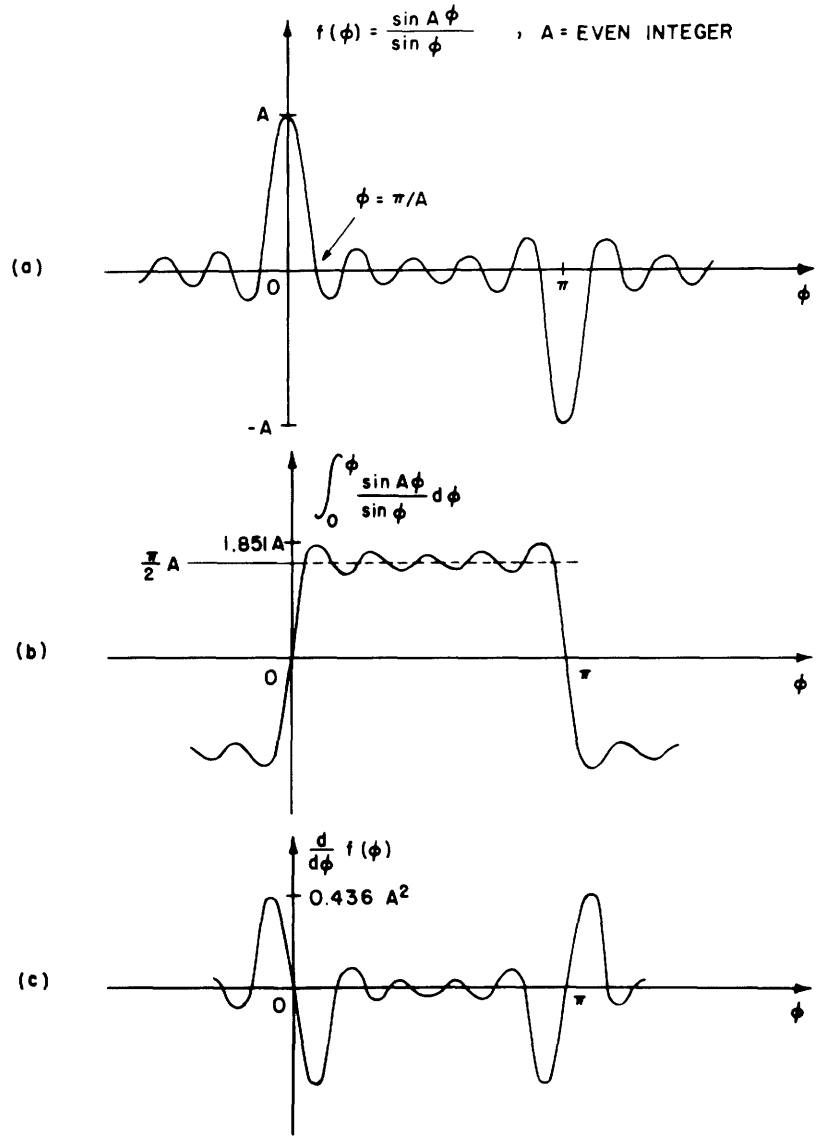

Fig. 6. Properties of the function $f(\phi)$ defined in Eq. 33.

Finally, let

$$\frac{d\theta}{dt} = \frac{2\Delta\Omega}{\pi A} \int_0^{\phi} \frac{\sin A\phi}{\sin \phi}\, d\phi \ , \qquad \phi = rt \tag{36}$$

in which A will be restricted to even integer values. From Fig. 6b, it is clear that Eq. (36) is a convenient approximation to a square-wave frequency modulation. The instantaneous frequency shifts from approximately $\mp\ \Delta\Omega$ to approximately $\pm\ \Delta\Omega$ in an interval of approximately $2\pi/Ar$ sec. The steepest slope of each edge is given by

$$\left| \theta''(t) \right|_{max} = \frac{2}{\pi} \Delta\Omega r$$

and condition (10) becomes

$$\frac{k}{2\pi} \cdot \frac{2\Delta\Omega}{(BW)} \cdot \frac{r}{(BW)} \ll 1$$

In general, if a specific test video waveform (which need not be periodic) is prescribed, it is a simple matter to spot and evaluate the steepest slope that must be used in condition (17) in order to determine the conditions under which a given transmission filter or a discriminator will follow the corresponding frequency modulations and reproduce them within prescribed tolerances. In the examples given, this slope depends upon the fundamental repetition frequency r, which accounts for the appearance of r in the prescribed conditions. Such a dependence, however, will not necessarily always arise.

VI. INSTANTANEOUS FREQUENCY OF FILTER RESPONSE

The substitution of

$$Z(j\omega_i) = A(\omega_i)e^{j\phi(\omega_i)} \tag{37}$$

in Eq. (5), enables us to express E(rt) in the form

$$E(rt) = \left[E_r + j\ E_i\right]e^{j\phi(\omega_i)} \tag{38}$$

where E_r and E_i are infinite sums of real terms. If ω_{io} denotes the instantaneous frequency of the response, it is readily shown that

$$\omega_{io} = \omega_i + \phi'(\omega_i)\frac{d\omega_i}{dt} + \frac{E_r\dfrac{dE_i}{dt} - E_i\dfrac{dE_r}{dt}}{E^2_r + E_i^2} \tag{39}$$

To simplify the mathematical reasoning, we assume that E(rt) in Eq. (5) is adequately approximated by some finite number of terms, the first rejected term being an upper bound on the error involved in this approximation. E_r and E_i then become approximated by finite sums. The first few terms in the expansion for ω_{io} are given by

$$\begin{aligned}\omega_{io} = \omega_i &+ \phi'(\omega_i)\theta''(t) \\ &+ \frac{1}{2}\left(\phi'^2 - \frac{A''}{A}\right)\Bigg[\theta'''(t) - \frac{1}{2}\theta''^3\Bigg(\phi''' \\ &+ 2\,\frac{AA'' - A'^2}{A^2}\,\phi' + 2\frac{A'}{A}\phi''\Bigg) - \frac{1}{2}\theta''\,\theta'''\left(\phi'' + 2\frac{A'}{A}\,\phi'\right)\Bigg] \\ &+ \frac{1}{2}\left(\frac{A'A'' - A'''A}{A^2} + 2\phi'\,\phi''\right)\theta''^2(t) \\ &+ \dots\end{aligned} \tag{40}$$

A similar expansion has been presented by Stumpers (5). The terms following ω_i in expansion (40) may be properly classified as distortion terms introduced by the filter. It is of interest to observe that the first-order distortion term is given by the product of the slope of the filter phase characteristic and the instantaneous rate at which the instantaneous frequency is varied. It is also noteworthy that the effect of nonlinearities in the phase characteristic appears explicitly first in the third term, whereas the effect of nonlinearities in the amplitude characteristic presents itself at the second term. The first two distortion terms have also been discussed by van der Pol (4).

Although the use of Eq. (5) in deriving Eq. (40) implicitly assumes that $\theta(t)$ is periodic, Eq. (40) holds also for a more general type of angular modulation, provided that $\theta(t)$ can be expanded in a power series. The use of Eq. (5) offers a valuable guide in error estimation because the error is much more difficult to estimate from the van der Pol - Stumpers expansion when more than the first two terms are used.

Analysis of Residual Distortion

When the filter response is closely approximated by Eq. (11), the instantaneous frequency of the response is approximately

$$\omega_{io}(t) = \omega_c + \frac{d\theta}{dt} + \phi'(\omega_i)\frac{d^2\theta}{dt^2} \tag{41}$$

where, if desired, $\phi'(\omega_i)$ may be called the "instantaneous time delay." It is clear from this expression, that, even under conditions of quasi-static response, the filter will introduce some distortion into the instantaneous-frequency waveform. This distortion is a function only of the slope of the phase characteristic of the filter and the rate at which the instantaneous frequency of the excitation is varied. The maximum value of this first-order distortion term is readily expressed in terms of the tolerance ϵ and the characteristics of either the filter or the modulating wave. This maximum value is given by

$$D_{max} = \left|\phi'(\omega)\right|_{max} \cdot \left|\theta''(t)\right|_{max}$$

The quantity $\left|\phi'(\omega)\right|_{max}$ can be expressed in the form $k_d/(BW)$, where k_d is a constant, peculiar to the filter which may be called the "delay index." If we replace $\left|\theta''(t)\right|_{max}$ by its value from Eq. (24), the maximum value of the first-order instantaneous-frequency distortion term becomes

$$D_{max} = 2\left(\frac{k_d}{k}\right)\epsilon\,(BW)$$

For a single-tuned circuit, $k_d = 2$, $k = 8$, and the maximum value of the first-order distortion term is ϵ times the half-bandwidth between half-power points.

This distortion, however, is a combination of time delay in the transmission plus the more objectionable waveform alteration. If the phase-shift characteristic is linear (and has a small slope) in the vicinity of the point that corresponds to the unmodulated carrier frequency ω_c, the residual distortion under quasi-stationary response conditions amounts, essentially, to a time delay of the frequency modulation given by $\phi'(\omega_c)$, without change in the waveshape of this modulation. To show this, Eq. (41) can be manipulated into the form

$$\omega_{io}(t) = \omega_c + \theta'(t) + \phi'(\omega_c)\,\theta''(t) + \theta''(t) \sum_{\nu=1}^{\infty} \frac{1}{\nu!} \phi^{(\nu+1)}(\omega_c)[\theta'(t)]^{\nu}$$

$$= \omega_c + \theta'\big(t + \phi'(\omega_c)\big) + \theta''(t)[\phi''(\omega_c)\theta'(t) + 1/2\, \phi'''(\omega_c)\theta'^2(t) + \ldots]$$

$$- [1/2\, \theta'''(t)\phi'^2(\omega_c) + 1/3!\, \theta^{iv}(t)\, \phi'^3(\omega_c) + \ldots] \tag{42}$$

from which it follows that

$$\omega_{io}(t) \cong \omega_c + \phi'\big(t + \phi'(\omega_c)\big)$$

if $\phi(\omega)$ is linear and $\phi'(\omega_c)$ is small (see 44 below). This shows that although the effect of a nonlinearity in $\phi(\omega)$ does not appear explicitly in the first distortion term (the quasi-stationary term) in Eqs. (40) and (41), this first distortion term does carry implicitly a waveshape alteration (in addition to the delay) when $\phi(\omega)$ is not linear.

The constituents of the distortion term in the quasi-stationary solution are explicitly separated in Eq. (42) into a part that isolates the unavoidable delay, a part that brings out the effect of nonlinearities in $\phi(\omega)$, and a part that brings out the effect of the value of the slope of $\phi(\omega)$. The importance of the parts that represent waveshape alterations is best brought out by writing

$$\omega_{io}(t) = \omega_c + \theta'(t) + \phi'(\omega_c + d\theta/dt)\,\theta''(t)$$

$$= \omega_c + \theta'\big(t + \phi'(\omega_c)\big) - \sum_{q=2}^{m-1} \frac{1}{q!}\, \theta^{(q+1)}(t)[\phi'(\omega_c)]^q$$

$$- R_{m\theta'} + \theta''(t) \sum_{\nu=1}^{n-1} \frac{1}{\nu!}\, \phi^{(\nu+1)}(\omega_c)(\theta')^{\nu} + \theta''(t) R_{n\phi'} \tag{43}$$

where

$$\left| R_{m\theta'} \right| \le \left| \frac{1}{m!}\, \theta^{(m+1)}(t)\phi'^{\,m}(\omega_c) \right|$$

and

$$\left| R_{n\phi'} \right| \leq \left| \frac{1}{n!} \phi^{(n+1)}(\omega_c)\ \theta'^{n}(t) \right|$$

The error in neglecting the distortion due to $\phi'(\omega_c)$ is

$$\mathcal{E}_{\phi'} \leq \left| R_{2\theta'} \right| \leq \frac{1}{2} \left| \theta'''(t) \cdot \phi'^{2}(\omega_c) \right| \tag{44}$$

For a single-tuned circuit and $\theta'(t) = \Delta\Omega \sin \omega_m t$ the relative error is

$$\frac{\mathcal{E}_{\phi'}}{\Delta\Omega} \leq \frac{1}{2} \left(\frac{\omega_m}{\alpha} \right)^2 = \frac{1}{2} \left(\frac{\omega_m}{\Delta\Omega} \right) \epsilon \ll \epsilon \text{ for } \frac{\Delta\Omega}{\omega_m} > 5$$

It must be pointed out, however, that the appearance of waveform distortion terms that are caused by the presence of a nonzero time delay (or slope of the phase characteristic) is strictly a mathematical peculiarity that results from approximating the complex envelope, E(t), of the total dynamic response by the quasi-stationary term, $Z(j\omega_i)$. From a study of a more general expansion for ω_{io} (t) we have detected the appearance of a series of terms in powers of $\phi'(\omega_i)$ that represent a Taylor's series expansion of θ' (t + ϕ' The quasi-stationary solution retains only the first two terms in this expansion, as is evident from Eq. (41). The third term is also present in Eq. (40). If $\phi'(\omega_i)$ changes with ω_i, then $\theta'(t + \phi'(\omega_i))$ will carry waveform distortion as well as time delay. If $\phi(\omega_i)$ is a linear function of ω_i over the range of frequencies covered by ω_i, $\phi'(\omega_i)$ equals a constant which we can write as $\phi'(\omega_c)$. The distortion terms in powers of $\phi'(\omega_c)$ in Eqs. (42) and (43) will thus be cancelled out by terms that the total dynamic solution would include.

The error incurred in neglecting the distortion caused by nonlinearity of $\phi(\omega)$ is

$$\mathcal{E}_{\text{nonl.}\,\phi} \leq \frac{1}{2} \left| \theta''(t) \right| \left| \theta'(t) \right|^{2} \left| \phi'''(\omega_c) \right| \tag{45}$$

if ω_c equals the center frequency of the filter. For a single-tuned circuit and $\theta'(t) = \Delta\Omega \sin \omega_m t$,

$$\frac{\mathcal{E}_{\text{nonl.}\,\phi}}{\Delta\Omega} \leq \left(\frac{\Delta\Omega}{\alpha} \right)^2 \left(\frac{\omega_m}{\alpha} \right) = \frac{\Delta\Omega}{\alpha} \epsilon < \epsilon$$

Harmonic and Intermodulation Distortion

We shall next compute the harmonic and intermodulation distortion components that will appear in the quasi-stationary response of the filter when the filter phase characteristic is not a linear function of frequency in the neighborhood of the unmodulated carrier frequency. For this purpose, we note, with reference to Table V, that the simplest functional form for $\phi(\omega)$ that is of practical importance is

$$\phi(\omega) = \tan^{-1} y \qquad (46)$$

where

$$y = bx = b\,\frac{\omega - \omega_o}{\alpha}$$

and

$$\begin{aligned} b &= 1 && \text{for single-tuned circuit} \\ &\cong \sqrt{2} && \text{for second-order Butterworth} \\ &\cong 2 && \text{for third-order Butterworth} \end{aligned}$$

Table V

Expressions for $\phi(\omega)$ for Use in Computation of Residual Distortion

FILTER	PHASE CHARACTERISTIC, $\phi(\omega)$ $x \equiv \Omega/\alpha$, $\Omega = \omega - \omega_o$
High-Q parallel-resonant circuit	$-\tan^{-1} x$
Second-order Butterworth*	$-\tan^{-1} \dfrac{\sqrt{2}\,x}{1 - x^2} = -\tan^{-1} \sqrt{2}\,x$ if $\lvert x \rvert < 1/3$
Third-order Butterworth*	$-\tan^{-1} \dfrac{2x - x^3}{1 - 2x^2} = -\tan^{-1}\ 2x$ if $\lvert x \rvert < 1/4$
*An additive constant phase shift is neglected in the expression for $\phi(\omega)$.	

The indicated approximations are justified by the fact that the results of the computation will be of practical value only when the interest centers on designing the filters so as to insure a minimum of distortion in the reproduction of the frequency modulation of the excitation. Consequently, it may be assumed that condition (15) or (17) is satisfied, and that the implications of this condition could be incorporated into the computation by way of appropriate simplifying approximations.

With a filter phase characteristic that is essentially linear and with an excitation whose instantaneous frequency is given by

$$\omega_i(t) = \omega_c + \Delta\Omega \sin \omega_m t \tag{47}$$

the distortion term in the quasi-stationary response is (see Eq. 42)

$$\begin{aligned} D(t) &= \phi'(\omega_i) \frac{d\omega_i}{dt} \\ &= \phi'(\omega_c) \frac{d\omega_i}{dt} \\ &= \omega_m \Delta\Omega \phi'(\omega_c) \cos \omega_m t \end{aligned}$$

This distortion term will be present whether or not $\phi(\omega)$ is linear. As we have already noted, it has the significance of a delay of the frequency modulation function when $\mathcal{E}_{\phi'}$ in Eq. (44) is negligible.

Now let $\phi(\omega)$ be given by Eq. (46). With the modulation indicated in Eq. (47), we have

$$y/b = \frac{\omega_i - \omega_o}{\alpha} = \frac{\omega_c - \omega_o}{\alpha} + \frac{\Delta\Omega}{\alpha} \sin \omega_m t$$

and the distortion term is given by

$$\begin{aligned} D(t) &= \frac{1}{1 + y^2} \frac{dy}{dt} \\ &= b\omega_m \cdot \frac{\Delta\Omega}{\alpha} \frac{\cos \omega_m t}{1 + b^2 \left(x_c + \frac{\Delta\Omega}{\alpha} \sin \omega_m t \right)^2} \end{aligned} \tag{48}$$

where $x_c = (\omega_c - \omega_o)/\alpha$. If the signal is not well centered within the filter passband (i.e. if $x_c \neq 0$), then it is evident from Eq. (48) that both even and odd harmonics of the modulation frequency, ω_m, will be present in the instantaneous frequency of the response. A Fourier analysis that shows the effect of detuning upon the distortion has been carried out by this author and by others. Our analysis has been guided by condition (15), and its limitations are therefore more clearly brought out. The results of this analysis will be presented in a later report along with detailed experimental findings. In this report we shall consider only the situation in which $x_c = 0$, i.e. the "well-tuned case."

When $x_c = 0$, the analysis is greatly simplified by rewriting Eq. (48) in the form

$$\frac{\delta}{2\gamma\omega_m} D(t) = \frac{\cos \omega_m t}{1 - \gamma \cos 2\omega_m t} \tag{49}$$

where

$$\gamma \equiv \frac{\frac{1}{2}\delta^2}{1 + \frac{1}{2}\delta^2}, \quad \text{and } \delta \equiv b\left(\frac{\Delta\Omega}{\alpha}\right)$$

Since γ is always less than unity, Eq. (49) can be rewritten in the form

$$\frac{\delta}{2\gamma\omega_m} D(t) = \cos \omega_m t \sum_{n=0}^{\infty} \gamma^n \cos^n 2\omega_m t$$

$$= \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{\gamma}{2}\right)^n \sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!}$$

$$\cdot \left\{\cos\,[2(n-2k)+1]\,\omega_m t + \cos\,[2(n-2k)-1]\,\omega_m t\right\} \tag{50}$$

where we have used the relation

$$\cos^n \phi = 2^{-n} \sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!} \cos (n-2k)\,\phi$$

which holds for any positive integral n.

Closed-form expressions for the harmonic-component amplitudes may be derived as follows. First, we express the right-hand side of Eq. (49) in the form

$$\frac{\cos \omega_m t}{1 - \gamma \cos 2\omega_m t} = \sum_{n=1}^{\infty} \alpha_n \cos n\omega_m t \tag{51}$$

where

$$\alpha_n = \frac{1}{\pi} \int_0^{2\pi} \frac{\cos \phi \cos n\phi}{1 - \gamma \cos 2\phi} \, d\phi$$

If we let $z = e^{j\phi}$, then this integral can be rewritten in the form

$$\gamma \alpha_n = \frac{1}{2\pi j} \oint_{\substack{\text{unit} \\ \text{circle}}} \frac{(z^2 + 1)(z^n + 1/z^n)}{-z^4 + \frac{2}{\gamma} z^2 - 1} \, dz \tag{52}$$

In addition to the n^{th}-order pole at $z = 0$, the integrand has poles at

$$z^2{}_1 = \frac{1}{\gamma}\left[1 - \sqrt{1 - \gamma^2}\right] < 1$$

and

$$z^2{}_2 = \frac{1}{\gamma}\left[1 + \sqrt{1 - \gamma^2}\right] > 1$$

The poles at $z^2{}_2$ lie outside the unit circle and are, therefore, of no interest. But the two simple poles that lie (on the real axis) at

$$z = z_{r,\ell} = \pm \left[\frac{1}{\gamma}\left(1 - \sqrt{1 - \gamma^2}\right)\right]^{1/2} \tag{53}$$

are enclosed within the unit circle. The residues in these poles are given by

$$R_r(n) = \frac{z^2{}_r + 1}{\left(z_2{}^2 - z_r{}^2\right) 2z_r} \left(z_r{}^n + \frac{1}{z_r{}^n}\right) \tag{54}$$

and

$$R_{\ell}(n) = (-1)^{n+1} R_r(n) \tag{55}$$

The residue in the n^{th}-order pole at $z = 0$ is given by

$$R_o(n) = \frac{1}{(n-1)!}\left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{z^2+1}{-z^4+\frac{2}{\gamma}z^2-1}\right)\right]_{z=0} \tag{56}$$

and is best evaluated by noting that it is the coefficient of z^{n-1} in the power series expansion of

$$\psi(z) = \frac{z^2+1}{-z^4+\frac{2}{\gamma}z^2-1}$$

This power series expansion is most expediently obtained by rearranging the numerator and denominator polynomials in ascending powers of z and carrying out a process of long division. Note that only powers of z^2 appear in the expression for $\psi(z)$. Therefore, $R_o(n) = 0$ for all <u>even</u> integral values of n. If the indicated procedure for obtaining $R_o(n)$ is carried out, it becomes quickly obvious that (for the <u>odd</u> integral values of n)

$$R_o(n) = \frac{2}{\gamma} R_o(n-2) - R_o(n-4) \tag{57}$$

The first few values of $R_o(n)$ are given by

$$R_o(1) = -1$$

$$R_o(3) = -\left(\frac{2}{\gamma}+1\right)$$

$$R_o(5) = -\left[\frac{2}{\gamma}\left(\frac{2}{\gamma}+1\right)-1\right]$$

$$R_o(7) = -\left\{\frac{2}{\gamma}\left[\frac{2}{\gamma}\left(\frac{2}{\gamma} + 1\right) - 1\right] - \left(\frac{2}{\gamma} + 1\right)\right\}$$

We now have for the amplitude of the n^{th} harmonic component in D(t), of Eq. (49),

$$A(n\omega_m) = \begin{cases} \dfrac{2\omega_m}{\delta}\left[2R_r(n) + R_o(n)\right] , & \text{for } n \text{ odd} \\ 0 , & \text{for } n \text{ even} \end{cases} \tag{58}$$

A simplification of the expression for $A(n\omega_m)$ will result if it is noted that $\Delta\Omega/\alpha$ will usually be kept less than unity under the condition for quasi-stationary response, and γ will thus be $< 1/3$ for a single-tuned circuit. The reader is, however, cautioned against neglecting too many terms at the start in the expression for $2R_r(n)$, because cancellations of the larger terms by terms in $R_o(n)$ will leave only second and higher-order terms in the final expressions. Further discussion of these results, accompanied with correlation with experimental data, will be presented in a later report.

We now turn to the evaluation of some intermodulation-distortion products. For this purpose, we assume that

$$\omega_i(t) = \omega_c + \Delta\Omega_1 \cos \omega_1 t + \Delta\Omega_2 \cos \omega_2 t$$

and that $\phi(\omega_i)$ is an inverse-tangent function (as in Eq. 46). In many important applications, satisfactory results are readily obtainable by expanding $\tan^{-1}y$ in powers of y, manipulating the trigonometric terms into the desired form, and then differentiating the result. A summary of the results of a computation in which the expansion was carried out to the term in y^5 is presented in Table VI.

Before we conclude this discussion, it is of interest to note that the amplitudes of the harmonic and intermodulation distortion components that will be present in the instantaneous frequency of the response of a cascade of N identical filters will be N times the corresponding amplitudes in the response of the first filter in the cascade.

Table VI

FM Intermodulation Distortion Components Under Conditions of Quasi-Stationary Response

$$\omega_i(t) = \omega_c + \Delta\Omega_1 \cos \omega_1 t + \Delta\Omega_2 \cos \omega_2 t$$

$$\phi(\omega_i) = \tan^{-1} y \quad , \quad y = \delta_1 \cos \omega_1 t + \delta_2 \cos \omega_2 t \quad , \quad \delta = b\frac{\Delta\Omega}{\alpha}$$

$$\omega_{io}(t) = \omega_i(t) + \frac{d}{dt}\phi(\omega_i) \quad , \quad \delta_1 + \delta_2 < 1$$

$$\underline{\sin \omega_{1,2}t} \; : \; -\omega_{1,2}\delta_{1,2}\left[1 - \frac{1}{4}\delta^2_{1,2} - \frac{1}{2}\delta^2_{2,1} - \frac{1}{8}\delta^4_{1,2} - \frac{3}{8}\delta^4_{2,1} - \frac{3}{4}\delta^2_1\delta^2_2\right]$$

$$\underline{\sin 3\omega_{1,2}t} \; : \; 3\omega_{1,2}\delta^3_{1,2}\left[\frac{1}{12} + \frac{1}{16}\delta^2_{1,2} + \frac{1}{4}\delta^2_{2,1}\right]$$

$$\underline{\sin 5\omega_{1,2}t} \; : \; 5\omega_{1,2}\,\delta^5_{1,2} \cdot \frac{1}{80}$$

$$\underline{\sin(2\omega_{1,2} \pm \omega_{2,1})t} \; : \; (2\omega_{1,2} \pm \omega_{2,1})\,\delta^2_{1,2}\delta_{2,1}\left[\frac{1}{4} + \frac{1}{4}\delta^2_{1,2} + \frac{3}{8}\delta^2_{2,1}\right]$$

$$\underline{\sin(3\omega_{1,2} \pm 2\omega_{2,1})t} \; : \; (3\omega_{1,2} \pm 2\omega_{2,1})\,\delta^3_{1,2}\delta^2_{2,1} \cdot \frac{1}{8}$$

$$\underline{\sin(4\omega_{1,2} \pm \omega_{2,1})t} \; : \; (4\omega_{1,2} \pm \omega_{2,1})\,\delta^4_{1,2}\delta_{2,1} \cdot \frac{1}{16}$$

VII. CONCLUDING REMARKS

In this report, we have attempted to cover various aspects of the theory of the response of linear filters to variable-frequency excitations. This topic has been the subject of many published papers — some original, some interpretive. The following brief comments are intended to clarify a few of the special contributions of this report that might be less obvious to the unoriented reader.

As a preamble to the main discussion, Section II shows how two important expansions in FM theory can be derived by identical arguments that differ only in the initial arrangement of the integrand in the expression for the filter response. A new derivation is thereby presented for the Carson and Fry expansion, but the argument that yields the van der Pol-Stumpers expansion is adapted from Clavier's work (6). Our proposed unified argument is not only more straightforward (and teachable!) than the original arguments of van der Pol and Stumpers and of Carson and Fry, but also the limitations are more clearly brought out. This has enabled us to make certain definite statements about the convergence properties of these expansions that have never — to our knowledge — been made before. For example, it should now be clear under just what conditions these expansions become asymptotic, and that they are not asymptotic as they stand. This point, we feel, was not clarified by van der Pol and Stumpers in their papers (4, 5). (Incidentally, in Table II we have carried out the van der Pol-Stumpers expansion to more terms than has been done thus far, and we have pointed out how the n^{th} term can be generated with a minimum of manipulation.)

Of central importance in this report are the analyses presented in Sections III and IV. In general, the usefulness of a series expansion is severely hampered by the absence of explicitly stated bounds on the penalties involved in using only a few of its leading terms. Helpful statements to this effect are conspicuously lacking in all papers (that we have encountered) that discuss or use the Carson and Fry or the van der Pol-Stumpers expansions. We therefore feel that our derivations in Sections III and IV will be found to fill an important gap in this theory and should therefore enhance the usefulness of these expansions. Moreover, we feel that the sequence of terms in the asymptotic expansion Eq. (6) is the proper arrangement for the error criterion of asymptotic expansions to apply. However, Stumpers, who was the first to derive the expansions in their asymptotic forms, did not invoke this criterion in his paper.

The condition we have derived in Section IV is invaluable to FM theory. The most important and novel aspect of this condition is that it specifies an upper bound on the error that will be incurred in assuming that the filter will respond in a quasi-static manner to the applied FM excitation. Quasi-static response is a necessary condition for low-distortion transmission and FM-to-AM conversion. The condition specified by inequality (17) (or its milder form, 15) should therefore form the solid basis for prescribing linear-system bandwidths to meet the requirements of low-distortion transmission and of FM-to-AM conversion. In essence, the satisfaction of this inequality is the condition for the validity of using the instantaneous-frequency concept as a tool in studying the response of a linear system to an FM signal. It should therefore resolve many arguments on the significance of the concept of instantaneous frequency; on how fast is fast frequency sweep of a filter; and how sluggish is sluggish. This condition should thus show clearly when the so-called dynamic and static responses of transmission filters and of discriminators become essentially one and the same for a specified class of frequency modulations. (Such a class of frequency modulations would be specified by a bound on the maximum slope of the instantaneous-frequency waveform.)

In his paper (4), van der Pol does postulate that the second term in his expansion must be negligible in comparison with the first (which is really what is said by our inequality (16)) for quasi-stationary response, but it is not evident from his analysis exactly why the negligibility of the second term will insure the negligibility of any or all of the later terms.

In Section V, we introduce and stress the usefulness of the concept of sluggishness ratio of a filter. The sluggishness ratio plots that are presented should serve as a guide and aid to engineers in the design and evaluation of FM circuits. Existing approaches to the problem of specifying filter bandwidths in FM systems seem to concentrate on the spectrum of the FM signal and to ignore the type of filter that will be used. Our results do not only point out the pitfalls of this attitude, but also they show that the more complete criterion (that involves the significant characteristics of the filter and of the modulation jointly) is simpler and more compact.

Finally, our analysis of the instantaneous frequency of the response in Section VI is intended to emphasize that low-distortion reproduction of a specified frequency modulation requires, first and foremost, quasi-stationary response and, second, linear phase-versus-frequency behavior. The residual distortion (in the instantaneous-frequency response of a filter that is intended for low-distortion transmission) is studied in general terms and the evaluation of the harmonic and intermodulation distortion components is illustrated.

References

1. E. J. Baghdady, Frequency-modulation interference rejection with narrow-band limiters, Proc. IRE 43, 51-61 (1955).

2. E. J. Baghdady, Interference rejection in FM receivers, Technical Report 252, Research Laboratory of Electronics, M.I.T., Sept. 24, 1956.

3. J. R. Carson and T. C. Fry, Variable-frequency electric circuit theory, Bell System Tech. J. 16, 513-540 (1937).

4. Balth. van der Pol, The fundamental principles of frequency modulation, J.I.E.E. (London), Part III, 93, 153-158 (1946).

5. F. L. H. M. Stumpers, Distortion of frequency-modulated signals in electrical networks, Commun. News, 9, 82-92 (1948).

6. A. G. Clavier, Application of Fourier transforms to variable-frequency circuit analysis, Proc. IRE 37, 1287-1290 (1949).

7. K. Knopp, Theory and Application of Infinite Series (Hafner Publishing Co., New York, n. d.), p. 180.

8. T. J. I'a Bromwich, Theory of Infinite Series (Macmillan Company, London, 1908).

9. E. J. Baghdady, Theory of stronger-signal capture in FM reception, Proc. IRE 46, 728-738 (1958).

10. E. J. Baghdady, Theory of feedback around the limiter, IRE Convention Record, Part 8, 1957, pp. 176-202.