AN ANALYSIS AND SYNTHESIS PROCEDURE FOR FEEDBACK FM SYSTEMS

ANDRZEJ WOJNAR

TECHNICAL REPORT 415

SEPTEMBER 30, 1963

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

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) under Contract DA36-039-sc-78108, Department of the Army Task 3-99-25-001-08; and in part by Signal Corps Grant DA-SIG-36-039-61-G14.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 415 September 30, 1963

AN ANALYSIS AND SYNTHESIS PROCEDURE FOR FEEDBACK FM SYSTEMS

Part A: Conventional FM Systems

Part B: Feedback FM Systems

Andrzej Wojnar

(Manuscript received February 27, 1963)

Abstract

An investigation of frequency-compressive feedback FM systems has been made, with emphasis placed on threshold behavior.

In Part A, following a survey of the existing noise analysis of FM systems, use is made of a recent evaluation of the impulsive noise component which is due to Rice. It is shown that this excess-noise component predominates in the threshold region of most conventional FM systems. A new analytical expression (which agrees with experimental evidence) has been found for the location of the noise threshold.

In Part B, two possible mechanisms causing noise threshold in a feedback FM system are examined. It is shown that the feedback threshold results in an abrupt breakdown of system performance, and therefore should not be approached too closely. It is therefore recommended that feedback systems be designed with the conventional threshold predominant. In this case, the analysis of feedback systems approaches that of conventional systems. Some modified formulas for feedback FM system performance at threshold are proposed; new bounds on the performance of the systems with optimum feedback filters are derived. The maximum obtainable threshold power-bandwidth tradeoff implied thereby is also determined. Experimental results with certain feedback FM configurations are reported. The experimental investigation verifies some synthesis rules and certain of the analytical results, but some aspects of the system threshold behavior still require further investigation.

TABLE OF CONTENTS

PART A: CONVENTIONAL FM SYSTEMS

PART B: FEEDBACK FM SYSTEMS

PART A: CONVENTIONAL FM SYSTEMS

1. Noise Performance

Ultimate limitations in the performance of communication systems are imposed by random noise at the receiving end. In the first part of this report we shall present an analysis of conventional FM systems disturbed by additive fluctuation noise. In the second part, we consider the frequency-compressive feedback case.

In order to analyze an FM system, we shall use an idealized mathematical model. The model (Fig. 1) consists of a symmetrical narrow-bandpass filter, an ideal broad-band limiter with zonal filter, and an ideal frequency detector that operates in quasi-stationary fashion and is followed by a lowpass filter.

Under certain plausible assumptions[1,2] the input noise can be considered to be derived from a white Gaussian noise source, and the input to our frequency demodulator to be the sum of an FM signal and narrow-band Gaussian noise. The analysis is then reduced to differentiation in time of the phase angle of the composite wave.

After the detector output wave has been expressed as an implicit time function, the Fourier transform of its correlation function yields the spectrum of the demodulated noise. A straightforward filtering operation then gives the output spectrum and the total output noise power.

A great deal of information has resulted from analyses of this sort by Rice,[3] Stumpers,[4] and Wang[5] with regard to the rectangular and normal-law bandpass filter cases. Extended analysis by Middleton[6-8] has also covered more general cases of nonideal amplitude limiting and several different filter structures.

The nonlinear interaction between signal message and noise has been carefully investigated, in this same way, but leads to prohibitively complicated mathematical expressions. Some important cases have been evaluated by numerical computation and shown in graphs of input-output relations.[4,7,8] The analysis of limiting cases has always been quite well advanced; thus a general picture of the whole area emerges as a composite of several different regions, each characterized by different components of output noise. For this reason, different definitions of the noise threshold are also possible.[2,9]

Early attempts to translate the analysis into system-oriented signal-to-noise ratio formulas and charts[7,10] were adequate to meet some needs of conventional FM

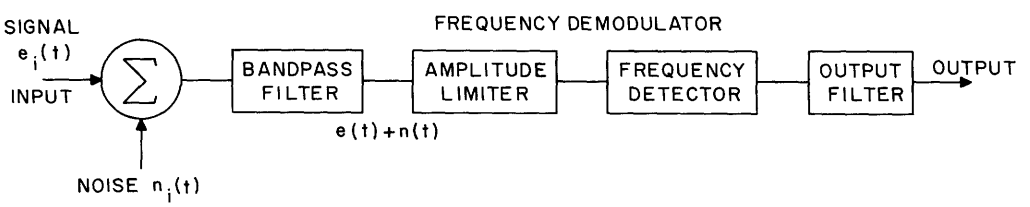

Fig. 1. Model of the FM receiving system.

engineering until the advent of feedback FM systems. New problems are now arising from new system principles and, thanks to phenomenological investigations, more insight into the physical character of output noise has been gained. Therefore a strong trend toward modifying the conventional (correlation function) analytic approach – which is more suitable for linear-modulation systems – parallels the research on hitherto unexplored (or unsatisfactorily explored) areas. As a result, important new contributions that are concerned with the structure of the nonlinearly processed noise are beginning to appear.

In this report, we shall first survey and explore certain new results of an analysis which facilitate the resolution of output noise into different components. Looking upon the input-output relations in an additively disturbed FM system, we propose consideration of the following regions of interest (Fig. 2):

(a) Strong carrier, weak noise: The output noise is nearly Gaussian, and linear superposition with message holds. This is the only region of unrestricted practical usefulness.

(b) Region of noise threshold: Two sharply rising components (quasi-Gaussian and impulsive) of output noise are superimposed on the output signal. Practical interest is confined to the boundary between regions 1 and 2.

Below the threshold, superposition does not hold any more; a pronounced fall in output signal power occurs until the message is fully captured by the disturbance. The well-defined limiting case of noise alone (no carrier) is sometimes used as a reference point (cf. the noise-quieting sensitivity).

For the model considered above, the input carrier-to-noise power ratio (CNR) is the only major parameter influencing the system performance. Conventionally, the latter has been expressed in terms of the output signal-to-noise power ratio (SNR). But it does not seem meaningful to cover both regions of interest with this single measure of performance.

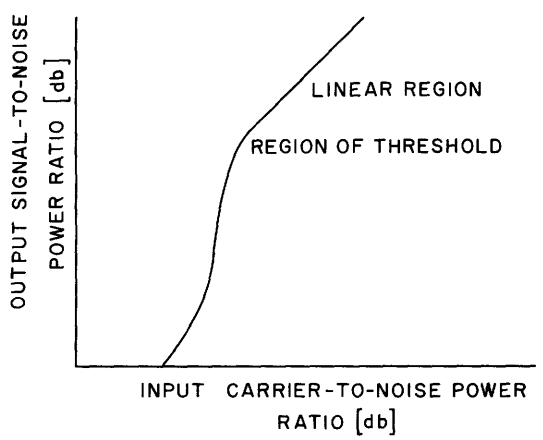

Fig. 2. Input-output relation in an FM system.

Additional hazard is also involved in plotting the signal-to-noise ratio over a wide range of variables, when the available numerical results are, in fact, limited to small ranges of values. For example, exploiting the analysis and computations by Rice,[3] Skinner in an unpublished memorandum[10] aimed at graphical presentation of the system performance over a wide range of CNR. Unfortunately, his plot was obtained by fairing together the different available graphs for fragments of the individual regions. In the important region of threshold, the necessary accuracy was missing. As a result, the threshold curve derived therefrom by Replogle[11] is far from accurate and can be misleading, as we shall see in Fig. 5.

This report will, in general, be confined to the analysis of regions 1 and 2 described above, both of which are characterized by the absence of first-order interaction between noise and signal. After recapitulation of known data, some extension of the existing noise analysis will be presented. The effective disturbance in these regions can be attributed to an additive noise rather than involving nonlinear distortion of the message also.

2. Simplified Analysis of the Strong-Carrier Case

Consider our idealized FM system in Fig. 1 to be disturbed by white Gaussian noise with spectral power density equal to N so that, after bandpass filtering, narrow-band Gaussian noise of the form

$$n(t) = n_c(t) \cos \omega_c t - n_s(t) \sin \omega_c t \tag{1}$$

enters the frequency demodulator. The FM signal at the same point may be expressed as

$$e(t) = Q \cos [\omega_c t + \phi(t)], \tag{2}$$

where the carrier amplitude Q and the carrier frequency $\omega_c = 2\pi f_c$ are constants. The message is bandlimited and is contained in the time derivative of the signal phase angle, $\phi'(t)$.

Assuming that the filtered noise is weak relative to the carrier, and denoting the demodulator input sum as

$$e(t) + n(t) = V(t) \cos [\omega_c t + \phi(t) + \theta(t)], \tag{3}$$

we are interested first in determining the additive disturbance $\theta'(t)$ at the output of the demodulator. Assume that part of the output filter suffices to pass all the message, but sharply cuts off the higher frequency components of $\theta'(t)$, in order to improve the output SNR. We shall evaluate the effect of this filtering in terms of output noise spectrum and power.

Let us further denote the (positive frequency) power density spectrum of the I-F noise by w(f) which (within a multiplicative constant) equals the squared magnitude response of the bandpass filter. With the demodulator input average CNR denoted by r,

we have the obvious relation

$$r = \frac{Q^2/2}{\overline{n^2(t)}} = \frac{Q^2}{2\int_0^\infty w(f)\,df} \tag{4}$$

which holds with arbitrary modulation (and also for $\phi = 0$).

We first state that region 1 is characterized by $r \gg 1$ so that the noise magnitude

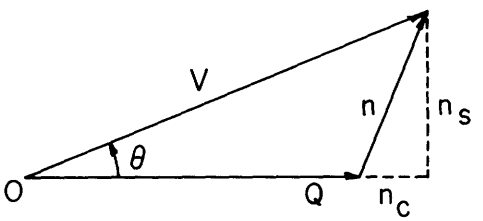

Fig. 3. Phase relations between carrier and noise.

$|n(t)|$ is nearly always much smaller than Q. In the simplest case of an unmodulated carrier, the input signal and noise combine as visualized in Fig. 3, and we easily find that

$$\theta(t) = \tan^{-1}\frac{n_s(t)}{Q + n_c(t)} \approx \frac{n_s(t)}{Q}. \tag{5}$$

Rigorously, this approximation holds true within an additive multiple of 2π.

When the filter response is symmetrical about f_c, the power density spectrum of $n_c(t)$ and $n_s(t)$ has been found[3] to be $2w(f_c+f)$. Therefore the normalized* power density spectrum of the derivative $n_s'(t)$ equals $f^2 \cdot 2w(f_c+f)$, and the power density spectrum, W(f), of the disturbance $\theta'(t) \approx n_s'(t)/Q$ can be expressed approximately as

$$W(f) \approx f^2 \times 2w(f_c+f)/Q^2, \tag{6}$$

provided that $r \gg 1$.

The above-mentioned treatment of FM noise was initiated by Crosby,[12] who considered the simplest case of rectangular filters and white noise. Denoting the input (intermediate-frequency) filter passband by B and the output lowpass band by f_a, we may repeat Crosby's integration in order to find the output power N of the (weak) Gaussian noise. We obtain

$$N = \int_0^{f_a} W(f)\,df = \frac{f_a^3}{3rB} \tag{7}$$

*Here normalization consists in assuming that the demodulation constant and the load resistance are unity and dimensionless.

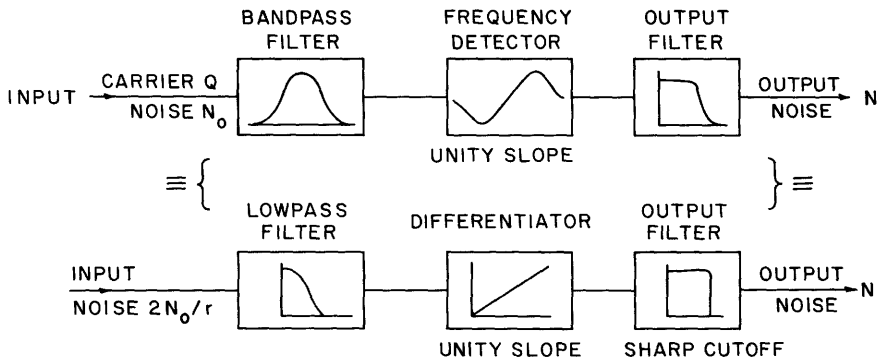

Fig. 4. Linear model for noise transmission.

in normalized units.

It is clear that the noise output (7) from an FM system in region 1 is identical to that produced at the output of the three linear filters shown in Fig. 4, in which the demodulator has been replaced by a differentiator weighting the original spectrum by $(f-f_c)^2$. Thus, as far as noise analysis (with large CNR) is concerned, we can replace the actual nonlinear receiver of Fig. 3 by the linear model of Fig. 4, operating upon a white spectrum of input noise.

Although a rectangular bandpass filter may be regarded as a rough approximation for a chain of double-tuned intermediate-frequency filters (used in most conventional systems), it is not the only good noise-filtering arrangement. Let us define the equivalent noise bandwidth B of the input filter as

$$B = \frac{1}{w(f_c)} \int_0^\infty w(f)\, df, \tag{8}$$

and consider a family of different symmetrical narrowband input filters characterized by varying w(f) under the constraint that B remains fixed. It is conjectured from FM distortion theory[13] that over a broad class of reasonable variations, the message transmission does not undergo significant changes. On the other hand, it is easy to see that a slow roll-off filter spreads the energy of FM noise outside of the output band, and thus contributes to the decrease of output noise power.

One convenient measure of the noise-power transfer through the first two units of the filter chain in Fig. 4 can be conceived[2] in terms of p, the radius of gyration about the symmetry axis, defined by

$$p^2 = \frac{\int_0^\infty (f-f_c)^2\, w(f)\, df}{\int_0^\infty w(f)\, df}. \tag{9}$$

Here, the denominator describes the noise entering the demodulator, and the numerator

the noise output from the demodulator.

In the complete filter chain of Fig. 4, the effect of the sharp-cutoff output filter is to chop off the output spectrum at f_a. This, however, can also be accounted for by imposing a finite limit of integration in the numerator of Eq. 9. We therefore define for the baseband analog another coefficient p_a, describing the noise transfer in the entire chain:

$$p_a^2 = \frac{\int_0^{f_a} 2f^2 w(f_c+f)\, df}{\int_0^{\infty} w(f)\, df}. \tag{10}$$

All of these definitions apply to the output of any symmetrical bandpass filter. They have been introduced primarily because the output noise power of the system can then be expressed by means of the simple relation

$$N = p_a^2/2r, \tag{11}$$

which follows directly from (7), with appropriate substitution of (4), (6), and (10). Notice that for $f_a \gg \frac{1}{2}B$ we have in the limit $p_a = p$; more important is the case of narrow-band lowpass filtering with $f_a \ll \frac{1}{2}B$ because, then, to a first-order approximation, the noise output does not depend greatly on the shape of the bandpass spectrum w(f), provided that the former is symmetrical and flat in the middle.

It will prove useful in the sequel to present in Table 1 a compilation of the properties of some filtering arrangements. The normal-law filter has been included as a good approximation to the cascade of many identical single-tuned filters; the single-tuned filter was evaluated because of its importance for the feedback FM systems. With our interests limited to the region $f_a < \frac{1}{2}B$, we have a well-defined p_a and the output power is bounded, even for the single-tuned filter. Also, the single-tuned filter really has good filtering capability in FM systems, as can be observed by comparing the values of p_a. Or, still better, consider the ratio p_a/B, which is one figure of merit for the filtering of noise, provided that B is regarded as a measure of the bandwidth required for tolerable transmission of the signal message.

3. Simple Approach to the Threshold Problem

The exact graphical representation of noise output power as a function of input CNR (see, for instance, the work of several authors[3-8]) clearly exhibits a relatively sharp break region. The question arises whether we can sensibly describe the location of the break, called the threshold, in the linear input-output characteristic of the system by one discrete value of input CNR. This approach, although mathematically questionable, nevertheless has important advantages in the analysis and synthesis of systems, and therefore seems to be worth following.

We are therefore in search of a border line to the linear (weak-noise) region. Recall

that the approximated linear equation (5),

$$\theta(t) \approx n_s(t)/Q,$$

holds in region 1, provided that $r \gg 1$. Actually, three simplifying steps (corresponding to various excess-noise components) have been made in the approximation:

$$\tan^{-1} \frac{n_s(t)}{Q + n_c(t)} \approx n_s(t)/Q, \qquad \text{if and only if } \overline{n_c^2} \ll Q^2,\ \overline{n_s^2} \ll Q^2.$$

The first approximation consists of neglecting the relatively small summand in the denominator; the second replaces the arc tangent function by its argument – without adding the multiple of 2π, which is the third approximation. The usual procedure would be to determine the most restrictive of these three approximations, and to find its region of validity. All three simplifying approximations may be significant, and considering but one of them may be misleading.

Let us, for the moment, follow the classical approach, disregarding the multivaluedness of the arc tangent function. In order to find a significant departure from linearity, we shall choose an expansion for which the right-hand side of (5) represents the leading term. Such a result may be derived in terms of the mean-square values from Rice's asymptotic expansion[3] of the autocorrelation function of θ, given in 1948. This expansion is expressed in terms of the autocorrelation function of the noise waveform $\overline{n(t)\, n(t+\tau)}$; at the origin ($\tau = 0$), by virtue of (4), we have

$$\overline{n^2} = \overline{n_s^2} = \overline{n_c^2} = Q^2/2r \tag{12}$$

in terms of the average carrier-to-noise power ratio r.

Evaluating Rice's equation (his Eq. 7.3) at the origin, we have in our notation:

$$\overline{\theta^2} = \frac{\overline{n^2}}{Q^2}\left[1 + \frac{\overline{n^2}}{Q^2} + \frac{8}{3}\frac{(\overline{n^2})^2}{Q^4} + \ldots\right]$$

or

$$\overline{\theta^2} = \frac{1}{2r}\left(1 + \frac{1}{2r} + \frac{2}{3r^2} + \ldots\right). \tag{13}$$

The same result can be obtained from the expansion of the arc tangent function in a power series, whereby the argument itself is represented by an expansion (compare also Eq. 10 in a recent paper by Baghdady[14]). Note that if the input is Gaussian, the first term of (13) represents Gaussian output noise. The first-order correction term is an important component of excess noise; it is not Gaussian, but we may call it quasi-Gaussian. Obviously, this noise term (when significant) will cause the input-output relation to depart from linearity. Now a simple inequality can be used to define the border of the linear

region, or the threshold caused by the quasi-Gaussian excess noise, say

$$1 \gg \frac{1}{2r} \quad \text{or} \quad r_T \geq \frac{1}{2}c, \tag{14}$$

where c is an arbitrary constant parameter, and the subscript "T" marks the threshold as a point rather than a region. A reasonable value for c, whereby (14) is satisfied, is c = 0.1; then we obtain $r_T = 5$.

In accordance with this choice of c, a carrier-to-noise power ratio of approximately 5 times, or 7 db, would be needed to pass the threshold. The reader may be skeptical that such a low value will suffice, but it must be borne in mind that (13) does not necessarily show all sources of excess output noise. When entering the threshold region from the linear region, we do, of course, encounter this first-order nonlinear term, derived from Gaussian noise; however, another noise term of quite different character will also appear, which we shall now discuss.

4. Analysis of the Impulsive Noise

A very important contribution by Rice[15] has recently provided a different approach to the analysis of noise in FM systems. In order to justify the significant conclusions drawn from Rice's argument, let us briefly summarize his new results.

Even well above the threshold, it sometimes occurs that over a short interval of time the magnitude of the random-noise vector in Fig. 3 exceeds the length of the carrier vector and, simultaneously, the noise phase angle $(\tan^{-1}[n_c(t)/n_s(t)])$ passes through $\pm\pi$. We can then expect that the sum of carrier and noise will encircle the origin, and that the phase angle θ will rapidly be increased by $\pm 2\pi$. Thus an impulse is produced in the $\theta'(t)$ waveform with an area of approximately $\pm 2\pi$ radians per second. After low-pass filtering, broadened impulses of either polarity appear in the output and are heard as clicks in the audio band, or seen as lengthy spots on the screen of an FM television system. Noise of this sort is related to the multivalued nature of the arc tangent function (which has been neglected in section 3).

Rice succeeded in determining the expected number of "clicks" per second in terms of the input carrier-to-noise power ratio and of the parameters of the input filter. Assuming a Poisson distribution of the arrival times of the pulses, Rice also computed the corresponding power spectral density and the output power in the output band.

Above the threshold region, the two excess output-noise components (quasi-Gaussian and impulsive) seem to be uncorrelated. This being the case, a comparison with earlier results obtained by autocorrelation analysis could be performed by addition of noise terms on the power basis. This reasoning has been checked by Rice with numerical computations extending slightly into the threshold region. It must, however, be borne in mind that the impulsive noise is very specific in character and spectral distribution.

On the other hand, it appears that Rice's evaluation of impulsive noise can be explored in another way, and leads to a new measure of noise threshold. The

fundamental formula[16] of Rice

$$\nu = p(1-\mathrm{erf}\sqrt{r}) = p\frac{2}{\sqrt{\pi}}\int_{\sqrt{r}}^{\infty} e^{-y^2}\,dy, \tag{15}$$

where p is the radius of gyration defined by Eq. 9, gives the expected total number, ν, of (upward and downward) clicks per second in the threshold region and above it, with unmodulated carrier. The gyration radius of the symmetrical bandpass filter is, in general, proportionally related to its noise bandwidth (Table 1) so that, for a given filter, ν/B depends only on the carrier-to-noise power ratio, r.*

The random succession of upward and downward impulses has a spectrum that is substantially flat at lower frequencies, whereas the Gaussian output noise spectrum is proportional to the square of the baseband frequency. In order to stress this difference, Rice denoted the constant power spectral density of the impulsive noise by W(0), and by straightforward reasoning obtained the result

$$W(0) \approx 2\nu = 2p(1-\mathrm{erf}\sqrt{r}). \tag{16}$$

With sharp-cutoff frequency of the output filter denoted by f_a, the output power of the impulsive noise is evidently

$$N_i \approx 2pf_a(1-\mathrm{erf}\sqrt{r}) \sim 2pf_a\frac{e^{-r}}{\sqrt{\pi r}}, \tag{17}$$

with no modulation present. Note, by comparing (17) with (13), that the impulsive excess-noise component differs substantially from the quasi-Gaussian one, which changes with input CNR as $1/r^2$.

In a further step, Rice computed the effect of modulation on the impulsive noise, and found a significant interaction between message and noise. The number of clicks increases considerably with the signal-frequency deviation. Rice made an approximate evaluation for the case of sinusoidal modulation, whereby the frequency is deviated over the entire band of a rectangular bandpass filter. Simplifying Rice's result (his Eq. 2.26), we find that a first approximation to the value of impulsive-noise output power is given by

$$N_i \sim \frac{2}{\pi}e^{-r}Bf_a, \tag{18}$$

provided that $r \gg 1$. A small correction term, diminishing the quasi-Gaussian noise in the same case, is also given by Rice.[15]

We shall now consider the location of the threshold in terms of the parameters of the impulsive noise. It will be shown that even with the noise power as a measure of

*For the mathematical model of the single-tuned circuit we have a singularity because the gyration radius increases without bounds; no method of overcoming this difficulty has been found (S. O. Rice, private communication, 1963).

Table 1. Noise-filtering properties of some bandpass filters.

		A. Rectangular Filter			B. Normal-Law Filter		C. Single-Tuned Filter	
		General Case	Limiting Cases $f_a = f_b$	Limiting Cases $f_a \ll f_b$	General Case	Limiting Case $f_a \ll f_b$	General Case	Limiting Case $f_a \ll f_b$
Input Spectrum	$w(f)$	$w_o,\ \lvert f-f_c\rvert < f_b$ $0,\ \lvert f-f_c\rvert > f_b$			$w_o \exp\left[-\left(\frac{f-f_c}{f_b}\right)^2\right]$		$\frac{w_o}{1+\left(\frac{f-f_c}{f_b}\right)^2}$	
Input Power	Bw_o	$2f_b w_o$	$2f_a w_o$	$2f_b w_o$	$\sqrt{\pi} f_b w_o$		$\pi f_b w_o$	
Gyration Radius	p	$\frac{f_b}{\sqrt{3}} = \frac{B}{\sqrt{12}}$	$\frac{f_a}{\sqrt{3}}$	$\frac{f_b}{\sqrt{3}}$	$\frac{f_b}{\sqrt{2}} = \frac{B}{\sqrt{2\pi}}$			
Output Spectrum	$W(f)$	$\frac{2w_o f^2}{Q^2} = \frac{f^2}{Br}$	$\frac{f^2}{2f_a r}$	$\frac{f^2}{2f_b r}$	$\frac{2w_o f^2}{Q^2} \exp\left(-\frac{f^2}{f_b^2}\right) = \frac{f^2}{Br} \exp\left(-\frac{f^2}{f_b^2}\right)$		$\frac{2w_o}{Q^2} \times \frac{f_b^2 f^2}{f^2+f_b^2} = \frac{B}{\pi^2 r} \times \frac{f^2}{f^2+f_b^2}$	
Output Power	N	$\frac{f_a^3}{3Br}$	$\frac{f_a^2}{6r}$	$\frac{f_a^3}{6f_b r}$	$\frac{B^2}{4\pi r} \operatorname{erf}\left(\frac{\sqrt{\pi} f_a}{B}\right) - \frac{Bf_a}{2\pi r} \exp\left(-\frac{\pi f_a^2}{B^2}\right)$	$\frac{f_a^3}{3Br}$	$\frac{B}{\pi^2 r}\left(f_a - \frac{B}{\pi}\tan^{-1}\frac{\pi f_a}{B}\right)$	$\frac{f_a^3}{3Br}$
Output Gyration Radius	p_a	$\sqrt{\frac{2}{3}\frac{f_a^3}{B}}$	$\frac{f_a}{\sqrt{3}}$	$\sqrt{\frac{f_a^3}{3f_b}}$	$\sqrt{\frac{2}{3}\frac{f_a^3}{B}\left[1 - \frac{3\pi}{5}\left(\frac{f_a}{B}\right)^2 + \frac{3\pi^2}{14}\left(\frac{f_a}{B}\right)^4 - \cdots\right]^{1/2}}$	$\sqrt{\frac{2}{3}\frac{f_a^3}{B}}$	$\frac{1}{\pi}\sqrt{2B}\left(f_a - \frac{B}{\pi}\tan^{-1}\frac{\pi f_a}{B}\right)^{1/2}$	$\sqrt{\frac{2}{3}\frac{f_a^3}{B}}$
Figure of Merit	$\frac{p_a}{B}$	$\sqrt{\frac{2}{3}\left(\frac{f_a}{B}\right)^3}$	$\frac{1}{2\sqrt{3}}$	$\sqrt{\frac{f_a^3}{12f_b^3}}$		$\sqrt{\frac{2}{3}\left(\frac{f_a}{B}\right)^3}$	$\frac{1}{\pi}\sqrt{\frac{2}{B}}\left(f_a - \frac{B}{\pi}\tan^{-1}\frac{\pi f_a}{B}\right)^{1/2}$	$\sqrt{\frac{2}{3}\left(\frac{f_a}{B}\right)^3}$

disturbance, the impulsive component is very significant, and, in fact, determines the extent of the linear region.

5. Evaluation of Threshold

Rice's results facilitate the physical interpretation of the threshold as being caused by the exponentially increasing impulsive component of output noise. As the first-order quasi-Gaussian excess-noise component is generally proportional to $1/r^2$, it will be predominant with a very strong carrier. As r decreases into the threshold region, however, we find that the impulsive contribution imposes a more severe limitation on the extent of the region of linearity. Also, its character is especially annoying with some types of messages.

Let us examine the simplest case of rectangular input filter and unmodulated carrier. We shall first determine that value of r for which the power of the impulsive noise is no longer negligible in comparison with the linear Gaussian noise (11).

The threshold inequality, approximated by setting

$$1 - \mathrm{erf}\sqrt{r} \approx \frac{1}{\sqrt{\pi r}} e^{-r}, \tag{19}$$

now is

$$2pf_a \frac{1}{\sqrt{\pi r}} e^{-r} \ll \frac{p_a^2}{2r}. \tag{20}$$

For the rectangular filter, we find from Table 1 that

$$p = \frac{B}{\sqrt{12}} \qquad p_a^2 = \frac{2}{3}\frac{f_a^3}{B}, \tag{21}$$

so that the condition of Eq. 20 becomes

$$\frac{e^r}{\sqrt{r}} \gg \sqrt{\frac{3}{\pi}} \left(\frac{B}{f_a}\right)^2 \approx \left(\frac{B}{f_a}\right)^2. \tag{22}$$

The inequality can be treated similarly to (14), and this yields

$$c\frac{e^r}{\sqrt{r}} \geqslant \left(\frac{B}{f_a}\right)^2. \tag{23}$$

With c = 0.1, as before, we obtain a transcendental equation determining the threshold CNR:

$$e^{\frac{1}{2}r_T} \approx \sqrt[4]{100 r_T}\, \frac{B}{f_a} \tag{24}$$

with the rectangular bandpass filter.

Note, first of all, that the location of the impulsive-noise threshold is now a monotonically increasing function of B/f_a. For the lowest practicable value, $\frac{B}{f_a} = 2$, we find that $r_T(2) = 4.35$, which is actually below the bound imposed by the over-simplified argument (14). With B/f_a as small as 2.58, however, the two terms of excess noise have equal power (cf. Eq. 14). It follows that (with customary values of $\frac{B}{f_a} \geq 3$) practically every FM system of interest will be limited by the impulsive-noise threshold.

For certain types of messages, the noise power is not the best measure of disturbance, especially if the noise consists of two components of different character. It is then not meaningful to manipulate with SNR below the threshold of an FM system.

This argument defines the threshold in terms of the departure from the linear power relation caused by the impulsive noise. As the masking effect of impulsive noise may be very annoying, even if the corresponding noise power is still low,[14] it may also be sensible to define the threshold in terms of the average number of clicks per second. Observe that Eq. 24 can be used for this purpose, and defines threshold as $\nu_T = f_a^2/60rB$ clicks per second.

The analysis for the rectangular filter can also be carried out for the normal-law filter. From Table 1 we have

$$p = \frac{B}{\sqrt{2\pi}} \qquad p_a^2 = \frac{2}{3}\frac{f_a^3}{B}\left(1 - \frac{3\pi}{5}\frac{f_a^2}{B^2} + \ldots\right). \tag{25}$$

Neglecting the correction term in parentheses, we find to a first-order approximation

$$e^{\frac{1}{2}r_T} \approx \sqrt[4]{200\, r_T}\,\frac{B}{f_a}. \tag{26}$$

Both Eqs. 24 and 26 have been numerically solved by the author; the result is plotted in Fig. 5. (The threshold curve by Replogle[12] is also shown for purposes of comparison.) Our curves are believed to be quite accurate over the entire range of values $r_T \geq 5$; below this region, that is, near the origin, our simplified analysis yields only a relatively crude estimate of the location of threshold.

Perhaps the most significant result is the threshold curve derived from the sinusoidal-modulation case, as given by Eq. 18. Using Eqs. 18 and 7 [or (11)], we obtain for the threshold condition with a rectangular bandpass filter

$$\frac{2}{\pi} e^{-rBf_a} \ll \frac{f_a^3}{3Br}. \tag{27}$$

Then, from reasoning identical to that leading to Eq. 24, we have

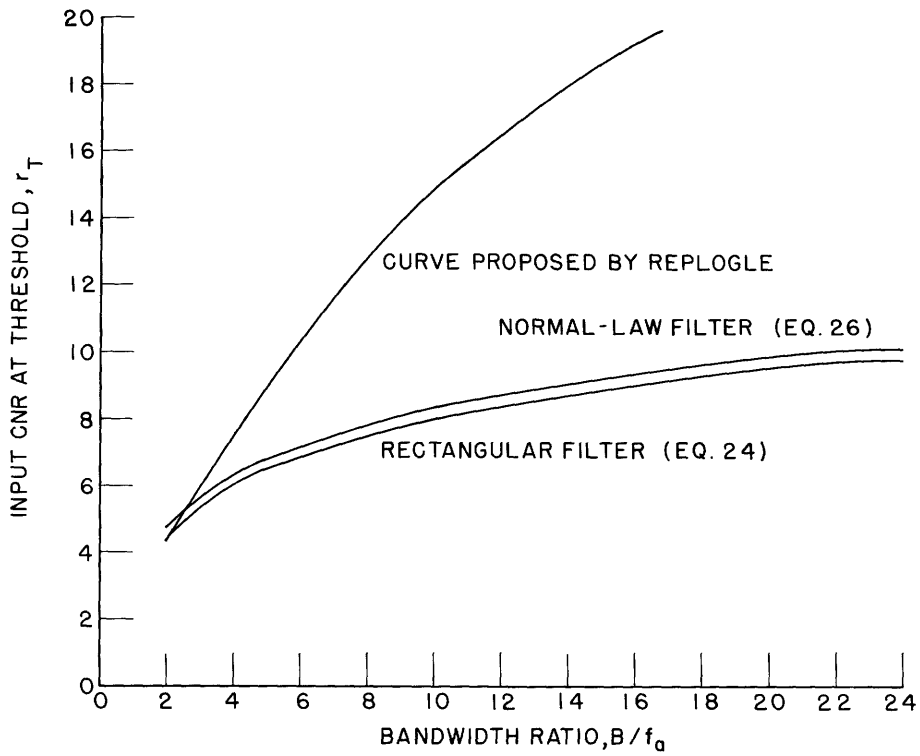

Fig. 5. Noise threshold curves for an unmodulated carrier.

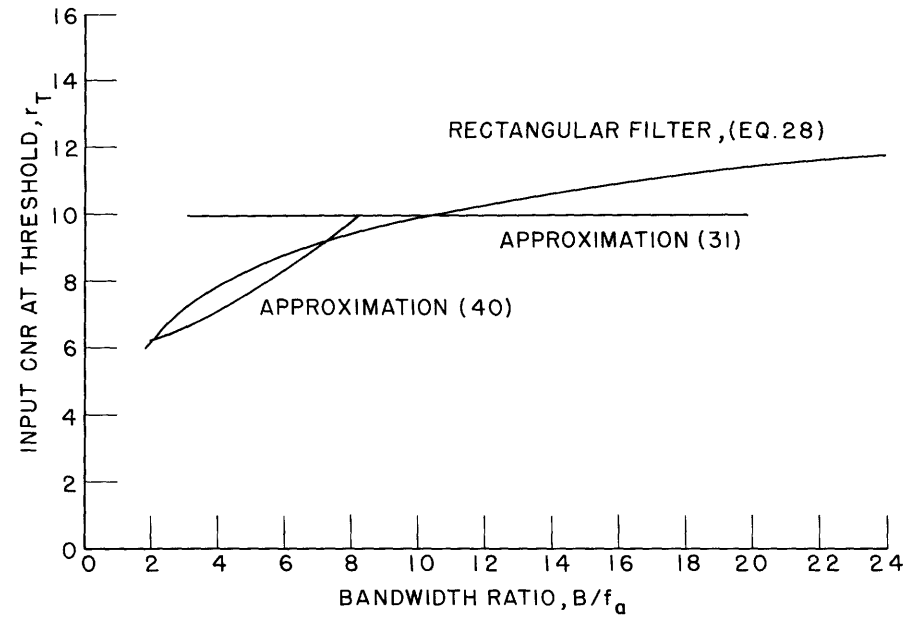

Fig. 6. Noise threshold curve for a modulated carrier.

$$\boxed{e^{\frac{1}{2}r_T} \sim \sqrt{20\, r_T \frac{B}{f_a}}} \,. \tag{28}$$

It is interesting to note that in this case r_T exceeds 5 for any value of $\frac{B}{f_a} \geq 2$, and is very close to 10 for $\frac{B}{f_a} = 10$. Thus, the widely held belief that approximately 10 db CNR is needed to pass the threshold in a conventional FM system finds confirmation.

In the sequel we shall use the solution of Eq. 28, shown in Fig. 6, as the location of the noise threshold in conventional FM systems. By itself, however, Eq. 28 does not facilitate a direct algebraic approach to the evaluation of the performance of such systems, and some further approximations will be needed.

6. Performance Analysis of Conventional FM Systems

It seems reasonable to assume that no FM system can satisfactorily operate below its (impulsive) noise threshold, since a sharp performance degradation is immediately noticed both with regard to the noise output power and to the appearance of noise clicks. Thus the performance of FM systems at and above threshold, as well as the location of the threshold in terms of received signal-power and noise-power spectral density, is of great interest. Since the minimum output signal-to-noise ratio, say, R_{min}, is usually specified for the system, it seems reasonable to compare R_{min} with the performance of the system at threshold, denoted here by R_T. Other quantities of importance that must be specified in advance are: the message bandwidth f_a, the system noise figure or noise power density N_o. The principal unknowns to be determined in system synthesis are: the modulation index m for the signal design, the minimum carrier power C for the transmitter design, and the filter bandwidth B for the receiver design.

The filter design will always be a compromise between minimizing the message distortion and obtaining the best noise filtering. Without entering the controversial area of nonlinear message distortion in linear FM systems, we shall confine our attention to the relation between the FM signal bandwidth and the filter bandwidth. Specifically, the maximum signal bandwidth will be considered to be determined by the modulation index $m = \frac{D}{f_a}$, where D denotes the maximum signal-frequency deviation with modulating frequency f_a. For purposes of engineering, it appears that the Van der Pol-Carson formula for the signal bandwidth,

$$B_s \approx 2f_a(m+1), \tag{29}$$

is accurate enough.*

*For the sake of precision, (29) may be extended as proposed by Manayev[17] so that it will account for all sidebands with amplitude exceeding 1 per cent of the unmodulated carrier amplitude. Then $B_s = 2f_a(1+m+\sqrt{m})$.

With large modulation index ($m \gg 1$), the term 1 in Eq. 29 can be neglected. Assume as a first measure for relatively undistorted signal transmission that the bandwidths of the signal and the bandpass filter are equal: $B_s = B$. Note also from Fig. 6 that, for large B/f_a, a fixed value is a good approximation for the threshold CNR in the range $7 \leq \frac{B}{f_a} \leq 17$, which corresponds to a modulation index in the range $2.5 \leq m \leq 7.5$. We therefore take

$$r_T \approx 10. \tag{30}$$

With actual systems, usually the input-noise power density N_o is fixed, rather than the power of noise filtered by the bandpass filter. We then have

$$r_T = \frac{C_T}{N_o B} \approx 10, \tag{31}$$

where C_T denotes the value of signal power C at the threshold. It is sometimes convenient to refer the CNR to the baseband width, f_a, in which case we define

$$r_a = \frac{C}{N_o f_a} = \frac{B}{f_a} r. \tag{32}$$

Observe that in the linear region the normalized signal-output power is

$$S = \frac{1}{2} D^2, \tag{33}$$

and with (7) we arrive at the well-established formula for the output signal-to-noise ratio

$$R = \frac{S}{N} \approx \frac{3rBD^2}{2f_a^3} = \frac{3}{2} m^2 r_a. \tag{34}$$

Thus, for the broadband FM system with nearly rectangular filters "matched" to the signal bandwidth (so that $\frac{B}{f_a} = 2m$), we have

$$R \approx \frac{3}{2} m^2 r \frac{B}{f_a} = \frac{3}{2} m^2 r \cdot 2m,$$

and at the threshold with (31),

$$R_T \approx 30m^3. \tag{35}$$

The "zero-order" solution of our system problems will then be

$$\boxed{\begin{aligned} m &= \sqrt[3]{R_T/30} \\ C_T &= 20 N_o m f_a \\ B &= 2mf_a \end{aligned}} \tag{36a, 36b, 36c}$$

Such a system exhibits a well-known inherent limitation of output performance, if carrier-power and noise-power density are fixed. From (31) and (32), we find

$$r_a = \frac{C}{N_o f_a} = \text{constant}, \tag{37}$$

so that the maximum value of

$$m = \frac{C}{N_o f_a r_T} \approx \frac{C}{10 N_o f_a} \tag{38}$$

is also limited. Consequently, for the maximum value of output SNR, we have

$$R_{max} \approx \frac{30}{r_T^3}\left(\frac{C}{N_o f_a}\right)^3 = 0.03\left(\frac{C}{N_o f_a}\right)^3. \tag{39}$$

This bound may be exceeded only by decreasing the noise-power density or by increasing the signal power.

For other systems, in which the modulation index is not necessarily much larger than one, say $m \leq 5$, we must take into account two corrections. The first must account for the significant fall in the threshold curve (Fig. 6); the second must allow for the signal bandwidth being larger than the full frequency swing $2D = 2mf_a$. In this case we use Eq. 29 without neglecting the unity term.

The transcendental character of the threshold equation (28) impedes a straightforward analytic approach. In looking for algebraic fits, it is convenient to select an expression leading to a manageable system analysis along the lines presented in Eqs. 30-36. We therefore suggest the following algebraic approximation for use in place of Eq. 28 in the range $\frac{B}{f_a} \leq 8$:

$$r_T = \frac{25}{32}\frac{B}{f_a}\left(2\frac{f_a}{B} + 1\right)^2. \tag{40}$$

Hence, substituting Eq. 29, after simplification we have

$$r_T = \frac{25}{16}\frac{(m+2)^2}{m+1}, \tag{41}$$

which is valid with $m \leq 3$.

The threshold CNR measured in the baseband width f_a according to (32) is

$$r_{a,T} = \frac{25}{8}(m+2)^2, \tag{42}$$

and the (linearized) output SNR at threshold (34) becomes

$$R_T = \frac{3}{2}m^2 r_{a,T} = \frac{75}{16}m^2(m+2)^2. \tag{43}$$

The following system equations follow immediately:

$$m = \sqrt{1 + \frac{4}{5}\sqrt{R_T/3}} - 1 \tag{44a}$$

$$C_T = \frac{25}{8} N_o f_a \left(\sqrt{1 + \frac{4}{5}\sqrt{R_T/3}} + 1 \right)^2 . \tag{44b}$$

In addition, as usual, the filter bandwidth equals the signal bandwidth (29), so that

$$B = 2f_a(m+1) = 2f_a \sqrt{1 + \frac{4}{5}\sqrt{R_T/3}} . \tag{45}$$

In order to represent all matched ($B=B_s$) systems in the wide range of modulation indices by one graph, we combine both approximations: (35) for $m > 3$, and (43) for $m < 3$. In our plot the input is represented by r_a, and the output $R(r_a,m)$ by a family of straight lines with m as a parameter (Fig. 7). They are cut off by the threshold locus (36b) or (44b), respectively, for $m \gtrless 3$.

The results, thus far, have been obtained under the assumption that the filters are

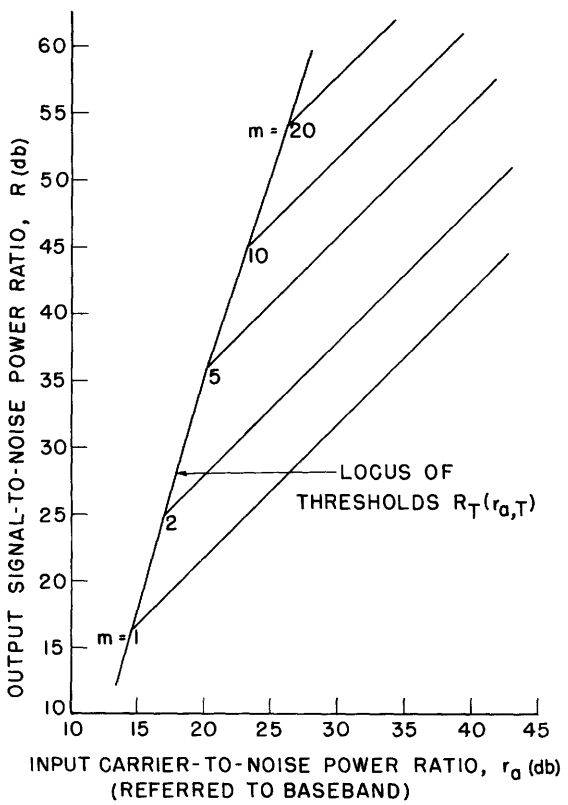

Fig. 7. Performance of matched FM systems.

TABLE 2. Correction factor for noise-power output.

Ratio of Bandwidths $\frac{B}{f_a}$	2	4	6	12	22
Modulation Index m	0	1	2	5	10
Normal-Law Filter	0.67	0.90	0.95	0.988	0.996
Single-Tuned Filter	0.44	0.75	0.86	0.92	0.988

Noise-power output of rectangular filter assumed to be unity for each B/f_a.
Correction factor equal to $(p_a/p_{a_{rect}})^2$.

nearly rectangular. As a final step, we should proceed to the general case for arbitrary filter shape. We have already investigated the noise-power output as influenced by the bandpass filtering (section 2). With the matching of bandwidths, $B_s = B$, still assumed to be valid, we do not introduce any significant changes in the signal-output power. Thus, modification of the gyration radius suffices to provide necessary correction terms for evaluating the output power of Gaussian noise, as well as the output SNR above the threshold.

Table 2 presents some results for cases of practical interest. Note that, in general, the correction factors are not very significant, except for the hypothetical limit $\frac{B}{f_a} = 2$, which corresponds to $m \to 0$. Therefore, most conventional systems can be analyzed in accordance with (40), (41), (42), (43), and/or Fig. 7. For FM systems with a single-tuned bandpass filter, a check with the aid of Table 2 is recommended, since the correction factor in some cases of small modulation index may be important.

PART B: FEEDBACK FM SYSTEMS

7. The Concept of Frequency-Compressive Feedback

The idea of a tracking FM system is not new; it originated, curiously enough, as a substitute for amplitude limiting in an FM receiver.[18] But important noise-combating properties were soon noticed,[19,20] and some successful applications were reported.[21-23] Later, realistic bounds on system performance were introduced[24,14] by taking into account the deficiencies inherently existing in a stable feedback loop.

The most important variety of a tracking FM system is the frequency-compressive feedback loop, which somewhat resembles a well-known feedback frequency-dividing scheme. In this report our interest will be strictly confined to the frequency-compressive FM demodulator as a device for decreasing the noise threshold in an FM system. Its mechanism is capable of reducing the bandwidth of threshold-causing noise without affecting the signal transmission and the output signal-to-noise power ratio. Thus an exchange between the power and the bandwidth of the input FM signal, or an increase in the usable range of a long-distance communication system with fixed signal power, is possible.

The frequency-compressive feedback loop of the usual type (see Fig. 8) has been looked upon either as a "frequency demodulator,"[24,14,25] or an "FM receiving system."[22,26,27] Less customary varieties have sometimes been mentioned in the literature, for instance, a loop without any amplitude limiter[18] or a loop driven from an amplitude limiter.[14] But these have been reported as being substantially inferior to their conventional counterpart.

Within the general structure of the frequency-compressive feedback loop (Fig. 8) some modifications are possible with respect to the location and the form of the output filter. It seems that it is most convenient to have the output filter independent of the feedback filter; both can be connected across the output of the frequency demodulator.

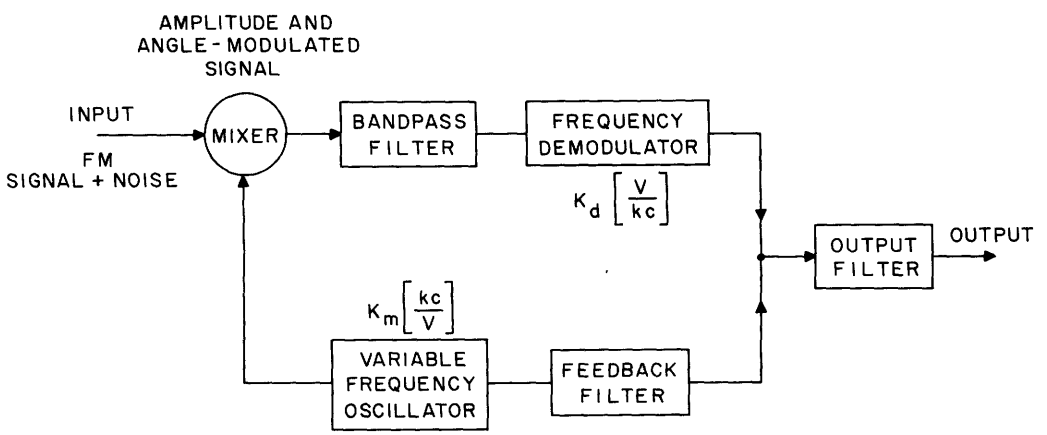

Fig. 8. Block diagram of a frequency-compressive feedback FM system.

Alternative arrangements introduce only a trivial difference in analysis and can be disregarded.

Even with the well-established form of the system under consideration, its analysis presents many essential difficulties. These are due partly to the general complexity of exponential modulation, which is a nonlinear transformation. Also, various regenerative phenomena exist within the loop, and affect its stability and the noise threshold of the system. It is, therefore, easy to understand why the analysis of this system has been less successful than its engineering applications, which were revealed in 1958,[21] 1960,[22] and 1961.[23] Important analytic progress by using linear approximations was initiated in 1962.[24,14]

It seems advisable to point out again that the validity of the linear analysis, in which a baseband analog is utilized, is restricted to cases of approximately linear modulation. Small rms index angle modulation resulting from a useful message or a weak noise disturbance is nearly linear and can be rigorously treated with the aid of the linear model. On the other hand, the linear model is invalid for the analysis of the message distortion in a practical feedback system, because the I-F signal – in spite of the frequency compression – is not small-index-modulated.

One more word of caution might be of use at the beginning of a study of this system. The noise threshold in the frequency-compressive feedback loop may be caused by two entirely different mechanisms: the conventional threshold, and the feedback threshold. Any system as a whole never exhibits more than one of these two noise thresholds. The question arises whether systems should be designed in which the feedback threshold dominates. In our opinion, the answer is no, since in this case it has been found experimentally that anomalies occur as a result of which the system performance is sharply degraded. Furthermore, when the conventional threshold does dominate, the behavior of the feedback system is, in fact, similar to the behavior of the same system with open loop, that is, of a conventional system with a specific type of bandpass filter.

Therefore, an approach to the analysis of feedback systems, whereby the areas of validity of the conventional theory are determined and the feedback threshold is branded as an anomaly to be avoided, is recommended here. This contention is largely backed up by our experiments. Our analysis concentrates on the optimum design of the feedback filter and on the determination of the obtainable power-bandwidth trade-off. However, an important area is still not covered by analysis: i.e., the regeneration of impulsive noise with strong feedback, although we have obtained some interesting experimental evidence regarding this phenomenon.

8. Noise Threshold in the Feedback System

It is now fully recognized that the random noise injected at the input to the frequency-compressive feedback loop (Fig. 8) can affect the system output in two different ways. Let us recall the normal behavior, common to all FM systems whether with or without feedback. We shall consider the case in which the input consists of a carrier wave and

a band of Gaussian noise.

If the feedback system is above its noise threshold with the loop open and certain gain-phase conditions for proper noise tracking are met, then closing the loop will reduce both the output signal and the output noise in the same way, that is, by the compression factor F for the instantaneous values. Therefore, the output signal-to-noise power ratio remains unchanged as the feedback factor F increases from 1 up to a certain limit.

When the input carrier power is diminished, some excess terms of the output noise begin to appear in addition to the linear Gaussian term. Ultimately, the impulse excess noise (which appears as impulsive "clicks") predominates and produces a sharp noise in the noise output: the point at which this occurs is defined as the conventional threshold. This threshold occurs at a specific value of the demodulator input CNR; it depends on the bandwidth B of the bandpass filter, but does not depend upon the feedback factor F (at least not up to a specific value F_{max}; see section 11). On the other hand, since the frequency compression in the feedback system allows us to make the I-F filter narrower than it is in the conventional FM system with the same input signal, a reduction of the threshold is possible. This is the main adavantage of the feedback system.

From the argument above, we shall see that the conventional theory of noise in FM systems, specialized for the case of the single-pole filter with a relatively narrow bandwidth, is sufficient to predict the performance of the feedback system above the conventional threshold. This is obvious for small values of the feedback factor; it will be shown, partly by experimental evidence, that the same type of behavior characterizes the frequency-compressive systems with stronger feedback. This occurs up to a certain value F_{max}, at which there is a rapid break of performance. Very pronounced bursts of impulsive output noise appear and at once degrade the message quality. The mechanism of this feedback threshold is not yet completely understood, but it appears that a cumulative action launches self-regenerating noise in the loop.

As in the conventional case, the precise definition of the threshold as a point rather than as a region is arbitrary: It can be approximately located as a specified departure from the linear input-output relation.

It should be clear from our description that any practical feedback system should not be brought up to the edge of breakdown; that is, the amount of feedback should be bounded by F_{max} in order not to enter the detrimental region of the feedback threshold. In other words, a well-designed frequency-compressive system will not exhibit the closed-loop threshold, since (by design) its conventional threshold predominates, similarly to the situation in a conventional system. Thus, a unified approach to all FM systems appears to be justified.

At present, the bound imposed by the noise regeneration in the feedback loop is known and has been roughly determined by analysis.[24,14] Let us briefly recall Enloe's argument[24] pertaining to the loop of Fig. 8, which is excited by a carrier wave and a band of weak Gaussian noise with spectral power density equal to N_o. After the bandpass filtering, the noise n(t) becomes narrow-band and can be represented by two

orthogonal components, say, $n_c(t)$ and $n_s(t)$.

Enloe's fundamental result,[24] shows that the phase of the resultant wave at the frequency demodulator is

$$\Psi(t) \sim [\hat{n}_s(t) - \hat{\Phi}(t)] - \widehat{n_c(t)\,\Phi(t)}, \tag{46}$$

where $\Phi(t)$ denotes the phase fluctuation (caused by the input noise, and normalized by the carrier amplitude) of the feedback oscillator output. The circumflex stands for the filtering operation in the I-F path (that is, for convolution with the bandpass filter impulse response). Observe that the bracketed term denotes the quadrature noise compressed by the feedback action, and that the last term is excess noise derived from the in-phase component of input noise. Enloe[24] presents arguments substantiating the fact that this last term is the most significant among second-order terms, and that the fluctuation of the envelope of demodulator input wave, which produces additional zero crossings, is negligible in the region of interest.

Under these conditions, it is the mean square of the VCO phase, $\overline{\Phi^2(t)}$, which determines the significance of the additional noise component, and hence the closed-loop threshold. Enloe has observed empirically that threshold occurs when $\overline{\Phi^2} \geq \frac{1}{3.11}\,\text{rad}^2$, which we may approximate by the condition $\overline{\Phi^2} \geq 1/10\ \text{rad}^2$. This is equivalent to the condition that the normalized mean square of the oscillator phase (caused by noise) is no longer small compared with unity. Enloe succeeded in translating this condition into a corresponding carrier-to-noise power ratio at the mixer input, whereby the amount of noise is determined by the closed-loop noise bandwidth. This makes the feedback threshold dependent on the feedback factor. Thus, the noise threshold of the system can be degraded by the feedback action. The primary objective of our analysis is to describe the different consequences – positive as well as negative – of feedback in the loop.

9. Analysis of the Feedback Threshold

The linear analysis of a compressive-feedback FM system is based upon a conventionally constructed baseband analog. Transforming the system in Fig. 8, we make the following substitutions:

A phase and/or frequency subtracter replaces the mixer;

A lowpass (analog) filter replaces the bandpass filter;

A differentiator replaces the frequency demodulator; and

An integrator replaces the variable oscillator.

We assume that our original FM signal is represented by its instantaneous phase as a time function, and that the input Gaussian noise is already narrow-band filtered. Thus it can be represented by its in-phase and quadrature components referred to the carrier frequency f_c; that is

$$n(t) = n_c(t)\cos\omega_c t - n_s(t)\sin\omega_c t. \tag{47}$$

The linear baseband model shown in Fig. 9 is valid when the composite angle modulation (by message, noise or both) is characterized by a small modulation index (rms). This enables us to represent an exponentially modulated wave by the modulating wave (cf. Eq. 5 in Part A). Specifically, we can use the linear model for the analysis of the weak noise disturbance and for determining the border of the linear weak noise region.

Following Enloe's argument, we shall write the open-loop transfer function of the model in Fig. 9 as

$$H_o(s) = K_d K_m A(j\omega)\, H(j\omega), \tag{48}$$

and then define the closed-loop transfer function for the transmission from signal input to the oscillator output

$$H_c(s) = \frac{H_o(s)}{1 + H_o(s)}. \tag{49}$$

Note that this will reduce to the usual input-output closed-loop transfer function, as defined in linear feedback theory, only if the feedback filter is also used as the output filter.

The feedback factor F, as usual, is related to the loop gain at the lowest frequencies; then with $A(0) \equiv H(0) \equiv 1$ we obtain

$$F \triangleq 1 + H_o(0) = 1 + K_d K_m, \tag{50}$$

and we immediately find that

$$H_c(0) = \frac{F - 1}{F}. \tag{51}$$

The two-sided noise bandwidth B_c of the closed loop is defined in the usual way, with the squared magnitude at the origin taken as the reference level:

$$B_c = \frac{1/2\pi j}{|H_c(0)|^2} \int_{-j\infty}^{j\infty} H_c(s)\, H_c(-s)\, ds. \tag{52}$$

Substituting (50) and (51) in (52), we have

$$B_c = \left(\frac{F}{F-1}\right)^2 \int_{-\infty}^{\infty} \left| \frac{K_d K_m A(j\omega)\, H(j\omega)}{1 + K_d K_m A(j\omega)\, H(j\omega)} \right|^2 df. \tag{53}$$

The main finding of Enloe determines C_F, the carrier power at the mixer input below which the noise regeneration in the loop begins to degrade performance:

$$C_F = \frac{N_o B_c}{2\overline{\Phi^2}} \left(\frac{F-1}{F}\right)^2. \tag{54}$$

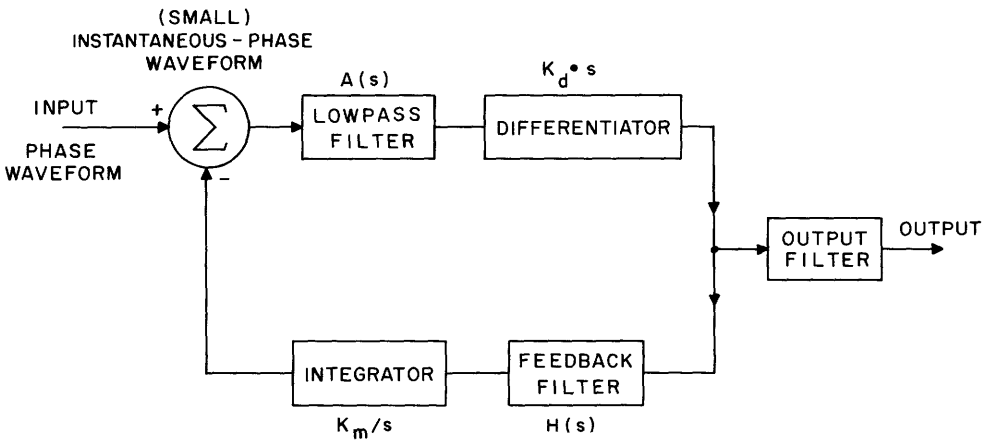

Fig. 9. Linear baseband analog of the frequency-compressive feedback system.

We have found it convenient to use the defining value $\overline{\Phi^2} = 0.1$, and to refer the input carrier-to-noise power ratio at this "feedback threshold" to the baseband width f_a. Then obviously

$$(CNR)_{f_a} \triangleq \rho_a = \frac{C_F}{N_o f_a}, \tag{55}$$

or

$$\boxed{\rho_a = 5\left(\frac{F-1}{F}\right)^2 \frac{B_c}{f_a}}. \tag{56}$$

Equation 56 locates the feedback threshold for a fictitious system without any conventional threshold. The structure of this expression resembles somewhat the structure of Eq. 32 in Part A.

In every real system, of course, the threshold occurs because of the mechanism that requires the higher value of comparable input CNR, measured at the same point and in the same bandwidth. With the linear model, transfer of the CNR from mixer input to demodulator input is straightforward, provided that the filtering is taken into account and a fixed reference bandwidth (preferably f_a) is conserved. Under these conditions, we can compare the effects of the two threshold-causing mechanisms on a frequency-compressive feedback system.

We have shown (section 5) that the conventional threshold of an FM system $r_{a,T}$, defined in terms of the demodulator input CNR in the baseband, depends uniquely on the noise bandwidth of the (broader) bandpass filter, B, normalized by the baseband width. On the other hand, the feedback threshold ρ_a depends primarily on the feedback factor F, its dependence on B being of quite secondary importance. These very slightly interrelated effects can be visualized by referring to Fig. 10, which shows also the crossover

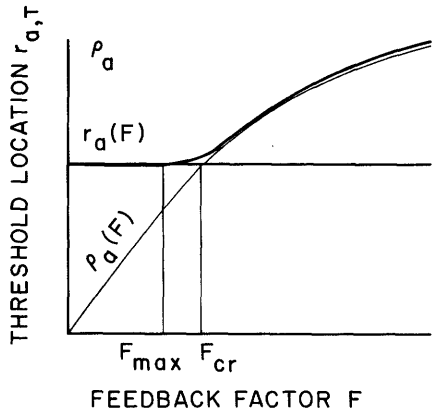

Fig. 10. Threshold location as a function of the feedback factor.

point defined by the value F_{cr}. The interrelation, if any, is expected to occur in the vicinity of this crossover.

The actual noise threshold in the system is found experimentally to be nearly independent of the feedback factor up to a value $F_{max} < F_{cr}$. From F_{max} on, the feedback begins to degrade the actual threshold; with $F = F_{cr}$ the system threshold is exclusively controlled by the excess closed loop noise that causes bursts of clicks in the output. We consider this threshold to be different from the normal behavior of FM systems, and much evidence for this claim will be shown. It was also predicted by an interesting analysis of Baghdady[14] (see his Eq. 111).

The main implication of this argument is a recommendation to design and operate the feedback system in the region $F \leq F_{max}$ where the threshold is not degraded by feedback. In order to establish a proper setting for $F_{max} = c \cdot F_{cr}$, the crossover F_{cr} of the two functions ρ_a and r_a must be examined. Of particular importance is the dependence of the feedback threshold on the parameters of the feedback filter, and this will be discussed first.

10. Analysis and Catalog of Feedback Filters

It has been concluded by independent investigators[14,27,28] that the bandpass filter in the feedback loop cannot have more than one pole, for stability reasons. In order to obtain a stable feedback loop with finite noise bandwidth, it is then necessary to interrelate the number of poles and the number of zeros in the transfer function $H_o(s)$, for the loop consisting of the two filters (shown in Figs. 8 and 9). In compliance with linear feedback theory, we can envisage two alternatives: the number of poles exceeds the number of zeros by one (this gives a better stability margin), or by two (this still gives stability, although with a smaller margin). Here, the chain of broadband amplifiers is not taken into account, mainly because it really should be avoided inside of the loop: otherwise, the parasitic time delay tends to destroy the stability of the system and proper noise tracking. Accordingly, it seems desirable to have high gain in the stages before the loop mixer.

Nonsingular zeros are not desirable in the bandpass filter. As for the low-pass feedback filter, its zeros play an important role either in canceling the bandpass filter pole (the "canceling" zero) or stabilizing the feedback loop without increasing its noise bandwidth (the "stabilizing" zero). Practical considerations with respect to parasitic circuit elements show that the accuracy of the cancellation is always poor, and that attempts to cancel usually cause the appearance of some spurious time delays, with the deleterious effect just discussed.

It seems, therefore, reasonable to recommend no more than one canceling zero; thus the total number of poles around the loop seems to be restricted (in actual engineering implementation) to no more than three. Accordingly, the following structures of the feedback filter can be envisaged: 1 pole; 1 pole and 1 zero (stabilizing); 2 poles and 1 zero (canceling); 2 poles, 1 canceling zero and 1 stabilizing zero.

We note that all of these cases can be described by one general type of the open-loop transfer function:

$$H_o(s) = (F-1) \times \frac{a}{s+a} \times \frac{b}{s+b} \times \frac{s+c}{c}, \tag{57}$$

where a is the complex frequency of a pole in the feedback filter, b is the complex frequency of a pole in the bandpass filter analog or in the feedback filter, and c is the complex frequency of the stabilizing zero in the feedback filter. (There is no need to account for the canceling zero located at c_{IF}, as long as it coincides with the pole of the I-F filter.) In practical calculations it may be convenient to normalize a, b, c, as well as s, the baseband width $\omega_a = 2\pi f_a$.

For the closed loop, we consider the transfer function (49) between the two inputs of the mixer:

$$H_c(s) = \frac{H_o(s)}{1 + H_o(s)}.$$

With some algebraic manipulation we obtain from (57) the general expression for the closed-loop transfer function:

$$H_c(s) = \frac{ab(F-1)(s+c)/c}{s^2 + \left[a + b + \frac{ab(F-1)}{c}\right]s + abF}. \tag{58}$$

The closed-loop noise bandwidth B_c defined by (53) is presented below in a form that stresses the influence of the zero in the feedback filter, and is easily reducible to the no-zero case:

$$B_c = \frac{ab}{2(a+b)} F \times \frac{c^2 + abF}{c^2 + \frac{abc}{a+b}(F-1)}. \tag{59}$$

It is easy to see from the denominator of (59) that with

$$c > \frac{F}{F-1}(a+b)$$

the zero added in the loop for stability can also diminish the closed-loop noise bandwidth, and thus reduce the feedback threshold. More interestingly, it appears that for any given value of the feedback factor F there is one position c_{opt} of the stabilizing zero which produces a minimum in the noise bandwidth.

Differentiating (59) with respect to c, we find this optimum value to be

$$c_{opt} = \frac{F}{F-1}\left[a+b+\sqrt{a^2+b^2+ab\left(F+\frac{1}{F}\right)}\right]. \tag{60}$$

The corresponding minimum of B_c is

$$B_{c_{min}} = \frac{abF^2}{(F-1)c_{opt}}$$

$$= \frac{abF}{a+b+\sqrt{a^2+b^2+ab\left(F+\frac{1}{F}\right)}}. \tag{61}$$

This simple but general result is of considerable interest in feedback-system synthesis. Observe that a minimum always exists for any choice of a and b in the left half-plane. The independent choice of a and b is secured only with the "2 poles and 2 zeros" structure of the feedback filter, so this class is expected to be the most promising one.

Further investigation of the minimum-noise-bandwidth feedback loop can be carried out by perturbing $\frac{a}{b}$ for a fixed value of F and with an optimum zero located by (60). In this optimizing procedure, necessary constraints must be established from bandwidth consideration of the open loop, which has to provide for nearly uniform feedback over the entire baseband.

Let us consider a family of different two-pole open loops (57) with the optimum zero (60), requiring first that the open-loop noise bandwidth be kept fixed. If the closed-loop noise bandwidth (61) is minimized as a function of $\frac{a}{b}$, an optimum value of $\frac{a}{b} = 1$ can be found. This defines the so-called binomial filter: its simplest realization calls for two cascaded, real poles a = b.

For the binomial filter with (negative) real poles we obtain

$$c_{opt} = a\frac{F+\sqrt{F}}{\sqrt{F}-1} \tag{62}$$

and

$$B_{c_{min}} = a \frac{F\sqrt{F}}{(\sqrt{F}+1)^2}. \tag{63}$$

Note that for large values of the feedback factor F the noise bandwidth increases only proportionally to the square root of F; without the (optimum) zero it would increase proportionally to F, as shown in Table 3.

Another possible optimization procedure would call for a fixed half-power bandwidth of the open loop. We then rigorously find that the Butterworth filter with $\frac{a}{b} = j$ exhibits a minimum value of the noise bandwidth. Besides being maximally flat in frequency response, this filter also has the remarkable property of smallest possible noise bandwidth with a second-order transfer function of fixed half-power bandwidth (in particular, smaller than the noise bandwidth of a binomial filter with equal half-power bandwidth).

The second-order Butterworth filter has two conjugate poles:

$$a = \frac{A}{\sqrt{2}}(1+j) \qquad b = \frac{A}{\sqrt{2}}(1-j).$$

It then follows that

$$c_{opt} = \frac{AF}{F-1}\left(\sqrt{F+\frac{1}{F}}+\sqrt{2}\right) \tag{64}$$

and

$$B_{c_{min}} = \frac{AF\sqrt{F}}{\sqrt{F^2+1}+\sqrt{2F}}. \tag{65}$$

Again, for large F the noise bandwidth increases proportionally to $\sqrt{F}$.

There is considerable numerical evidence for the superiority of the Butterworth filter over the binomial filter in the feedback path. This writer, for instance, has evaluated the frequency compression of the FM signal, under the assumption that the post mixer rms modulation index remains smaller than unity. The message compression factor at the highest modulation frequency f_a was compared for the two above-mentioned filter types with optimum stabilizing zero and equal noise bandwidth. The difference was immaterial for lower values of feedback factor F; however, the advantages of the Butterworth filter were clearly shown for $F \geq 5$. They are attributed mainly to the fact that the optimum stabilizing zero for the Butterworth filter, as located by Eq. 64, is much closer to the poles than for the binomial filter (Eq. 62). Therefore, it seems justified to claim that the maximally flat feedback-filter response leads to the minimum of the closed-loop bandwidth.

We thus consider our survey of feedback-filter types to be completed and list the appropriate entries in Table 3.

11. Synthesis of the Feedback System

For the synthesis of a feedback FM system we have to specify the quantities that we shall consider as given: the message bandwidth f_a, the lowest value of output

signal-to-noise ratio R_T, and the noise figure or the noise power spectral density N_o. Then the following characteristics of the system are still to be found: the (lowest) carrier power C_T, the signal modulation index m, the feedback factor F, and the bandwidth B of the I-F filter. Also, after some additional assumptions have been made, the structure of the feedback filter and its parameters have to be chosen.

For practical calculations it is convenient to establish the input CNR as a measure of the carrier power C_T; we will use the value at threshold, referred to the baseband width

$$r_{a,T} = \frac{C_T}{N_o f_a}. \tag{66}$$

A very general method of synthesis results if one assumes that the linear model is valid for the (not narrow-band) signal transmission, as well as for the noise filtering.[29] One obtains two threshold equations in terms of the closed-loop transfer function, under the constraint that the bandpass filter width is matched to the (not fully compressed) I-F signal at the highest baseband frequency. After $H_c(j\omega_a)$ has been found, we next determine the open-loop transfer function, break it down into two factors, and synthesize the feedback filter from its transfer function.

A more careful approach, however, restricts the validity of the linear model (even for the compressed post-mixer signal) to the evaluation of weak noise-processing only.

Table 3. Catalog of feedback-filter structures.

Case	Filter Type	Filter Structure: Poles	Filter Structure: Zeros	Closed-Loop Noise Bandwidth	Remarks
1	1p	a	—	$\frac{abF}{2(a+b)}$	General Expressions
2	1p – 1z(st.)	a	c	$\frac{abF}{a(a+b)} \times \frac{c^2 + abF}{c^2 + \frac{abc}{a+b}(F-1)}$	General Expressions
3	2p – 1z(c) binomial	$a_1 = a_2 = a$	c_{IF}	$\frac{aF}{4}$	Noise bandwidth minimized
4	2p – 1z(c) Butterworth	$a_1 = \frac{A}{\sqrt{2}}(1+j)$ $a_1 = \frac{A}{\sqrt{2}}(1-j)$	c_{IF}	$\frac{AF}{2\sqrt{2}}$	Noise bandwidth minimized
5	2p – 2z binomial	$a_1 = a_2 = a$	c_{opt} c_{IF}	$\frac{aF\sqrt{F}}{(1+\sqrt{F})^2}$	Noise bandwidth minimized
6	2p – 2z Butterworth	$a_1 = \frac{A}{\sqrt{2}}(1+j)$ $a_2 = \frac{A}{\sqrt{2}}(1-j)$	c_{opt} c_{IF}	$\frac{AF\sqrt{F}}{\sqrt{F^2+1}+\sqrt{2F}}$	Noise bandwidth minimized

Then the synthesis procedure must rely upon some previous knowledge of the closed-loop transfer function, as specified by its poles and zeros location (58). Just as we have no freedom to change the structure of the bandpass filter, the possible selection of feedback filter types is also restricted. Thus, with a catalog of the allowable filters in mind, we should not deviate far from the best Butterworth structure known (the second-order Butterworth, cf. Table 3).

In order to determine the main parameters of both the bandpass and feedback filters, we must meet two requirements, which seem to be generally agreed upon.[24,29-31]

(a) the open-loop transfer function should be essentially uniform over the baseband in order to have the post-mixer frequency deviation evenly compressed for all modulation frequencies;

(b) the closed-loop noise bandwidth should be minimized (as explained in Sec. 9) in order to reduce the excess-loop noise that causes the feedback threshold.

One more relation is needed to fully specify the problem, namely, the relative position of the feedback threshold. In Enloe's[28] approach, it should coincide with the conventional threshold. That is, the bandpass filter is required to be wide enough to pass the compressed I-F signal without, however, allowing the conventional threshold to predominate. As we have pointed out, we are inclined to modify this condition slightly, and to require that the conventional threshold predominate by 1-2 db.

There is also some justification (cf. Part A) in matching the bandwidth of the I-F signal B_s to the noise bandwidth of the I-F filter B. (Enloe proposes that the half-power filter bandwidth be matched to the full frequency swing $2D_{IF}$; this is somewhat optimistic for narrower I-F signals with lower m/F index.) We shall use again Carson's formula,

$$B_s = 2f_a\left(\frac{m}{F}+1\right), \tag{67}$$

with appropriate account taken of the frequency compression in the I-F path. Denoting the baseband location of the pole in the bandpass filter by b, we have for the noise bandwidth of the single-tuned structure

$$B = \pi b, \tag{68}$$

and the proposed matching yields

$$\pi b = 2f_a\left(\frac{m}{F}+1\right). \tag{69}$$

From Eq. 40, the conventional threshold in low-modulation-index systems is found to occur at

$$r_{a,T} = \frac{25}{32}\left(\frac{B}{f_a}\right)^2\left(2\frac{f_a}{B}+1\right)^2. \tag{70}$$

Combining (70) with (68) and (69), we obtain

$$r_{a,T} = \frac{25}{8}\left(\frac{m}{F}+2\right)^2 \tag{71}$$

In the frequency-compressive feedback system, the output SNR at and above threshold is given by the unmodified relation,[14,27]

$$R = \frac{3}{2}m^2 r_a \tag{72}$$

so that at threshold (using Eq. 71), we have

$$R_T = \frac{75}{16}m^2\left(\frac{m}{F}+2\right)^2. \tag{73}$$

Equation 73 involves the two unknowns m and F; we must therefore introduce our additional constraint upon the location of feedback threshold to solve the synthesis problem. Numerical machine computations or graphical methods described by Enloe for the condition of equal thresholds could certainly be adapted to our assumptions. We are, however, first inclined toward a very simple cut-and-try method, which is based upon some choice of maximum permissible value of the feedback factor, say, F_{max}. This might be determined by previous experience, or by experiments with loop stability.

Assuming a value of $F \leq F_{max}$, we observe from (28) that

$$m\left(\frac{m}{F}+2\right) = \frac{4}{5}\sqrt{R_T/3}. \tag{74}$$

Consequently,

$$m = F\left(\sqrt{1 + \frac{4}{5F}\sqrt{R_T/3}} - 1\right), \tag{75}$$

and it follows that

$$B = 2f_a\sqrt{1 + \frac{4}{5F}\sqrt{R_T/3}} \tag{76}$$

$$r_{a,T} = \frac{25}{8}\left(1 + \sqrt{1 + \frac{4}{5F}\sqrt{R_T/3}}\right)^2. \tag{77}$$

Now, a test is needed in order to ascertain that we do not approach too close to the critical value of F_{cr}; otherwise the system would be vulnerable to regenerative noise, with an intolerable probability of high-level noise spikes. For this purpose, we evaluate the theoretical location of the closed-loop threshold ρ_a and compare it with (77), requiring (for example) that

$$\rho_a \leq 0.8 r_{a,T}. \tag{78}$$

Here, the factor 0.8 is chosen as the highest possible value of a safety margin, leading to the maximum system sensitivity.

If relation (78) is not satisfied, it is necessary to resort to another, lower value of

F and to redesign the system.

If Eq. 78 is satisfied, it is easy to determine the power saving, as compared with the conventional (no-feedback) system. Assume as a constraint that the two systems to be compared have equal performance R_T at the noise threshold. We also assume in both cases that the signal is accommodated within the (respective) noise bandwidth of the bandpass filter. For the conventional system, the "zero-order" approximation in Part A shows that a threshold is attained with the input carrier-to-noise ratio equal to

$$r_{a,o} = 20\sqrt[3]{R_T/30}. \tag{79}$$

The example that follows illustrates the simple sequence of computations outlined above. Assume that a threshold performance of $R_T = 3 \times 10^4$ is required, and that a conventional system is therefore designed with

$$m_o = \sqrt[3]{\frac{R_T}{30}} = \sqrt[3]{\frac{3 \times 10^4}{30}} = 10$$

$$r_{a,o} = 20m_o = 20 \times 10 = 200.$$

Assume next that for the feedback system, $F = 10$ is known as a value not exceeding the condition for loop stability. Then we compute

$$m = 10\left(\sqrt{1 + \frac{4}{5 \times 10}\sqrt{\frac{3 \times 10^4}{3}} - 1}\right) = 20,$$

and with $m_{IF} = \frac{m}{F} = \frac{20}{10} = 2$ we have

$$B = 2f_a\left(\frac{m}{F}+1\right) = 6f_a.$$

Finally,

$$r_{a,T} = \frac{25}{8}\left(1 + \sqrt{1 + \frac{4}{5F}\sqrt{R_T/3}}\right)^2 = \frac{25}{8}\left(1 + \sqrt{1 + \frac{4}{5 \times 10}\sqrt{10^4}}\right)^2 = 50.$$

It seems that the power saving amounts to 6 db (4 times). However, a check for the feedback threshold is necessary. Our choice for the feedback filter will be the optimal Butterworth structure; with $A = 2\pi f_a$ and B_c computed from Eq. 65. Then

$$\rho_a \approx \frac{5 \times 2\pi F}{\sqrt{F + \frac{1}{F} + \sqrt{2}}} = \frac{10\pi \times 10}{3.18 + 1.41} = 68 > r_{a,T}.$$

Thus the threshold in the system would be caused by the predominating effect of noise regeneration in the feedback loop. Accordingly, we have to reduce the feedback factor, and design for a smaller value, say, $F = 7$. We obtain

$$m = 17.7$$

$$\rho_a = 53.7$$

$$r_{a,T} = 66.5.$$

Observe that the necessary closed-loop threshold relation (78) is now preserved.

For this corrected design we have approximately

$$\frac{r_{a,o}}{r_{a,T}} = \frac{200}{66.5} \approx 3$$

and the threshold extension by a factor of three (4.8 db) is obtained. It occurs at the expense of an increase in signal bandwidth which equals

$$\frac{1 + m}{1 + m_o} = \frac{1 + 17.7}{1 + 10} = 1.70,$$

or 70 per cent.

This example shows the simplicity of the proposed cut-and-try synthesis method. We shall, however, also investigate the feasibility of a straightforward synthesis of the system with optimum power-bandwidth trade-off.

12. Optimization of the Feedback FM System

Apart from the cut-and-try synthesis procedure, it is interesting to examine quantitatively the limitations imposed on the feedback system by the requirement of loop and noise stability. The main objective here will be the determination of the maximum allowable value of the feedback factor, and also the maximum obtainable power saving in comparison with a conventional FM system.

If we restrict our attention to the second-order Butterworth structure, we can obtain from Eq. 65 a single formula for the closed-loop threshold in terms of ρ_a, the mixer input CNR referred to the baseband width:

$$\rho_a = 10\pi \frac{(F-1)^2}{\sqrt{F^3 + F} + \sqrt{2F^2}}. \tag{80}$$

For the conventional threshold, measured in the same way, from (70), we have

$$r_{a,T} = \frac{25}{8}\left(\frac{m}{F} + 2\right)^2.$$

In accordance with (78), we feel that the maximum system sensitivity can be obtained with

$$0.8 r_{a,T} = \rho_a. \tag{81}$$

This yields the condition

$$2.5\left(\frac{m}{F}+2\right)^2 = 10\pi \frac{(F-1)^2}{\sqrt{F^3 + F + \sqrt{2F^2}}}, \tag{82}$$

wherefrom we obtain

$$m = 2\sqrt{\pi} \frac{(F-1)\cdot F}{\sqrt{\sqrt{F^3 + F + \sqrt{2F^2}}}} - 2F. \tag{83}$$

Equation 83 expresses m in terms of F, whereas our objective is to determine $m(R_T)$ and $F(R_T)$. Combining (74) and (83), we obtain the implicit relation

$$\sqrt{1 + \frac{4}{5F}\sqrt{R_T/3}} - F = 2\sqrt{\pi} \frac{F^2 - F}{\sqrt{\sqrt{F^3 + F + \sqrt{2F^2}}}} - 2F. \tag{84}$$

The optimization procedure can now be stated as follows: Given R_T, solve (84) for F and substitute it in (74) to obtain m.

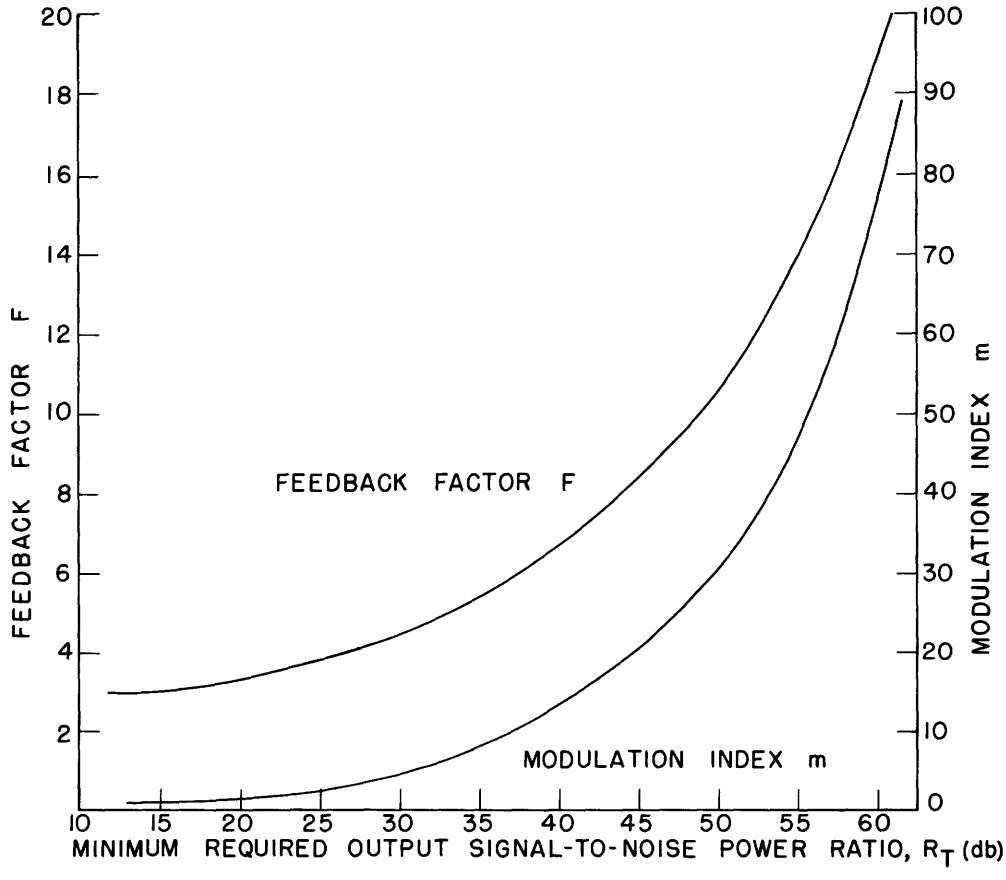

Fig. 11. Optimum parameters of the feedback FM system.

After numerical evaluation in the region of interest, we obtain the two relations shown in Fig. 11. It is now easy to observe that m/F in an optimum system does not vary much with R_T. It follows that the signal bandwidth B_s of Eq. 67, when equated to the I-F bandwidth B, yields

$$B = 2f_a\left(\frac{m}{F}+1\right) \tag{85}$$

which is also fairly insensitive to R_T.

It is of interest to evaluate – with optimal values of m and F – the threshold location $r_{a,T}$ and compare it with that of the conventional system, under the constraint of a fixed R_T. This is shown in Fig. 12, from which the power saving can be found directly in db.

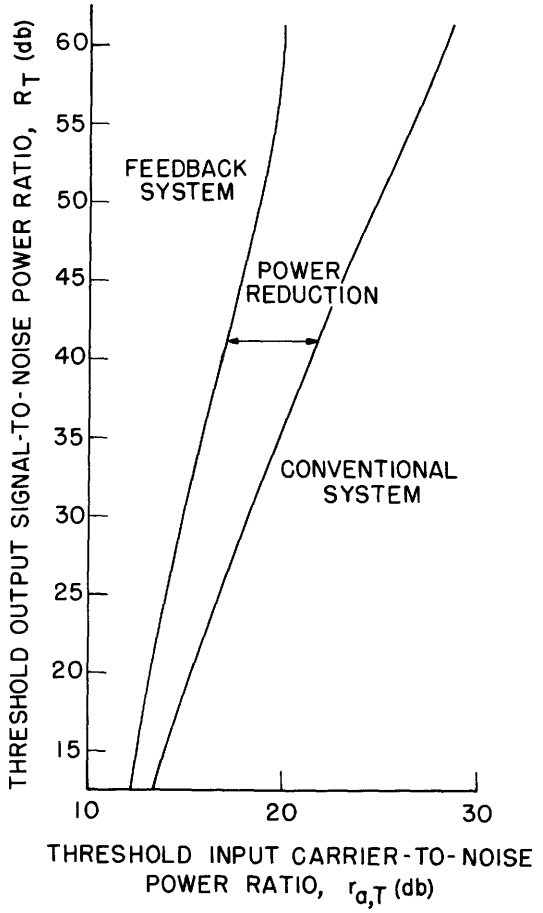

Fig. 12. Comparison of a feedback FM system with the conventional FM system.

The figures differ somewhat from those obtained with Enloe's synthesis procedure.

Finally, since the power consumption (proportional to r_a) is traded off against the signal bandwidth

$$B_s = 2f_a(m+1), \tag{86}$$

the efficiency of this trade-off might be of interest. In Fig. 13 we plot

$$\frac{r_{a,o} \cdot B_{s,o}}{r_{a,F} \cdot B_{s,F}}$$

as a function of R_T, for the sake of comparison of an optimal feedback system (indicated by the subscript "F") with a properly designed conventional system (indicated by the

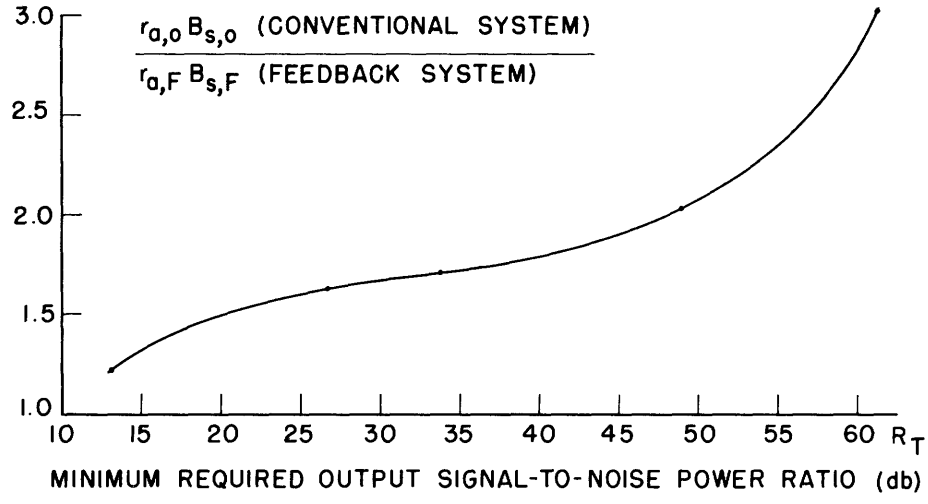

Fig. 13. Power-bandwidth trade-off in an optimum feedback FM system.

subscript "o"). It is interesting to point out that the last relation runs relatively flat, showing not much change in the efficiency of power-bandwidth exchange.

13. Experimental Program and Apparatus

We shall now report on the experimental research on feedback FM systems which was undertaken in order to check some analytical inferences, to examine some properties of the systems not amenable to analysis, and to verify some assumptions made in connection with the synthesis procedure.

Because of time limitations, it was not feasible to design and construct an optimized feedback system and thereby investigate directly the maximum obtainable amount of power-bandwidth trade-off. Nevertheless, the experiment contributed to an understanding of the system behavior, and resulted in significant modifications of the design procedure.

Much of the instrumentation arrangement originated in earlier research by W. Wilson.[31] The main part of the system under study was composed of a commercial high-quality FM tuner, in which the local oscillator was replaced by an external FM signal generator and one of the I-F filters was shunted by a highly selective single-tuned circuit. Different feedback filters were added between the discriminator output and the modulation input to the signal generator. The output of the system was preceded by an independent variable sharp-cutoff filter. The over-all block diagram of the system is shown in Fig.14.

For the most part, the system was excited with an input composed of a cw (or an FM)

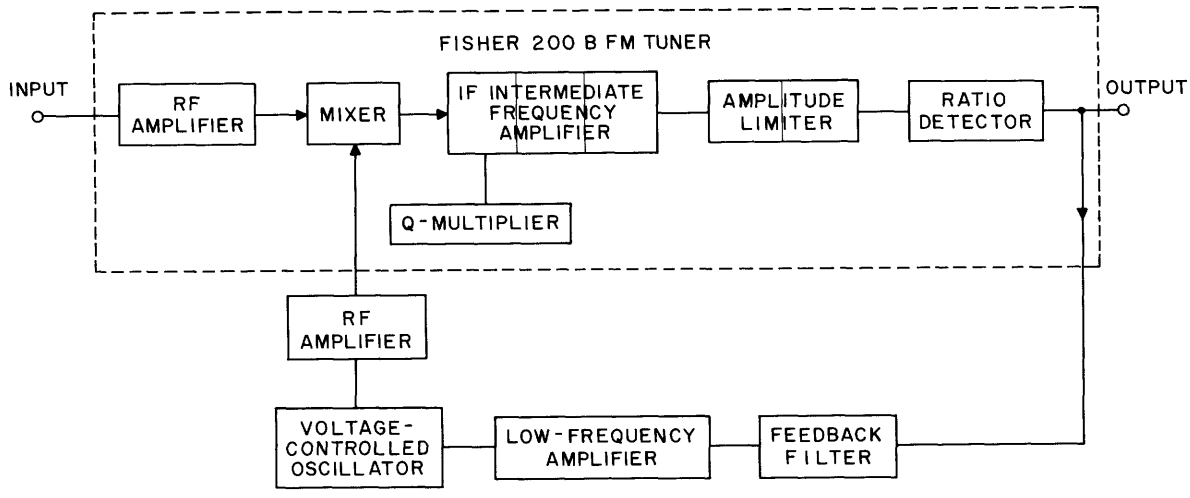

Fig. 14. Block diagram of the experimental feedback FM system.

signal and a strong Gaussian noise added together. A high-quality FM signal generator was used as the signal source, and the noise was generated by a tungsten filament noise diode which was amplified by a wideband amplifier so that a wide band of constant density Gaussian noise was available.

Provision was made for monitoring carrier and noise power before the demodulator input. The frequency response of the single-tuned circuit was also visually monitored by the sweep-generator method. Throughout, it was extremely difficult to obtain precise measurement in the I-F strip, and therefore the resulting data are often presented in relative rather than in absolute form.

The output characteristics of interest were monitored by a switched-beam wideband oscilloscope, and measured by means of a highly sensitive true rms voltmeter. For extraction of the sinusoidal message buried in noise a wave analyzer was used. A medium-speed pulse counter was utilized to count the "clicks" of impulsive noise; the level of counting was controllable over as wide a range as necessary. Figure 15 also shows certain other equipment that is customarily used in experiments of this type.

The feedback loop was constructed by using standard commercial FM equipment wherever possible and, accordingly, did not meet the more stringent requirements of a well-designed feedback system. The primary objective of having enough gain around the loop without much parasitic phase shift was, in fact, barely met. Since the intermediate frequency of the commercial tuner was the standard value, 10.7 kc, and the I-F amplifier was of standard structure, it was necessary to resort to a Q-multiplier in order to obtain a sufficiently narrow single-pole filter. Not much, however, could be done to decrease the parasitic phase shifts by reducing the gain and selectivity of the broadband I-F amplifier and the frequency demodulator: otherwise the output signal level would have been too low not to be corrupted by the residual noise of the low-frequency amplifier at the system output.

Therefore, the amount of feedback was sharply limited to approximately 16 db by the inherent phase shift of the broadband I-F stages and the demodulator inside the loop. Closing of the loop (as expected) reduced the usable baseband to approximately 5 kc for

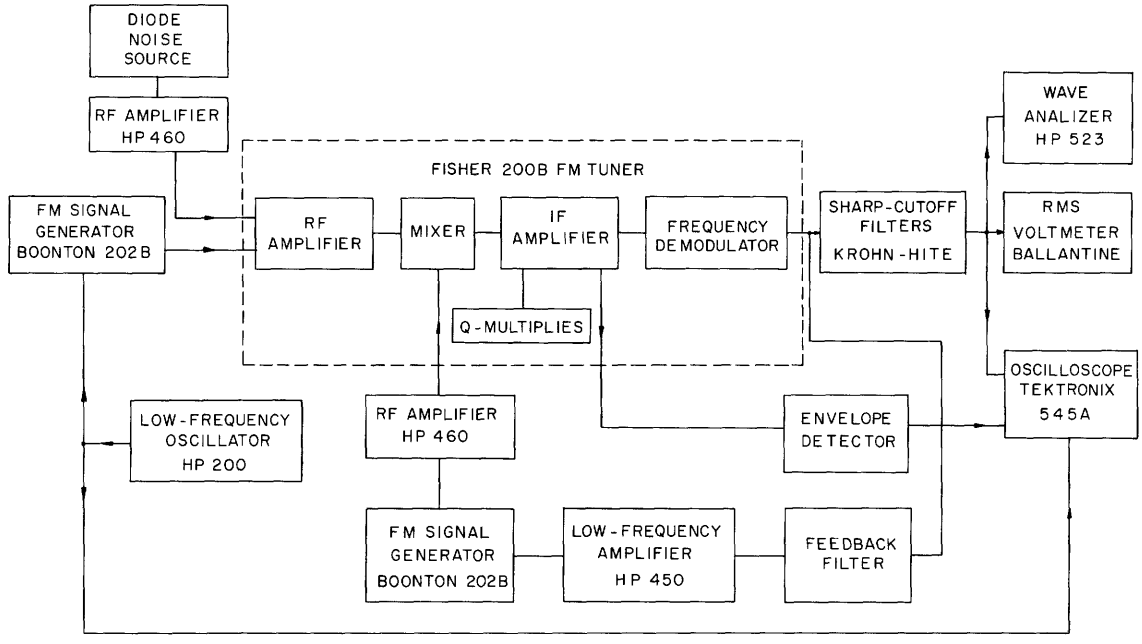

Fig. 15. Measurement and auxiliary apparatus used in these experiments.

stability reasons. The sensible way out of this trouble, of course, would be to construct a double conversion-receiving system, with most of the gain and some selectivity before the loop and with a low-noise I-F amplifier of higher selectivity in the forward path of the loop.

The program of experiments was partly determined by the deficiencies of the system under test. Specifically, it was recognized that the system stability was far from optimum, and not much could be done by a judicious choice of the feedback filter. Some quantitative evidence of the improvement of noise threshold is already available,[23,29] but this certainly could not be matched by standard commercial apparatus, as the earlier research had shown.[31]

On the other hand, even this unsatisfactory apparatus was capable of supplying information on certain specific features of the feedback systems, some of which were not mentioned heretofore in published analyses or experimental data. In particular, it was possible to show distinctly the existence of the feedback threshold, and to investigate the differences of system behavior with the two types of threshold; the postulated similarity of the conventional threshold with the loop open and closed was also evidenced. The system noise threshold as a function of the feedback factor was examined and it checked qualitatively with predictions resulting from our discussion in Section 10 and

Table 4. Parameter adjustments used in the experimental investigation.

	Experiment	Measured output quantity	Parameters of filters					Loop		Parameters of signal		Figure reference
			Bandpass filter		Feedback filter		Output filter	Circuit	Gain factor	Modulation frequency	Frequency deviation	
			$\frac{b}{2\pi}$ [kc]	B [kc]	$\frac{a}{2\pi}$ [kc]	$\frac{c}{2\pi}$ [kc]	f_a [kc]		F [db]	f_m [kc]	D [±kc]	
A	Open – loop systems	noise power N	Double tuned	~200	—	—	5	open	0	—	—	9a
			20	63	—	—	4	open	0	—	—	9b
			8.5	27	—	—	5	open	0	—	—	9c
		signal-to-noise power ratio SNR	Double tuned	~200	—	—	5	open	0	3	24	10a
			20	63	—	—	4	open	0	3	24	10b
			8.5	27	—	—	5	open	0	3	15	10c
B	Conventional threshold with loop closed	SNR	Double tuned	~200	8	—	5	open closed	0 12	3	24	11a
			20	63	0.7	4	4	open closed	0 11	3	24	11b
			8.5	27	8	12	5	open closed	0 10	3	15	11c
		S, N	20	63	0.7	4	4	open closed	0 11	3	24	12
C	Feedback threshold	SNR	8.5	27	8	12	5	open closed	0 16	—	—	13
D	Both types of threshold	N	8.5	27	8	12	5	closed closed	10 16	—	—	14
E	System noise threshold	F for $\nu=1$	8.5	27	8	—	4	closed	≤14	—	—	15
F	Number of noise clicks	ν	8.5	27	8	—	4	open closed	0 14	—	—	16

other analyses.[14,24,28] Finally, some observations were made with respect to the composition of output noise and occurrence of noise clicks.

In spite of the limitations in the experimental apparatus, it was still possible to obtain data confirming the analytical inferences, and to isolate certain aspects that require further investigation.

14. Summary of Experimental Results

After preliminary investigation of the stability of the system, it was found that a stabilizing zero in the feedback filter would be of value. The structure utilized in the experiments was, for the most part, of the "1 pole-1 zero" type (see Table 3). The pole location was determined by the time constant of the lowpass filter in the ratio detector, and was found to be located below 8 kc.

The bandpass I-F single-tuned filter bandwidth was made variable over the range $3 \text{ kc} < \frac{b}{2\pi} < 30 \text{ kc}$, where $b/2\pi$ denotes the (baseband negative real axis) location of its pole. The half-power bandwidth is then b/π, varied in the range between 6 kc and 60 kc. The corresponding noise bandwidth is $B = \frac{b}{2}$, so that we have $9.4 < B < 94$ kc. A reference value of $\frac{b}{2\pi} = 8.5$ kc, with B = 27 kc, was chosen in many of the experiments.

As for the output filter, electronically tuned sharp cutoff units with 24-db/oct slope were used, usually with two of them in tandem. The cutoff frequency corresponded to

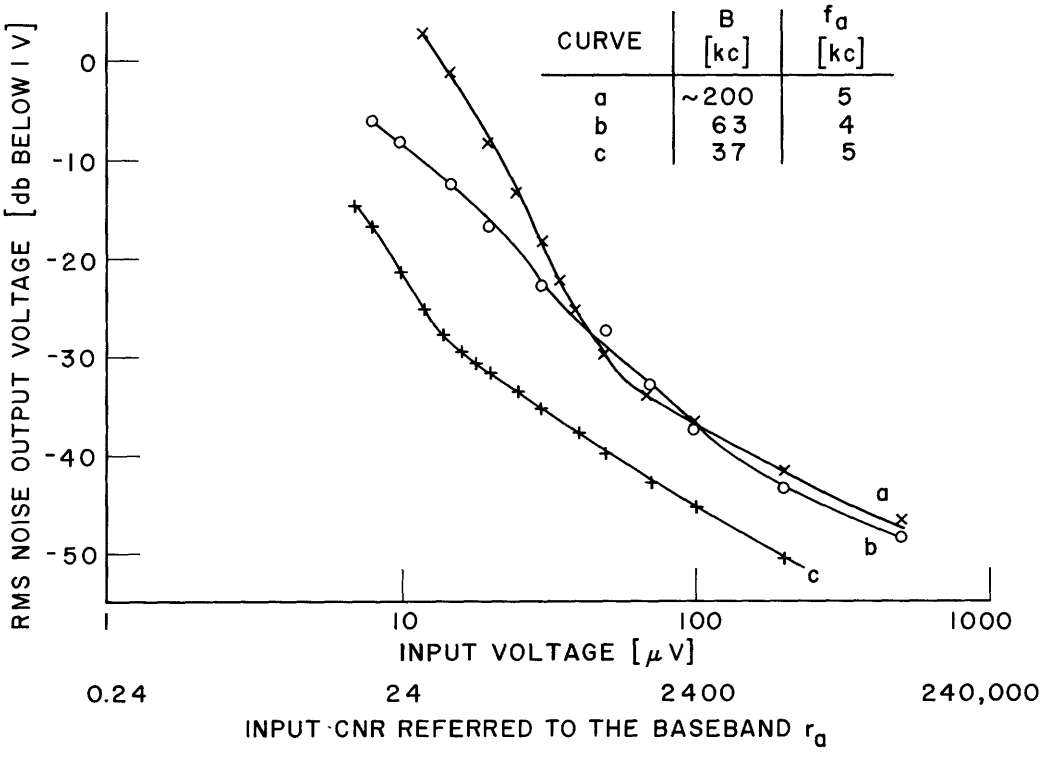

Fig. 16. Noise-quieting curves in open-loop (conventional) systems.

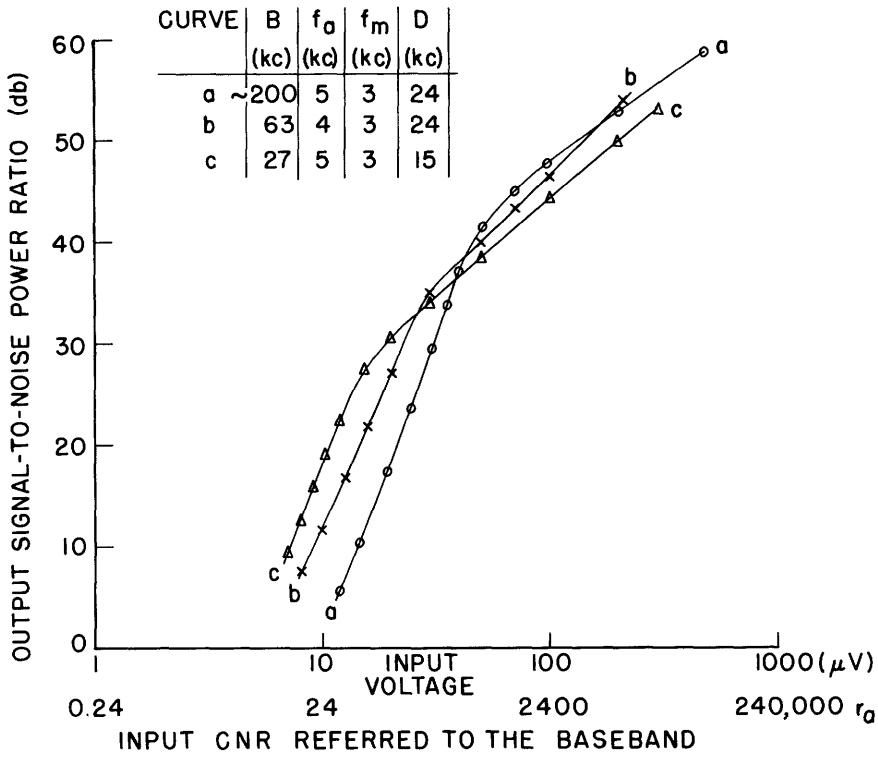

Fig. 17. Conventional threshold in open-loop systems.

the baseband chosen for reference, and was varied from 3 to 5 kc. The general plan of the experiments with corresponding values of parameters is presented in Table 4.

Experiment A. The first series of experiments was supposed to indicate the general behavior of conventional FM systems, both with the original arrangement of broad I-F filters and with the narrow single-tuned filter of variable width. Both the noise-quieting curves (Fig. 16) and the SNR-performance curves (Fig. 17) show a well-pronounced, sharp knee, indicating the conventional threshold. The location of the threshold expressed in terms of the input carrier power, C_T (or the corresponding CNR in a fixed band, $r_{a,T}$) depends, as expected, on the I-F filter bandwidth. This relation is stronger than linear in accordance with the analysis presented in Part A. On the other hand, the threshold location expressed in terms of the demodulator input carrier-to-noise ratio, r_T, should show only a slight dependence on the bandpass-filter width; evidence of this was not clearly observed, probably because of inaccuracies in the determination of the bandwidth, B.

Experiment B. In the second series of experiments, feedback systems exhibiting only the conventional threshold were carefully investigated with the loop open and closed. It was repeatedly shown (with three different combinations of bandpass and feedback filters) that the system performance as measured by the SNR does not depend on the feedback factor (up to a certain value, F_{max}, as expected) (see Fig. 18). On the other hand,

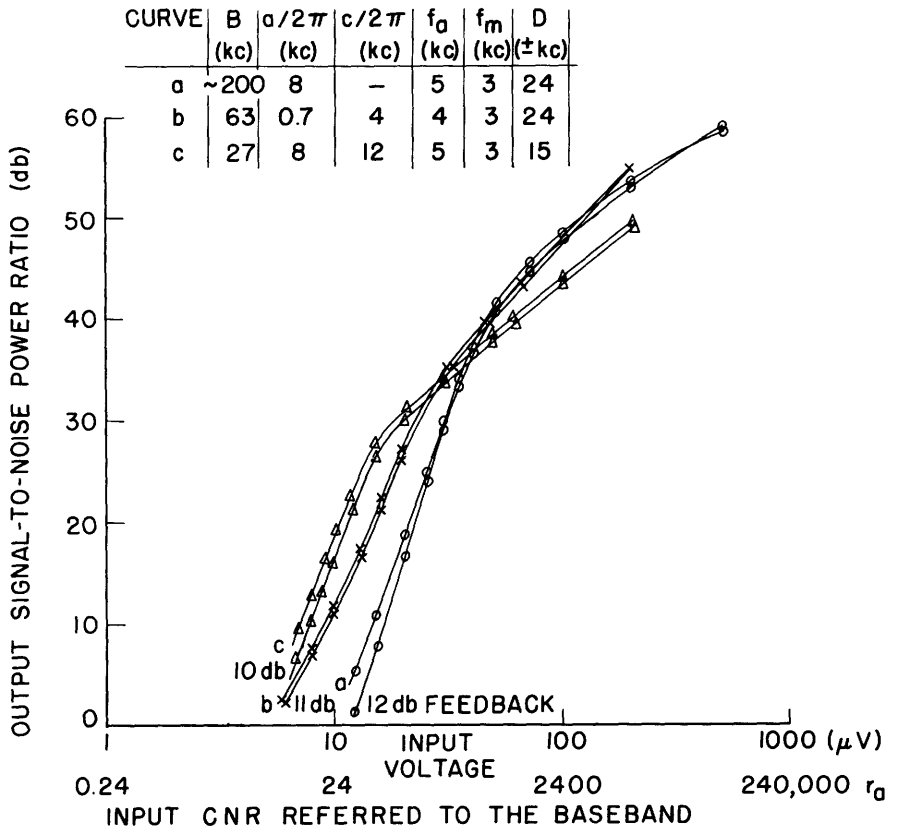

Fig. 18. Conventional threshold with open loop and closed loop.

there is some change of shape below the threshold of both the noise plot and the signal plot after the loop has been closed. One of the examples (Fig. 19) shows interesting evidence for the conjecture of Baghdady,[14] that the feedback augments both the weak signal and the strong noise in the threshold region. Thus, even with the threshold location unshifted and the SNR below conventional threshold unchanged, we can distinguish some differences in the behavior of the system: specifically, the below-threshold signal-to-noise ratio is now formed as a ratio of two quantities, each of which changes faster than it does with the loop open.

Experiment C. Subsequently, the closed-loop or feedback threshold was examined in systems in which it predominated over the conventional threshold. In qualitative agreement with Enloe's results, the system threshold in this case was degraded by the feedback (Fig. 20). The amount of degradation was found to be somewhat larger than expected; this can be attributed to deficiencies in both the analysis (second-order effects neglected) and in the instrumentation (parasitic phase shifts increasing the closed-loop noise bandwidth).

Experiment D. It was clearly observed that the performance breakdown at the feedback threshold is much more abrupt than at the conventional threshold. Thus, the measurement of output noise power with the loop open and closed suffices to determine what type

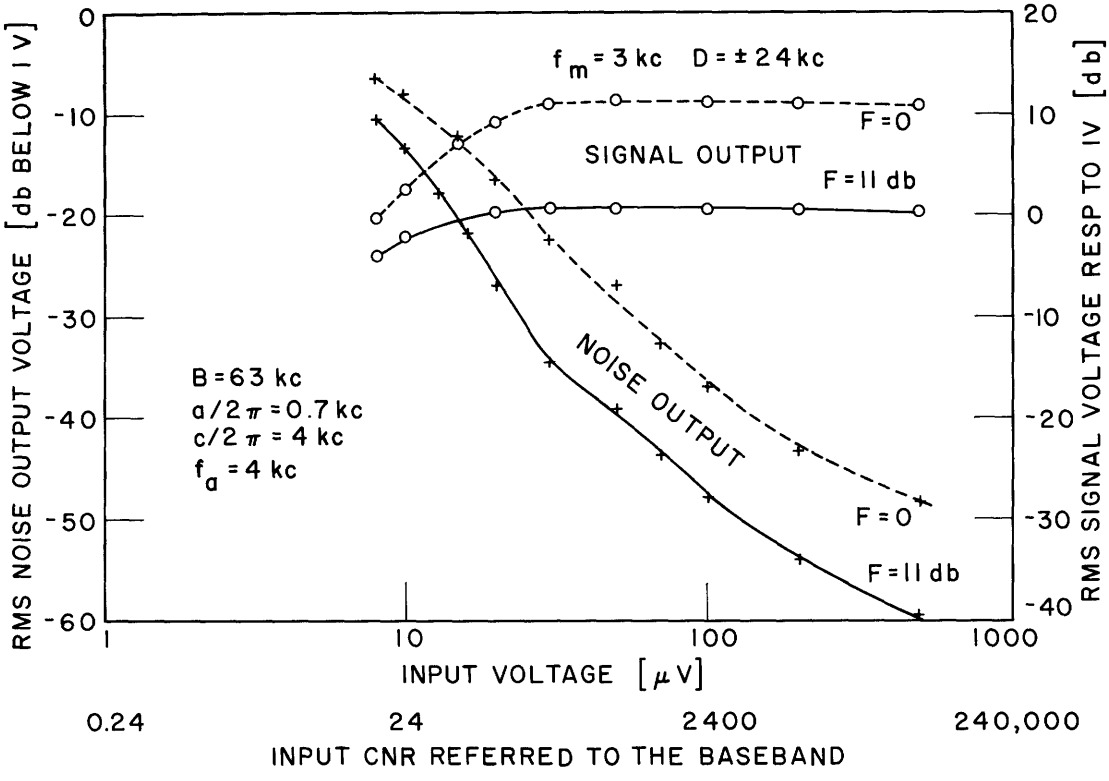

Fig. 19. The power of signal and noise in the threshold region as a function of feedback.

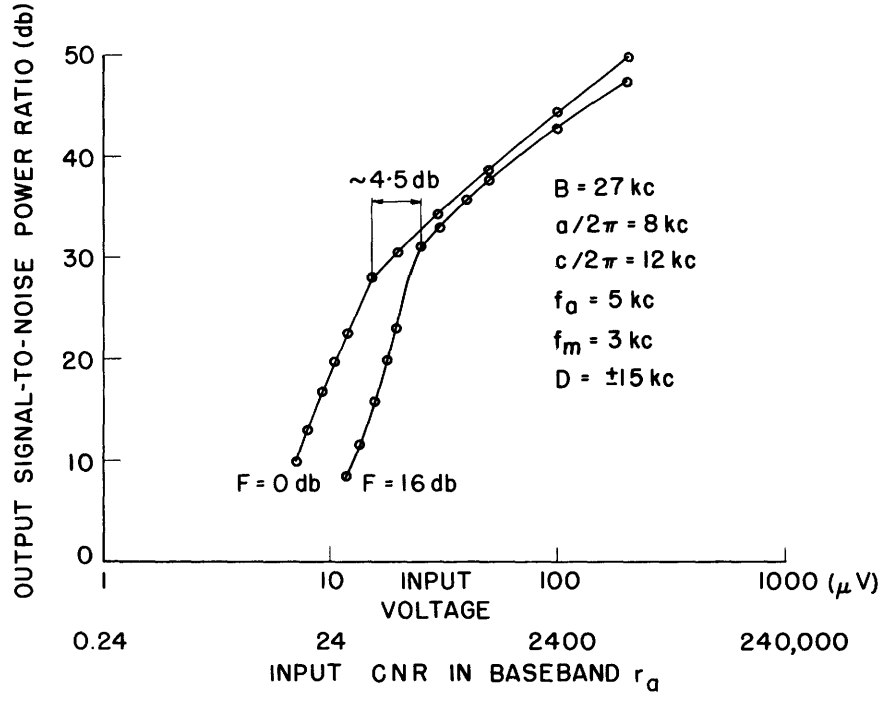

Fig. 20. The degradation of the closed-loop threshold by feedback.

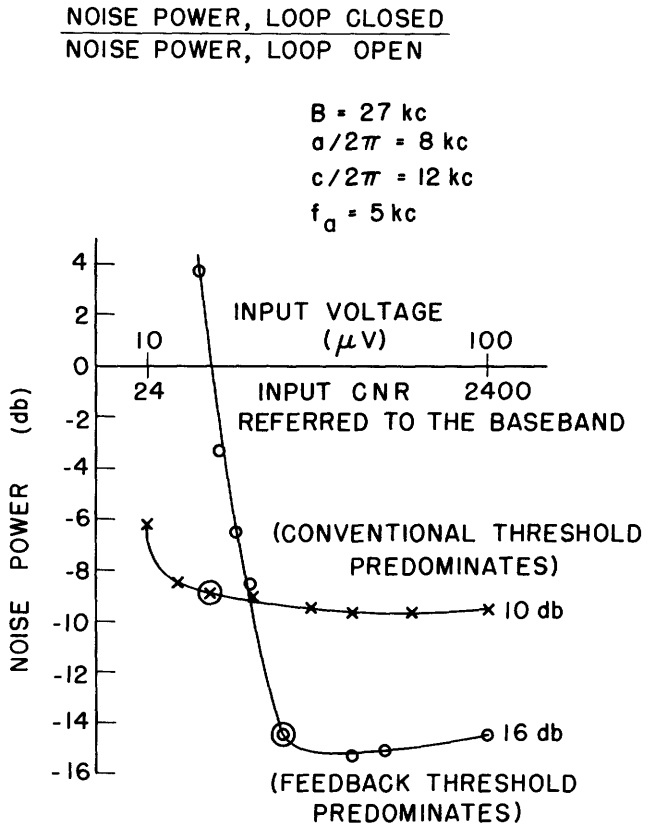

Fig. 21. An experimental method of recognizing the type of threshold in a feedback system. (O denotes the knee on the SNR performance curve.)

of threshold predominates in any feedback system. This is distinctly visualized by referring to Fig. 21, in which the sharp breakdown corresponds to the feedback threshold. The onset of the feedback threshold is also clearly visible when the output noise is monitored on the screen of an oscilloscope. Much larger relative amplitude of the noise spikes as compared with the Gaussian-noise background seems to confirm the conjecture about regeneration of at least one noise component in the loop.

Experiment E. It appears, therefore, that the onset of the impulsive noise is a good indication of closed-loop threshold in the feedback system. In fact, this onset is still sufficient to indicate the location of threshold with a smaller feedback factor, one not large enough to make the feedback threshold predominate. To supplement this observation, we made a rough estimate of the threshold as a function of the feedback factor, where threshold is defined by the occurrence of one noise impulse per second (on the average). No other parameter of the system was changed during this experiment, and, accordingly, we should have expected (and, in fact, observed) two distinct regions: one of substantial independence, and the other of pronounced dependence. Moreover, the transition region is of great interest for the system synthesis, and is also observable in the plot of Fig. 22.

Experiment F. The final part of the experimental program was concerned with the character of noise in the vicinity of the threshold. In order to show quantitative evidence of the character of output noise, we need a time or magnitude measure to define that

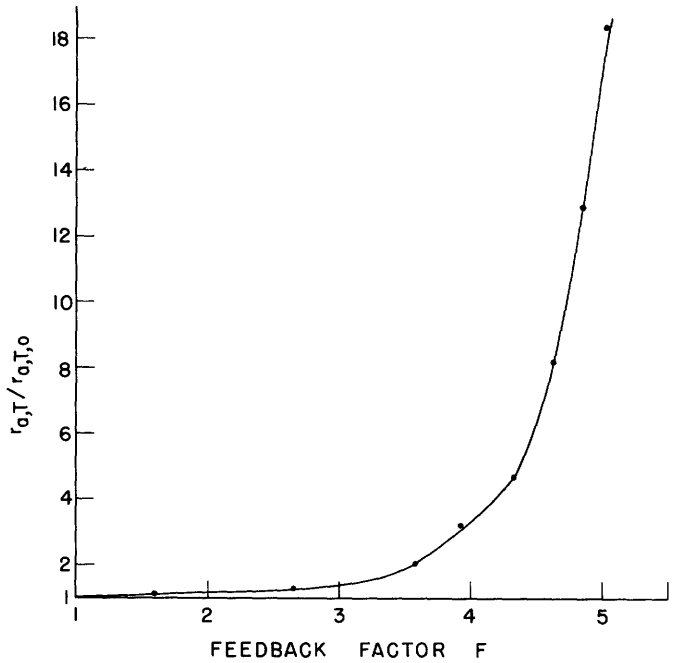

Fig. 22. Dependence of the location of threshold on the feedback factor.

which we choose to call an impulse. A level (amplitude) selection was chosen in order to separate the impulsive component. Since the measurement of the total power of output noise cannot tell the distinction between different types of noise, some measure is also needed to indicate the intensity of the impulsive noise. In accordance with Rice's analysis, the average number of noise clicks per second is a convenient measure for the intensity because this number is directly proportional to the average impulsive noise power: This relation holds, provided that the probability of clicks is small enough to insure the independence of the individual impulses in the train.

A proper amplitude selection was obtained by adjusting the counting level with the counter so that the contribution from the Gaussian and quasi-Gaussian output noise would be negligible. In accordance with the customary procedure, this selection level was related to the standard deviation σ of the Gaussian component, as determined from the true rms voltage measurement. The above-threshold Gaussian output noise level was linearly extrapolated below the threshold and the rms value 3σ was chosen for counting level. Since the speed of the available counter exceeded by far the I-F filter bandwidth, all noise clicks with amplitude above the 3σ level were counted (during a 10-sec interval). In order to indicate the striking difference in the onset of impulsive

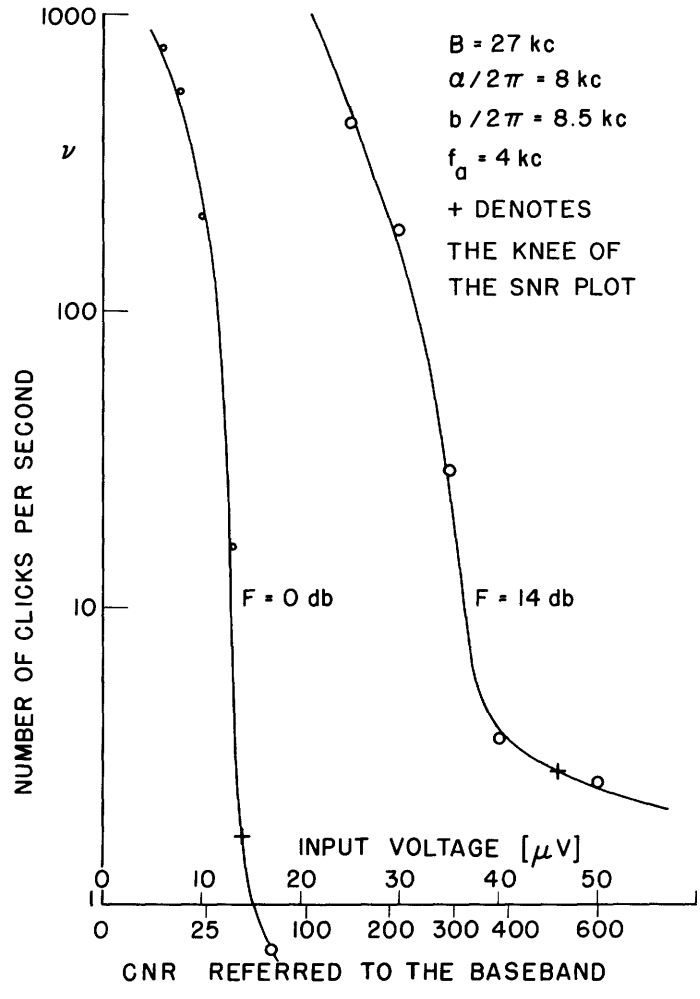

Fig. 23. Number of noise clicks per second with the loop open and closed. (+ denotes the knee on the SNR performance curve.)

noise, this experiment was performed with the system exhibiting the feedback threshold, and then with the same system with the loop open. The results are shown in Fig. 23, and lead to the conclusion that the frequency-compressive system cannot operate satisfactorily near its feedback threshold. The regenerative feedback impulses are very sharp, and therefore contribute little power to the total amount of output noise. But the number of clicks per second in the closed-loop case becomes objectionable at a carrier-to-noise ratio a few decibels above the knee of the SNR curve (see Fig. 23).

Acknowledgment

The author wishes to express his deep appreciation to Professor John M. Wozencraft for his attentive advice during the course of this research. Many helpful suggestions from Professor Edward M. Hoffstetter and Dr. Elie J. Baghdady are also gratefully acknowledged.

This project was started while the author, who was on leave from the Technical University and the Radio Institute, Warsaw, Poland, was holding a Ford Foundation Postdoctoral Fellowship. The research was completed under a fellowship from the Research Laboratory of Electronics of the Massachusetts Institute of Technology.

References

1. D. Middleton, An Introduction to Statistical Communication Theory (McGraw-Hill Book Company, Inc., New York, 1960).
2. E. J. Baghdady (ed.), Lectures on Communication Systems Theory (McGraw-Hill Book Company, Inc., New York, 1961).
3. S. O. Rice, Properties of sine wave plus noise, Bell System Tech. J. 27, 109 (1948).
4. F. L. H. M. Stumpers, Theory of frequency-modulation noise, Proc. IRE 36, 1081 (1948).
5. M. C. Wang, Chapter 13 in Threshold Signals, edited by J. L. Lawson and G. E. Uhlenbeck (McGraw-Hill Book Company, Inc., New York, 1950).
6. D. Middleton, The spectrum of frequency modulated waves after reception in random noise, Quart. Appl. Math. 7, 129 (1949); 8, 59 (1950).
7. D. Middleton, On theoretical signal-to-noise ratios in FM receivers, J. Appl. Phys. 20, 334 (1949).
8. H. W. Fuller and D. Middleton, Signals and Noise in a Frequency Modulation Receiver, Technical Report 242, Cruft Laboratory, Harvard University, 1956.
9. M. Schwartz, Signal-to-noise effects and threshold effects in FM, Proc. National Electronics Conference, Vol. 18, p. 59, 1962.
10. F. J. Skinner, Theoretical Noise Performance Curves for Frequency Modulation Receivers, Memorandum MM-54-2422-5, Bell Telephone Laboratories, Inc., New York, 1954.
11. L. H. Enloe, The synthesis of frequency feedback demodulators, Proc. National Electronics Conference, Vol. 18, p. 477, 1962.
12. M. G. Crosby, Frequency modulation noise characteristics, Proc. IRE 25, 472 (1937).
13. A. Wojnar, FM signal distortion in linear networks (in Polish), Zeszyty Naukowe Politechniki Warszawskiej, Elektryka (Warsaw) 3, 47 (1954).
14. E. J. Baghdady, The theory of FM demodulation with frequency-compressive feedback, Trans. IRE, Vol. CS-10, p. 226, 1962.
15. S. O. Rice, Noise in FM receivers, Proc. Symposium on Time Series Analysis, Brown University, Providence, Rhode Island, June 11-14, 1962.
16. Ibid., see Formula 1.4.
17. E. I. Manayev, On bandwidths with FM reception (in Russian), Radiotekhnika (Radio Engineering) 3, 54 (1948).
18. J. G. Chaffee, The application of negative feedback to frequency-modulation systems, Bell System Tech. J. 18, 404 (1939).
19. P. F. Panter and W. Dite, Application of negative feedback to frequency-modulation systems, Elec. Commun. (London) 26, 173 (1949).
20. J. G. Rodyonov, Optimum passband of a filter in the reception of FM signals with follow-up tuning (in Russian), Radiotekhnika (Radio Engineering) 15, 47 (1960).
21. M. O. Felix and A. J. Buxton, The performance of FM scatter systems using frequency compression, Proc. National Electronics Conference, Vol. 14, p. 1029, 1958.
22. M. Morita and S. Ito, High sensitivity receiving system for frequency-modulated wave, IRE Int. Conv. Record, Vol. 8, Pt. 5, p. 227, 1960.
23. C. L. Ruthroff, Project Echo, FM demodulators with negative feedback, Bell System Tech. J. 40, 1149 (1961).
24. L. H. Enloe, Decreasing the threshold in FM by frequency feedback, Proc. IRE 50, 18 (1962).

25. J. J. Spilker, Threshold, Comparison of Phase-Lock, Frequency-Lock and Maximum-Likelihood Types of FM Discriminators, WESCON Paper No. 14/2, 1961.

26. R. M. Gagliardi, The Design and Capabilities of Feedback FM Receivers, WESCON Paper No. 14/1, 1962.

27. E. Bedrosian, Power-Bandwidth Trade-offs for Feedback FM Systems: A Comparison with Pulse-Code-Modulation, Report RM-2787-NASA, Rand Corporation, 1961.

28. L. H. Enloe, The synthesis of frequency feedback demodulators, Proc. National Electronics Conference, Vol. 18, p. 477, 1962.

29. C. L. Ruthroff and W. F. Bodtmann, Design and performance of a broadband FM demodulator with frequency compression, Proc. IRE 50, 2503 (1962).

30. R. E. Heitzmann, Minimum power FM reception using frequency feedback, Proc. IRE 50, 2503 (1962).

31. W. J. Wilson, Threshold Investigation of FM Feedback Systems, S. M. Thesis, Department of Electrical Engineering, M. I. T., 1963.