Small-Signal Distortion in Feedback Amplifiers for Audio[1]

James Boyk[2] and Gerald Jay Sussman[3]

April 22, 2003

[2]Pianist in Residence, Lecturer in Music in Electrical Engineering, and Director of the Music Lab, California Institute of Technology

[3]Matsushita Professor of Electrical Engineering, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Abstract

We examine how intermodulation distortion of small two-tone signals is affected by adding degenerative feedback to three types of elementary amplifier circuits (single-ended, push-pull pair, and differential pair), each implemented with three types of active device (FET, BJT and vacuum triode). Although high precision numerical methods are employed in our analysis, the active devices are modeled with rather simple models. We have not investigated the consequences of more elaborate models.

Though negative feedback usually improves the distortion characteristics of an amplifier, we find that in some cases it makes the distortion "messier." For instance, a common-source FET amplifier without feedback has a distortion spectrum displaying exactly four spurious spectral lines; adding feedback introduces tier upon tier of high-order intermodulation products spanning the full bandwidth of the amplifier (as suggested by Crowhurst in 1957). In a class-B complementary-pair FET amplifier, feedback mysteriously boosts specific high-order distortion products.

The distortions we are dealing with are small, but we speculate that they may be psychoacoustically significant.

This work also casts light on the relative virtues of the three types of active devices and the three circuit types. For instance, a FET pair run in class-A produces zero distortion even without feedback.

Some experienced listeners report favorably on the sound quality of non-feedback amplifiers. This is surprising, because such amplifiers have much more nonlinear distortion than amplifiers that use negative feedback. Indeed, the appropriate use of negative feedback improves almost all of the theoretical and measurable parameters of an amplifier. Of course, the listeners may be mistaken. Alternatively, some subtle consequent of negative feedback may be responsible for the difference in perception. Here we investigate one such possibility.

Of the various proposals attempting to explain how these perceptual differences might arise, most have suggested bad design errors in the application of feedback. For example, "transient intermodulation distortion" occurs in amplifiers with inadequate slew rate;[1] but real-world transients have rise times that are easy to accommodate with modern circuits.[2]

Another suggestion is that feedback amplifiers are more sensitive to radio-frequency interference. The idea is that the output impedance of a feedback amplifier may be very low at audio frequencies, but rises as the frequency increases. This is because feedback amplifiers must be compensated to ensure stability, and the most common compensation scheme introduces a principal pole at low frequencies that lowers the loop gain as the frequency increases, so that the output impedance rises. If the impedance is high enough, strong radio-frequency fields, as occur in our environment, can come in through the output and wreak havoc by rectification, shifting the bias conditions of the amplifier. This is certainly possible, but a well-shielded amplifier with appropriate filters need not have this problem.

Yet another idea is that the clipping behavior of feedback amplifiers is different from that of non-feedback amplifiers: clipping is sharper and recovery from clipping may be problematic. This also is true, but it does not explain the perceptual differences that may remain even in the case that the amplifiers are not driven to clipping: listeners report differences in the low-level details and the sound of the "room," the recording venue. One of us (JB) has

[1]Transient intermodulation distortion occurs when the amplifier cannot slew fast enough to follow the transient. During such a transient, the amplifier is pinned to the slew trajectory and cannot follow variations in the input. Apparently this effect was known as early as Roddam in 1952 [8], but it only became widely known in audio circles with the work of Otala [7].

[2]Boyk [2] has made a survey of wideband spectra from real musical sources. Although some of these show significant energy above 20 kHz they put limits on the rate of change of the sound pressure level in most ordinary music waveforms.

observed that the introduction of feedback into one particular (microphone pre-) amplifier seems to "separate" the very high frequencies from the rest of the range, as though a badly-integrated super-tweeter had been added to the monitor system. This yields an unnatural sound that seems correlated with but disconnected from the program material.

We investigate the possibility that the difference in sound quality is not an accident of the particular design but is an inherent characteristic of negative feedback. The idea is not new with us. In 1957, Norman Crowhurst [4] observed that since the intrinsic nonlinearity of an amplifier must produce harmonic and intermodulation products from the components of the program material, feedback will combine these products with the program to produce further distortion products. Since many of the products in each "generation" are higher or lower in frequency than the signals that produce them, the effect will be to create products extending over the full bandwidth of the amplifier. Although the total amount of this distortion is very small—much smaller than the lower-order distortion produced by the same amplifier without feedback—Crowhurst observed, "The logical result of this process would be a sort of program-modulated, high-frequency 'noise' component, giving the reproduction a 'roughness'." We speculate that this "noise," constantly changing as it is (because it is correlated with the program material), may interfere with the listeners' perceptions of low-level detail. Such speculation is not new with us either. As far back as 1950, Shorter [9] was worried about the perceptual effect of high-order distortion products; and the idea has been periodically revisited by many authors, including, most recently, Daniel H. Cheever [3], who developed a new measurement strategy which attempts to quantify the effect of this kind of distortion on perception.

In what follows, we examine the responses of nine elementary circuits to two-tone inputs, comparing in each case the behavior without feedback to that with feedback (in some cases more than one amount of feedback). Three of the nine basic circuits are simple stages using a single FET, BJT or vacuum triode; another three use pairs of these devices in complementary (FET, BJT) or push-pull (triode) configurations; and the final three are differential-input circuits. Each circuit is studied at a signal level which best reveals the behavior of interest.

As usual in such work, the real subject of our study is the behavior of certain mathematical equations and relations. When we ascribe the behavior to circuits, we are assuming that the active devices are modeled perfectly by

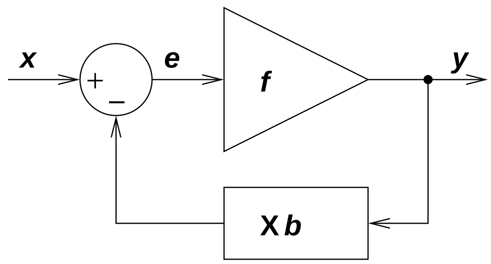

Figure 1: A feedback amplifier

the stated "laws" (square for FETs, three-halves for triodes, exponential for BJTs); that complementary pairs are perfectly symmetrical; that the tubes used in push-pull are identical; and that transformers are perfect. Though none of these is true, the results may yet be useful.

Our analysis is stateless; that is, we assume no frequency-dependent elements. Real amplifiers have such elements, but we can see the essential behavior without considering them.

The core of this paper is the numerically-derived spectra that we obtain for a variety of amplifiers, each being considered both with and without feedback. We begin, however, with a formal derivation of analytic estimates for the lowest order spectral lines, which we can use to check the validity of the numerical work; and a bit of circuit analysis. The reader may wish to skip this preliminary analysis and proceed to the discussion of the spectra that begins on page 10 under "Our spectra," touching down at figure 2 on the way.

Feedback in a nonlinear system

In figure 1 the amplifier is modeled by a function f that is in general nonlinear. The feedback path is assumed to be a linear path that multiplies by a constant b. Thus, the equation relating the output y to the input x is

$$y = f(x - by). \tag{1}$$

If f were linear, say $f(e) = Ae$ we could solve for y to get the familiar Black's formula,

$$y = \frac{Ax}{1 + Ab}. \tag{2}$$

If A is very large then $y \approx x/b$, allowing us to reliably make amplifiers with gain $1/b$ using amplifiers with large, but uncontrolled gain. However, the distortion we are interested in is due to the *nonlinearity* of f.

Assume that f may be expressed as a power series

$$y = A_1 e + A_2 e^2 + A_3 e^3 + \cdots \tag{3}$$

with no offset term, so if $e = 0$ then $y = 0$. To account for the feedback we can substitute $(x - by)$ for e to obtain

$$y = A_1(x - by) + A_2(x - by)^2 + A_3(x - by)^3 + \cdots. \tag{4}$$

In general, we can solve for y, producing a power series that represents the entire transfer function of the feedback amplifier. This series

$$y = a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{5}$$

can be obtained by taking derivatives of equation (4):

$$a_1 = \left.\frac{dy}{dx}\right|_{x=0} = \frac{A_1}{1 + A_1 b} \tag{6}$$

$$a_2 = \left.\frac{1}{2}\frac{d^2y}{dx^2}\right|_{x=0} = \frac{A_2}{(1 + A_1 b)^2} \tag{7}$$

$$a_3 = \left.\frac{1}{6}\frac{d^3y}{dx^3}\right|_{x=0} = \frac{(A_3 A_1 - 2A_2^2)b}{(1 + A_1 b)^5} \tag{8}$$

$$\cdots$$

We see that a_1 is the gain we would expect if the amplifier were linear, and the higher-order terms are the distortion. If we make the input a sinusoid $x(t) = C \cos \omega t$, expand powers using the multiple angle formulas,[3] and collect like terms, we get a Fourier series showing the harmonic components. Considering only the first three terms of equation (5) we get:

$$y = \left(Ca_1 + \frac{3}{4}C^3 a_3\right) \cos \omega t + \frac{1}{2}C^2 a_2 \cos 2\omega t + \frac{1}{4}C^3 a_3 \cos 3\omega t. \tag{9}$$

Thus, for small signals (C small) the relative size of the second harmonic and third harmonic distortion terms are:

$$\mathrm{HD}_2 \approx \frac{1}{2}C\frac{a_2}{a_1} = \frac{1}{2}\frac{A_2}{A_1(1 + A_1 b)} \tag{10}$$

$$\mathrm{HD}_3 \approx \frac{1}{4}C^2\frac{a_3}{a_1} = \frac{1}{4}\frac{(A_3 A_1 - 2A_2^2)b}{A_1(1 + A_1 b)^4} \tag{11}$$

[3]For example, $(\cos \alpha)^2 = \frac{1}{2} + \frac{1}{2}\cos 2\alpha$ and $(\cos \alpha)^3 = \frac{3}{4}\cos \alpha + \frac{1}{4}\cos 3\alpha$.

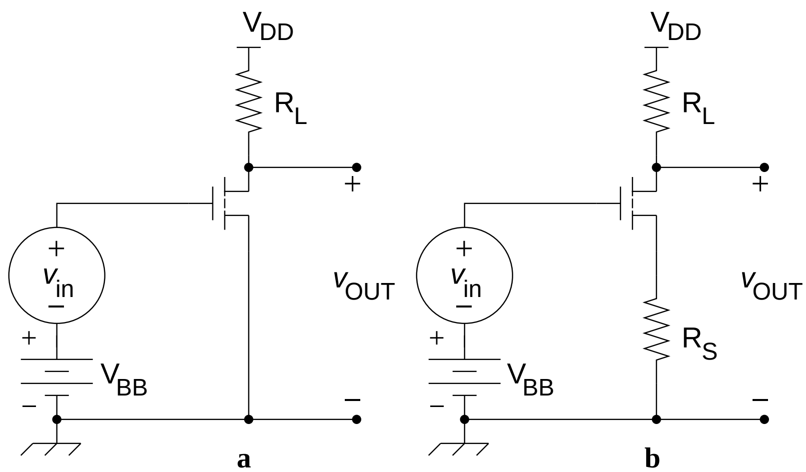

Figure 2: Amplifier **a** has no feedback; amplifier **b** has source feedback.

If we make the input the sum of two sinusoids of different frequencies, we can compute the intermodulation products as well. However this kind of algebra is not usually easy to understand past the first few terms. Since we are interested in the high-order terms, we will not pursue the algebraic approach, but will use it only to check the lowest terms of the numerical results, and to help us understand how the distortion terms are generated.

A tale of two FET amplifiers

The two FET circuits in figure 2 are identical except that amplifier **a** has no feedback while amplifier **b** has source feedback created by non-zero R_S. In the analysis below, we find that the amplifier without feedback introduces only second harmonics and first-order sums and differences of signal components; and that while adding feedback lowers these distortion products a bit, it also produces all orders of harmonic and intermodulation products in tiers stepping down in level but extending to the full bandwidth of the amplifier.

We will operate the Field-Effect Transistor (FET) in the saturated region, where it exhibits simple square-law nonlinearity: a good approximation for its drain current is

$$i_D = \frac{k}{2}(v_{GS} - V_T)^2, \tag{12}$$

where v_{GS} is the gate-to-source voltage and where k and V_T are parameters of the FET. The FET is in saturation so long as $V_T < v_{GS} < V_T + v_{DS}$, where v_{DS} is the drain-to-source voltage.

Analysis: the no-feedback case

To simulate the circuit we need the output voltage as a function of the input voltage. The output voltage is

$$v_{OUT} = V_{DD} - R_L i_D, \tag{13}$$

so we need the drain current i_D. For amplifier **a**, $v_{GS} = v_{in} + V_{BB}$, so substituting this into equation (12) and then plugging the resulting current into equation (13) we obtain an expression for the output voltage in terms of the input voltage:

$$v_{OUT} = V_{DD} - \frac{k}{2}(v_{in} + V_{BB} - V_T)^2 R_L. \tag{14}$$

Equation (14) is all we really need to obtain a numerical spectrum for any given input signal, as described on page10. But first, to see what we should expect, we do some analysis along the lines of equations (1–11), but now specific to these FET amplifiers.

We can rewrite equation (14) as

$$v_{out} = v_{OUT} - V_{OUT} = a_1 v_{in} + a_2 v_{in}^2, \tag{15}$$

a simple quadratic, where

$$\begin{aligned} V_{OUT} &= V_{DD} - \frac{k}{2}(V_{BB} - V_T)^2 R_L & (16) \\ a_1 &= -k(V_{BB} - V_T) R_L & (17) \\ a_2 &= -\frac{1}{2} R_L k. & (18) \end{aligned}$$

The incremental gain of amplifier **a** is thus

$$\left.\frac{\partial v_{OUT}}{\partial v_{in}}\right|_{v_{in}=0} = -k(V_{BB} - V_T) R_L = a_1 \tag{19}$$

This helps us choose circuit values to obtain a given gain.

Because this amplifier exhibits a simple quadratic law equation (15), its distortion products can only be second harmonics and sums and differences of the Fourier components.[4] In this simple case it is easy to work out the spectrum symbolically. (We will use this result on page 11 to check our numerical simulation.)

We define the excess gate bias $V_B = V_{BB} - V_T$, and the corresponding drain bias current $I_D = \frac{1}{2}kV_B^2$. Then the total drain current can be rewritten as

$$i_D = I_D\left(1 + 2\frac{v_{in}}{V_B} + \left(\frac{v_{in}}{V_B}\right)^2\right). \tag{20}$$

If we drive this amplifier with a sinusoid

$$v_{in} = A\cos\omega t \tag{21}$$

we obtain

$$i_D = I_D(b_0 + b_1\cos\omega t + b_2\cos 2\omega t) \tag{22}$$

where

$$b_0 = \frac{1}{2}\left(1 + \left(\frac{A}{V_B}\right)^2\right) \tag{23}$$

$$b_1 = 2\frac{A}{V_B} \tag{24}$$

$$b_2 = \frac{1}{2}\left(\frac{A}{V_B}\right)^2. \tag{25}$$

This Fourier series has only three terms. Comparing the magnitude of the second harmonic component to the magnitude of the fundamental we obtain

$$\mathrm{HD}_2 = \frac{b_2}{b_1} = \frac{A}{4V_B}. \tag{26}$$

We could work out the sizes of the Fourier components for the two-tone stimulus, getting terms for the sum and difference as well as the two second harmonics, but for more complicated circuits this would be much harder.

[4] Remember that squares of weighted sums of sinusoids can be expressed as weighted sums of sinusoids with angles that are sums and differences of the given angles.

The feedback case

Amplifier **b** is a bit more complicated. As before, we need an expression for v_{OUT} in terms of v_{in}. The output voltage is

$$v_{OUT} = V_{DD} - R_L i_D, \tag{27}$$

so we need the drain current i_D. But to compute the drain current we must solve a quadratic equation:

$$\frac{k}{2} R_S^2 i_D^2 - (k R_S v_G + 1)\, i_D + \frac{k}{2} v_G^2 = 0, \tag{28}$$

where $v_G = v_{in} + V_{BB} - V_T$. We pick the correct root so that as $R_S \to 0$ the circuit approximates the behavior of amplifier **a**. This gives

$$i_D = \frac{1}{R_S}\left(v_G + \frac{1}{R_S k}\right) - \frac{\sqrt{2}}{R_S\sqrt{R_S k}}\sqrt{v_G + \frac{1}{2R_S k}}. \tag{29}$$

Substituting into equation (27) we obtain

$$v_{OUT} = V_{DD} - \frac{R_L}{2kR_S^2}\left[\sqrt{1 + 2kv_G R_S} - 1\right]^2, \tag{30}$$

which is what we need for the simulation.

The incremental gain is

$$\left.\frac{\partial v_{OUT}}{\partial v_{in}}\right|_{v_{in}=0} = -\frac{R_L}{R_S}\left(1 - \frac{1}{\sqrt{1 + 2R_S\,(V_{BB} - V_T)\,k}}\right), \tag{31}$$

which goes to $-R_L k(V_{BB} - V_T)$ as $R_S \to 0$, as required by the condition that circuit **a** is a special case of circuit **b** ($R_S = 0$).

If we expanded equation (30) as a power series we would see that the square-root term expands into all powers of the incremental input voltage v_{in}. So by contrast with the simple amplifier **a** the feedback amplifier **b** produces not just second harmonics but all orders of harmonics; not just simple sums and differences, but all orders of intermodulation products. This illustrates the idea behind the claim that the feedback amplifier produces a more complex spectrum than the simple amplifier. However, to learn more we have to be quantitative.

Our spectra

Using numerical simulation, we compared the behavior of these amplifier topologies when stimulated by a two-tone signal. As for all examples in this paper, the two tones were at frequencies 3 and 5. Because the simulations are done with no frequency-dependent elements the units of the frequencies do not matter; all that matters is their ratio. The frequencies were chosen to be relatively prime (they are both actually prime) so as to show the maximum number of independent components. Perhaps it would be better to use incommensurate frequencies, such as 3 and 3ϕ, where $\phi = (1 + \sqrt{5})/2$, the golden ratio; but we chose integer frequencies and an integer timespan so that the spectra would come out as clean lines, without spectral leakage or skirts due to the window function. The small errors introduced by choosing integers rather than incommensurate numbers are not significant in our results.

All of the spectra in this paper were developed using numerical-analysis procedures written by one of the authors (GJS). For each circuit, a numerical procedure was written to determine the output voltage in terms of the input voltage, for each moment of time. This was easy when the relationship was given by a single equation, such as equation (14) for the no-feedback single-ended FET stage. Other cases, however, required solving simple nonlinear systems of equations, such as equations (28–30) for the same FET stage with feedback.

In general, equation solutions are accurate to one part in 10^{14} (-280 dB). Other errors sometimes contribute to bring the noise floor 15 dB higher. The two-tone input was generated in a time span of 16 and sampled with 4096 points. The output voltage was computed for each of these input points and transformed with a 4096-point transform to obtain the frequency spectrum, with a maximum representable frequency of 128. However, our spectral plots only show frequencies up to 32. In fact there are no distortion components above the noise floor in our data above a frequency of 127, so our graphs are not contaminated by aliases.

Two-tone spectra of FET amplifiers

The amplifiers were designed to have an incremental gain of -10 (that is, 20 dB, inverting), using formulas (19) and (31). The FETs were assumed to have $k = 0.002$ A V^{-2} and $V_T = 1.0$ V.[5] The other device parameters and operating conditions are given in the table below. We see that both amplifiers are comfortably biased into the saturation region, and that the source resistor in circuit **b** produces only a small amount of feedback (about 1.8 dB). The bias current in amplifier **a** is 1.0 mA; and in **b**, about 0.814 mA. The table shows parameters for these two amplifiers, and for amplifier **c**, with the same topology as **b** but more feedback (about 9.5 dB). We include **c** to demonstrate the robustness of the conclusions.

	a	**b**	**c**	
V_{BB}	2.0	2.0	3.0	V
R_L	5000	6800	15000	Ω
R_S	0	120	1000	Ω
I_D	1.0	0.814	1.0	mA

In these experiments the two tones of the stimulus are of equal amplitude (.05 peak volts), at frequencies 3 and 5. As mentioned above, we use the same stimulus frequencies throughout this paper, though the amplitudes may vary; and all single-ended amplifiers studied have incremental gains of -10 (that is, 20 dB, inverting).

In the spectrum of amplifier **a** (figure 3) the fundamental components (frequencies 3 and 5) have been normalized to 0 dB. The second harmonics (frequencies 6 and 10) are at -38 dB; and the sum and difference intermodulation products (frequencies 2 and 8) are at about -32 dB.

Here we can check the simulations against the theory: The simulations show that the second harmonics are down by -38 dB. If we evaluate the ratio of the second harmonic to the fundamental using equation (26), plugging in 0.05 for A and 1.0 for $V_B = V_{BB} - V_T$, we find that the ratio is 0.0125 or -38.06 dB.

The spectrum of the feedback amplifier **b** (figure 4) is more complicated, as expected. The second harmonics are at about -40.5 dB, and the sum and difference frequencies are at about -34.5 dB. This is a small improvement—about 2.5 dB in each line—compared to the amplifier without feedback. However, there is a new tier of components with peaks at about -80 dB; and even more components down around -120.

[5] These paramters are typical for an N-channel enhancement-mode MOSFET when used for small-signal amplification.

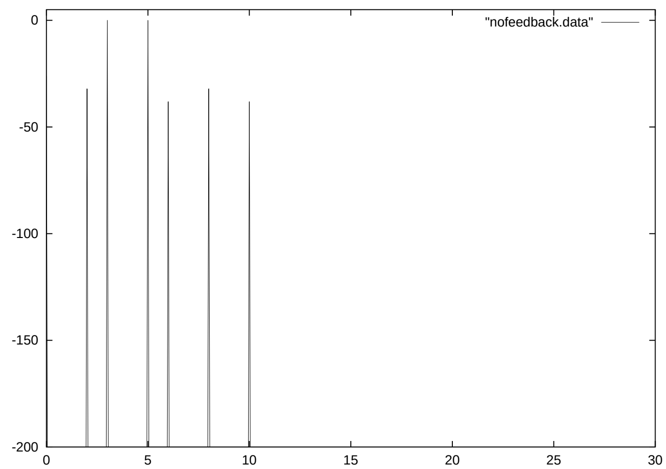

Figure 3: The spectrum of amplifier **a**: a single-ended FET without feedback; the two-tone input has 0.05 peak volts in each component.

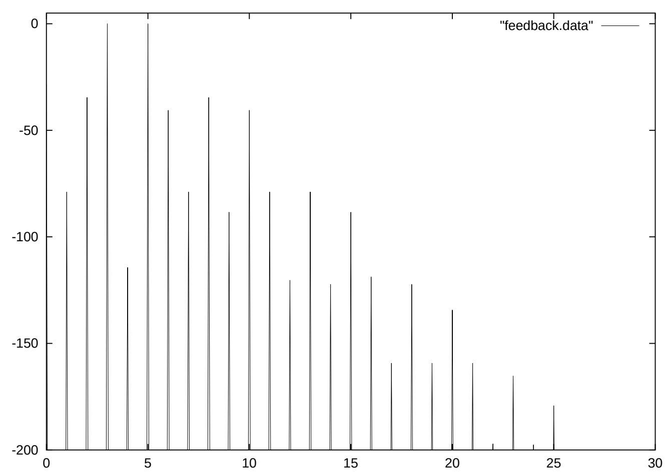

Figure 4: The spectrum of amplifier **b**: a single-ended FET with source feedback; the two-tone input has 0.05 peak volts in each component.

The spectrum of amplifier **c** is not shown. Its additional feedback compared to **b** makes for a distortion spectrum which is similar except that all of the products are pushed down in level.

Returning to feedback amplifier **b**, in figure 5 we expand the vertical scale of the spectrum to see the structure more clearly. We see many tiers of distortion products, each produced by an additional circulation around the feedback loop and 40 dB below the previous tier. The noise just above −300 dB is due to numerical error in the equation solver.

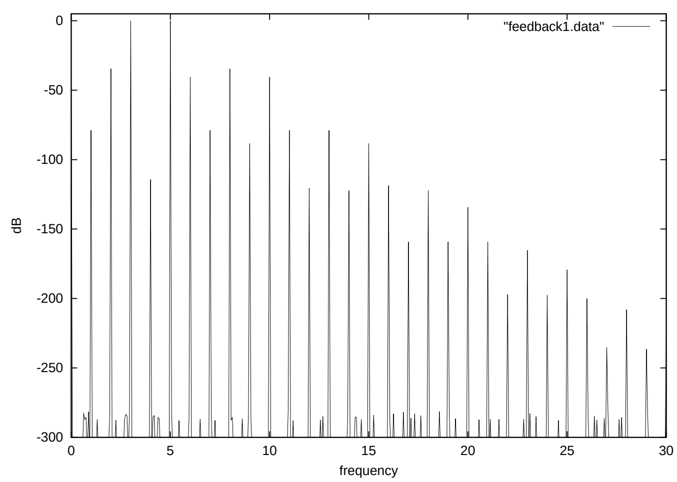

Figure 5: Expanded spectrum of amplifier **b**: a single-ended FET with source feedback; the two-tone input has 0.05 peak volts in each component.

Thus we observe that adding negative feedback to a FET amplifier, while decreasing the overall amount of distortion, significantly changes the distribution of the distortion products. In two-tone tests, feedback introduces new tiers of products, most very weak, but not necessarily insignificant perceptually, as they produce a noise floor *correlated with* the program material. And not only does the amplitude of this noise floor rise and fall with the amplitude of the program material, but its character changes as it rises and falls, higher-order products being more volatile than lower-order ones. If we increase the drive of the FET amplifier by 12 dB, from 0.05 peak volts to 0.2 peak volts, the distortion at frequency 2 rises by 24 dB (12 dB relative to the input signals), which we may take simply as due to higher signal levels. But the components at frequencies 17 and 19 increase by 60 dB (48 dB relative to