Audio power amplifier design — 5

Negative feedback and non-linearity distortion

Why does the low note contain the sound of the high note? — ARISTOTLE

by **Peter J. Baxandall,** B.Sc. (Eng), F.I.E.E., F.I.E.R.E.

The July article in the present series concluded the treatment of the basic techniques for achieving feedback-loop stability. Attention will now be given to the effects of negative feedback on non-linearity distortion, and it will be shown that some of the ideas involved are more subtle than is sometimes appreciated.

THE following treatment, which has gradually become clarified and extended in scope over a period of many years, will, it is hoped, enable the reader to see what the answers to questions such as the following should be:

(a) Is it a valid criticism of the use of large amounts of negative feedback that it converts moderate amounts of low-order harmonic and intermodulation distortion into a multitude of small-amplitude distortion products of high order, which may be subjectively more significant?

(b) Is it always desirable to design a feedback audio amplifier to have a nearly-level audio-frequency response before feedback is applied?

(c) Does plenty of feedback at medium audio frequencies, assuming there are no slew-rate or other overload effects, necessarily ensure that two or more signal components near the top of the audio band will give rise to negligible intermodulation products at medium frequencies?

(d) Is it important for an audio amplifier to give low distortion when signals at frequencies lying outside the audio band are fed into it?

Obviously, in many amplifier circuits, owing to the presence of capacitors or transformers, or because of insufficient bandwidth in transistors, frequency-dependent effects will have to be invoked when considering distortion mechanisms. In some practical audio circuits, however, such effects may be negligible. The following treatment will initially assume no significant frequency-dependence, and will provide a foundation of theoretical understanding which may later be extended to include the influence of frequency.

Amplifier with parabolic transfer characteristic

Consider the basic feedback amplifier configuration shown in Fig. 1. The voltage symbols represent instantaneous voltages, and each polarity marked is that which exists when the corresponding symbol has an instantaneously positive value. For the feedback to be negative, either A or β must be negative. (For a defence of the sign convention adopted, see page 41 of the March 1978 issue.) For present purposes it will be convenient to take A as being positive, so that β will be negative.

The simplest form of non-linearity to consider is that in which the transfer characteristic of the amplifying device, i.e. the graph of instantaneous output voltage (or current) against instantaneous input voltage (or current), departs from being a straight line only because of the presence of a square-law term in the corresponding equation*. Thus, referring to Fig. 1, let

$$v_{out} = Av' + \alpha\,(Av')^2 \qquad (1)$$

The graph of this equation is the transfer characteristic shown in Fig. 2. Plotted on this convenient basis, with equal scales on the two axes, the 45° broken line represents the slope at the origin, the actual characteristic departing from the ideal straight line by the amount $\alpha\,(Av')^2$ as shown. Because equation (1) is a quadratic equation, representing a parabola (of which only part is drawn in Fig. 2), the graph is sometimes called a quadratic transfer characteristic.

If there is no feedback in the Fig. 1 circuit ($\beta = 0$), v' becomes equal to v_{in}, and the complete circuit then has a transfer characteristic equation as in (1) but with v_{in} written for v'. Suppose now that v_{in} is a sine-wave signal given by

$$v_{in} = \hat{V}_{in}\sin\,\omega t \qquad (2)$$

Substituting this for v' in equation (1) gives

$$v_{out} = A\hat{V}_{in}\sin\omega t\; + \alpha A^2\hat{V}_{in}^{\,2}\sin^2\omega t \qquad (3)$$

(No feedback)

The first term represents the wanted fundamental output, the other term representing the second-harmonic distortion because

$$\sin^2\omega t = \tfrac{1}{2} - \tfrac{1}{2}\cos 2\omega t \qquad (4)$$

This elementary trigonometry formula may be illustrated graphically as in Fig. 3. (I trust that those readers highly familiar with such elementary ideas will bear with me until more interesting topics are reached – I have assumed that some readers will welcome a rather slow and basic approach.)

* It is tempting to call this equation the 'transfer function', but this usage is better avoided because the term has an almost universally accepted meaning in a somewhat different context, as explained in the March 1978 article. It is thus better to refer simply to 'the equation of the transfer characteristic'.

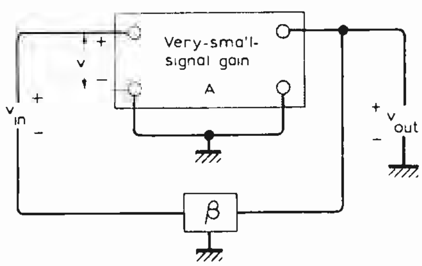

Fig. 1 *Basic feedback-amplifier configuration.*

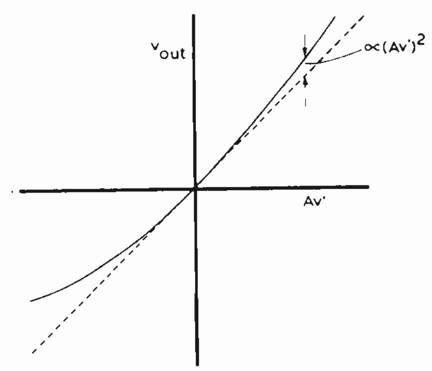

Fig. 2 *Simple parabolic, or quadratic, transfer characteristic. α is a constant determining the degree of non-linearity, and A and V' are as in Fig. 1.*

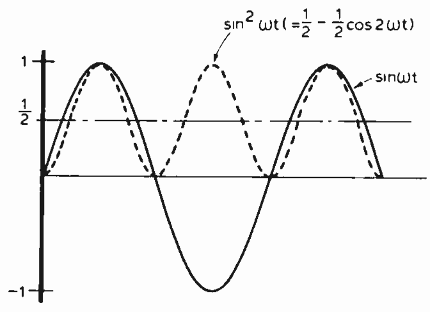

Fig. 3 *Waveforms illustrating a basic trigonometry formula.*

If (4) is substituted for $\sin^2\omega t$ in (3), it will be seen that the magnitude of the second-harmonic output voltage component is given by

$$\hat{V}_{2nd} = \tfrac{1}{2}\alpha A^2 \hat{V}_{in}^2 \qquad (5)$$
(No feedback)

The magnitude of the fundamental output is given by

$$\hat{V}_{fund} = A\hat{V}_{in} \qquad (6)$$
(No feedback)

Dividing (5) by (6) and multiplying by 100 gives the percentage second-harmonic distortion as

$$\%2nd = \tfrac{1}{2}\alpha A\hat{V}_{in} \times 100 \qquad (7)$$
(No feedback)

Thus, from (5) and (7), the absolute magnitude of the second-harmonic output voltage is proportional to the square of the input (or fundamental output) voltage, whereas the percentage second-harmonic distortion is linearly proportional to the input voltage itself. This is a property of any circuit or device in which square-law distortion is dominant. (It may here be mentioned that a statement such as "the distortion is proportional to the square of the output voltage" is really quite ambiguous, for "the distortion" can be taken to mean either "the distortion voltage" or "the percentage distortion". This ambiguity often appears in the literature and sometimes causes very real confusion. A plea is therefore made to authors to say what they mean!)

The problem now to be considered is the effect on distortion of making β finite in Fig. 1, i.e. applying negative feedback, still assuming a parabolic transfer characteristic for the basic amplifier. This problem may be approached from several different angles, and, as is often the case, adopting more than one viewpoint is helpful in providing a more complete understanding of the principles involved.

First of all it is possible to construct, point by point, a graph of v_{out} against v_{in} with feedback operative, and to show that it is much more nearly linear than without feedback. To do this, a particular value of v' is taken, and from equation (1), assuming A (the gain for very small signals) is known, v_{out} is calculated. Then βv_{out} is determined. Finally, with due care over signs, v_{in} is obtained from the relationship

$$v' = \beta v_{out} + v_{in} \qquad (8)$$

A little thought will show that as the magnitude of A or β is increased, the resultant transfer characteristic becomes more and more nearly a perfect straight line. With very large A or β, v_{in} becomes enormously greater than v', and the overall gain is then given very nearly by

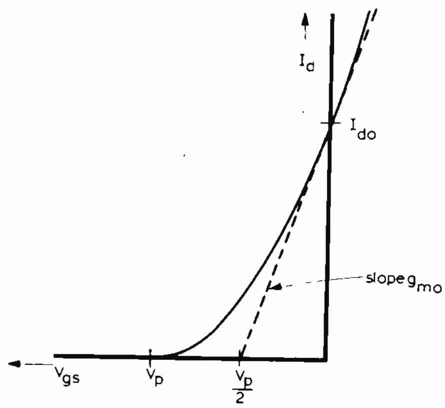

Fig. 4 *Ideal parabolic characteristic for f.e.t.*

$$v_{out}/v_{in} = -1/\beta \qquad (9)$$
(Infinite feedback)

The change from a parabolic transfer characteristic to a straight line as the loop gain is increased from zero to infinity is a smooth and gradual process. All the intermediate transfer characteristics are absolutely smooth curves, quite free from any suggestion of kinks or other blemishes. But is each one still a parabola, of lesser curvature?

The answer to the above very important question is "no", and an indication that this must be so can be obtained without actually working out the equation of the new transfer characteristic. Start with $\beta = 0$ (no feedback). With a sine-wave input at frequency f, the output will contain components at f and $2f$. As soon as β is made finite, some of the $2f$ component will be let through into the input circuit, so that the basic amplifier will now be receiving inputs at f and $2f$.

Now any device with a parabolic, or quadratic, transfer characteristic, when fed with two inputs at different frequencies, generates intermodulation products at the sum and difference frequencies – and the sum frequency in the present case is $3f$. (This arises from the fact that $(\sin\omega_1 t + \sin\omega_2 t)^2$ gives a term $2\sin\omega_1 t\sin\omega_2 t$ which is equal to $\cos(\omega_1 - \omega_2)t - \cos(\omega_1 + \omega_2)t$.) Thus, while the amplifier without negative feedback gives nothing but second-harmonic distortion on a single sine-wave input, as soon as a little feedback is applied, a third-harmonic output appears. This is not the end of the story, however, for this third harmonic, like the second harmonic, gets fed via the β-network into the input circuit, where sum and difference products are again generated. This time the sum products are at $f + 3f$, which gives a fourth harmonic, and $2f + 3f$, which gives a fifth harmonic. Clearly there is theoretically no end to this process – every new harmonic considered, when fed back, gives rise to harmonics of yet higher order. Before too hastily condemning feedback, however, it is wise to consider the magnitude of these effects, and also to question whether assuming a purely parabolic transfer characteristic is sufficiently closely related to the behaviour of practical devices to be of much value. Maybe they already produce comparable amounts of high-order harmonics before feedback is applied? It is evident that a fully satisfactory understanding of the problem can best be reached by a combination of theory and experiment.

Before presenting experimental results for comparison, the theory of feedback over an ideal parabolic device will be pursued further, to obtain the actual magnitudes of the various harmonics generated. The full analysis is somewhat tedious, but an outline of the approach adopted is as follows. The aim is to obtain an expression for the closed-loop transfer characteristic in the form of a power series

$$v_{out} = a_1 v_{in} + a_2 v_{in}^2 + a_3 v_{in}^3 + \ldots \qquad (10)$$

Then $v_{in} = \hat{V}_{in}\sin\omega t$ is put in this and the resultant harmonic magnitudes are obtained. To obtain the power series, the starting point is equation (8), the value of v' there given being substituted in equation (1). This produces a quadratic equation relating v_{in} and v_{out} which can be solved to give v_{out} as a direct function of v_{in}. The function, however, contains a square-root sign and is not in itself a power series. The binomial theorem is then used to obtain the wanted power series. Substituting $v_{in} = \hat{V}_{in}\sin\omega t$ in this series gives terms in $\sin\omega t$, $\sin^2\omega t$, $\sin^3\omega t$ etc. As illustrated in Fig. 3, $\sin^2\omega t$ produces second harmonic, and elementary extension of this principle shows that the $\sin^3\omega t$ term produces third harmonic,† and so on. The various harmonic amplitudes are thus obtained as functions of the peak input voltage, $\hat{V}_{in}$. More conveniently, however, for practical purposes, the harmonic magnitudes are expressed as functions of $\hat{V}_{out}$, on a percentage basis. This is preferable, because in assessing the performance of a feedback amplifier, one is interested in the percentages of the various harmonics present at known *output* levels, and how these vary with the amount of negative feedback used. The results of the analysis are given in Table 1. α is the "distortion constant" of equation (1), A is the amplifier forward gain for very small signal levels, and β is the feedback factor.

It is instructive to plot curves from the Table 1 formulae and to see how they compare with curves based on measurements using an approximately

† Some third-harmonic is also produced by the $\sin^5\omega t$ term, but in view of the much smaller magnitude of this contribution except at signal levels approaching the overload point, it may reasonably be neglected. The output level used in the tests is just low enough to avoid serious errors from this cause.

Table 1. Theoretic distortion formulae for feedback amplifier with parabolic forward transfer characteristic.

Harmonic number	Percentage of fundamental	Ratio of harmonic amplitudes	
		Harmonics	Ratio
2	$\frac{50\alpha\hat{V}_{out}}{1-A\beta}$	2nd : 3rd	$1 \times \frac{1-A\beta}{\alpha\hat{V}_{out}\lvert A\beta\rvert}$
3	$\frac{50\lvert A\beta\rvert\alpha^2\hat{V}_{out}^2}{(1-A\beta)^2}$	3rd : 4th	0.800 " "
4	$\frac{62.50A^2\beta^2\alpha^3\hat{V}_{out}^3}{(1-A\beta)^3}$	4th : 5th	0.714 " "
5	$\frac{87.50\lvert A^3\beta^3\rvert\alpha^4\hat{V}_{out}^4}{(1-A\beta)^4}$	5th : 6th	0.667 " "
6	$\frac{131.25A^4\beta^4\alpha^5\hat{V}_{out}^5}{(1-A\beta)^5}$		

quadratic device such as an f.e.t. Now it will be noticed that the product $\alpha\hat{V}_{out}$, raised to various powers, occurs throughout the formulae, and a value for this must be decided upon before a set of curves, such as those shown in Fig. 7, can be drawn. A convenient procedure is to choose the value of α so that the theoretical percentage second-harmonic distortion without feedback, given by the Table 1 formula as $50\alpha\hat{V}_{out}$, is the same as the measured second-harmonic distortion at the value of $\hat{V}_{out}$ adopted. This effectively matches the value of α to that of the practical circuit, and is more convenient than determining α by other means.

F.e.t. characteristics

Most text books give the following equation for the drain current, I_d, of an f.e.t. whose drain-to-source voltage, V_{ds}, is well in excess of pinch-off voltage, V_p.

$$I_d = I_{do}\left[\frac{V_{gs}}{V_p} - 1\right]^2 \qquad (11)$$

This is a parabolic relationship, as illustrated in Fig. 4, and from the geometry of this diagram it follows that

$$g_{mo} = \frac{2I_{do}}{V_p} \qquad (12)$$

An f.e.t. would therefore appear to be the ideal parabolic device for checking the distortion theory evolved above.

However, several years ago, it struck me that there would be something rather queer about a device if it accurately followed a law as given by equation (11), the reasoning being as follows. Differentiating (11) gives

$$g_m = \frac{dI_d}{dV_{gs}} = \frac{2I_{do}}{V_p} \times \left[\frac{V_{gs}}{V_p} - 1\right] \qquad (13)$$

But from (11)

$$\frac{V_{gs}}{V_p} - 1 = \sqrt{\frac{I_d}{I_{do}}}$$

and substituting this in (13) leads to

$$g_m = \frac{2I_{do}}{V_p}\sqrt{\frac{I_d}{I_{do}}}$$

Finally, using the relationship (12), this becomes

$$g_m = g_{mo}\sqrt{\frac{I_d}{I_{do}}} \qquad (14)$$

According to this equation, as the working drain current I_d is reduced, g_m falls off in proportion to the square root of I_d. Now for a junction transistor g_m varies with collector current I_c according to the relationship

$$g_m = I_c \times \frac{q}{kT} \qquad (15)$$

where q = charge of an electron, k = Boltzmann's constant, and T = absolute temperature.

Here g_m is proportional not to the square root of the collector current, but to the collector current itself, and with silicon planar transistors the relationship holds accurately in practice down to currents of less than a nanoamp. Thus, while an f.e.t. will normally have a lower g_m than a junction transistor at, say, 1mA, the more gradual fall-off in g_m with working current for the f.e.t. would, if continued, give it a much larger g_m than a junction transistor when operated at a low enough current. In view of the very basic quantities involved in equation (15), I felt this result was probably too good to be true! A measurement of g_m for an f.e.t. over a wide range of drain current was therefore made, and gave the result shown in Fig. 5. Thus it seems that a law of nature does indeed come into action to prevent the g_m of an f.e.t. exceeding that of a junction transistor. It will be seen that the steeper broken-line asymptote is fairly closely that expected for a junction transistor, and would, if continued to the right, give a g_m of nearly 40mA/V at 1mA.

Because of the above discrepancy between the usual text-book equation (11) and what is found to happen in practice, if for no other reason, one would not expect the transfer chracteristic, corresponding to Fig. 4, for an actual f.e.t., to be quite precisely parabolic. Consequently, even without negative feedback, harmonic components in addition to second harmonic must be expected to appear to some extent.

However, despite the above, the assumption that the transfer characteristic for an f.e.t. is as given by equation (11) is quite near enough to the mark to permit the magnitude of the second-harmonic distortion without feedback to be fairly accurately calculated – provided the working current is not excessively small (see Fig. 5). It may be deduced from equation (11) that

$$\%2nd = 12.5\frac{\hat{I}}{I_{dc}} \qquad (16)$$

(f.e.t. without feedback)

where $\hat{I}$ is the peak fundamental drain-current excursion and I_{dc} is the working d.c. drain current.

Equation (16) may be compared with the result for an ideal voltage-driven junction transistor, which is

$$\%2nd = 25\frac{\hat{I}}{I_{dc}} \qquad (17)$$

(Junction transistor without feedback)

In this latter case an alternative formula [1,2] is

$$\%2nd = \hat{V}_{in} \qquad (18)$$

where $\hat{V}_{in}$ is the peak signal input vol-

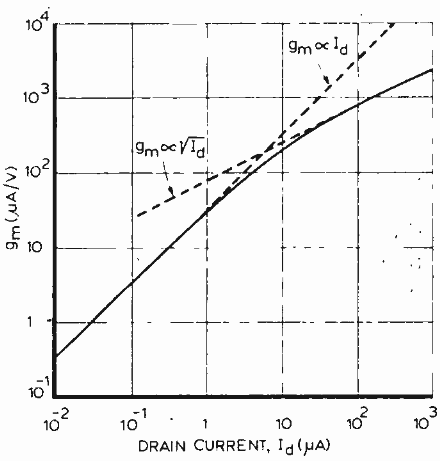

Fig. 5 *Measured mutual-conductance characteristic for an f.e.t.*

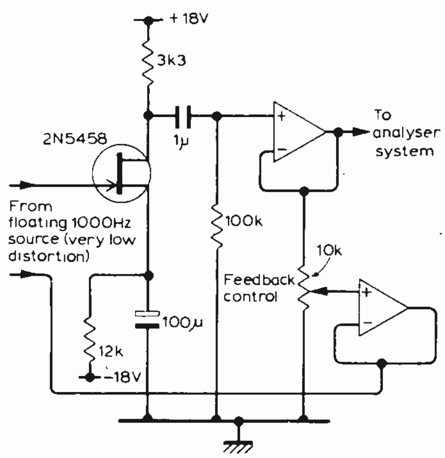

Fig. 6 *Test circuit for harmonic-distortion measurements.*

tage in millivolts. But no such delightfully simple result applies to the f.e.t.

F.e.t test circuit

The experimental circuit used for distortion measurements on an f.e.t. amplifier stage, with and without negative feedback, is shown in Fig. 6. No very expensive measuring equipment was used. The 1000Hz signal source consisted of a home-made low-distortion R-C oscillator feeding a Quad 50E amplifier, an air-cored tuned circuit purifying arrangement being connected to its floating output winding. The analyser system consisted of a parallel-T 1000Hz notch filter, whose output fed an R-C oscillator modified to function as a very highly selective amplifier feeding a c.r.o. For all measurements except second-harmonic, a passive notch circuit tuned to the second-harmonic frequency was inserted in front of the selective amplifier. Having tuned in a particular harmonic, the analyser system input was then switched to another oscillator, at the harmonic frequency, the known output of this oscillator being adjusted to give the same size of c.r.o. picture as before. With due care to avoid r.f. interference and hum problems, this set-up was both highly sensitive and of satisfactory accuracy. A test was done in which the signal source, at an enhanced level, was fed via a 3.3kΩ resistor straight to the integrated-circuit follower. The harmonic readings at the output of either integrated circuit, as the same fundamental voltage as before, were then negligible compared with those obtained with the f.e.t. in operation.

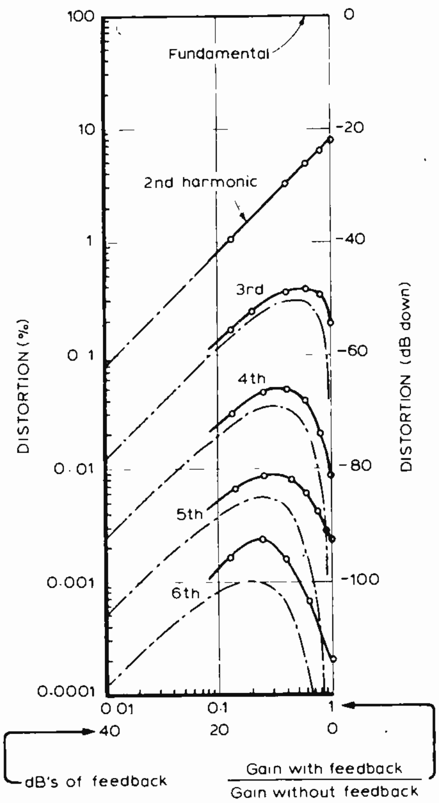

Fig. 7 *The full-line curves represent distortion measurements using the Fig. 6 set-up. The chain-dotted curves relate to calculated distortion, assuming an ideal parabolic f.e.t. characteristic as shown in Fig. 4. All curves are for a fundamental output level of 3V peak.*

Consideration of results

Fig. 7 shows, in full-line, the results of measurements using the Fig. 6 circuit, the chain-dotted curves being calculated from the formulae in Table 1. All curves relate to a fundamental output voltage of 3 volts peak. (A convenient fact is that, even with a large second-harmonic present, the peak value of the fundamental is accurately equal to half the peak-to-peak value of the total output waveform.)

The mean drain current in Fig. 6 is 1.55mA. The a.c. drain load is 3.2kΩ, giving a peak fundamental drain current, at 3 volts peak, of 0.94mA. Equation (16) above thus predicts a percentage second-harmonic distortion without feedback of $12.5\times(0.94/1.55)$ = 7.6%. It will be seen that the measured value is very close to this. As expected, however, the f.e.t. without feedback shows itself to be by no means ideally parabolic in transfer characteristic, so that appreciable amounts of higher-order harmonics are measured – though the largest of these, the third harmonic, is only 0.19% despite the quite large output level.

When feedback is applied, the magnitude of the measured third harmonic, conveniently expressed as a percentage of the constant fundamental output, at first rises, as more and more second harmonic is fed back into the input circuit to intermodulate with the fundamental voltage existing between gate and source and thus generate a sum component at third-harmonic frequency. As the feedback is further increased, the resulting improved linearity of the amplifier soon becomes the dominating influence and, when the amount of feedback is large, the third-harmonic output (at constant fundamental output) becomes directly proportional to $1/(1-A\beta)$. Similar effects occur also for the other harmonics, and it will be seen that the measured distortions, when the feedback is large, approximate closely to those calculated assuming a purely parabolic transfer characteristic. Thus, for an f.e.t. at least (though actually it applies also for a junction transistor), the main distortion mechanism for the production of third and higher harmonics, once plenty of feedback is applied, is the intermodulation one mentioned, rather than the existence of cubic and higher terms in the power series representing the transfer characteristic.

Conclusions

Some important conclusions that may be drawn from the above are:

- Even f.e.ts, used without feedback, generate high-order harmonics — and therefore, on programme, high-order intermodulation products.

- A small amount of negative feedback (e.g. 6dB) in a single-ended stage, though reducing the second-harmonic distortion, and also the total (unweighted) distortion, by about 6dB, will increase the higher-order distortion, and the quality of reproduction may well become worse as judged subjectively.

- If enough negative feedback is applied, all significant harmonics (and corresponding intermodulation products) can be reduced to a far lower level than without feedback, though the amount of feedback required to achieve this becomes larger the higher the order of the harmonic considered. (For example, referring again to Fig. 7, 16.5dB of feedback is sufficient to reduce the third harmonic to the same level as it has without feedback, whereas about 35dB is required for reducing the sixth harmonic to its no-feedback level.)

- The magnitude of harmonics of extremely high order will be increased by the application of negative feedback, no matter what practical amount of feedback is employed, but this is of no consequence if, when thus increased, they are, say, 120dB below the fundamental.

- Fig. 7, as already stated, applies at a particular output level of 3V peak in the Fig. 6 circuit, the peak drain current excursion being about 60% of the working drain current — in other words, it is high-level class A operation. When the signal level is reduced, the various harmonics fall off at different rates, as may be seen from Table 1. The percentage second-harmonic is proportional to $\hat{V}_{out}$, whereas the percentage fifth-harmonic, for example, is proportional to $\hat{V}_{out}^4$. On a logarithmic plot, as in Fig. 7, the effect of reducing the output signal level is that all the curves remain of the same shape, but each curve shifts downwards by a distance proportional to $(n-1)$, where n is the order of the harmonic, so that the spacing between the curves becomes wider. Thus at a reduced output level the higher-order harmonics rapidly become negligible.

(To be continued)

References

1. Baxandall, P. J., "Low-distortion amplifiers – Part 2", *J. British Sound Recording Association*, Nov. 1961, pp. 246-256.
2. Taylor, E. F., "Distortion in low-noise amplifiers", *Wireless World*, Aug. 1977, pp. 28-32.