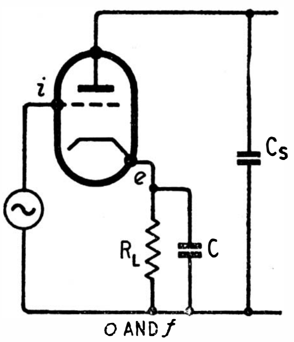
The Nyquist Diagram at Work

Dealing with Feedback over more than One Stage

By "CATHODE RAY"

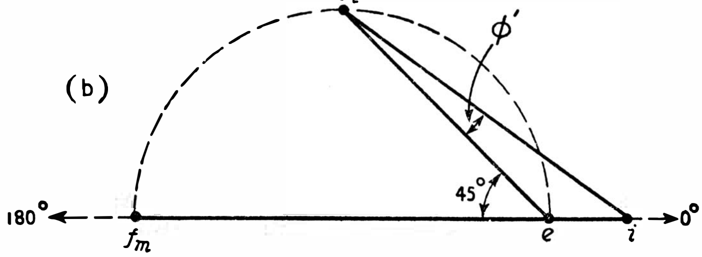
CONSIDERING that a moving-coil loudspeaker was patented in 1888 and transatlantic radio was achieved in 1901, it is surprising that it was not until 1934 that anybody pointed out the usefulness of negative feedback. Another surprising thing about it is how much has sprung from such an extremely simple idea. So much, in fact, that the hi-fi fan who chooses to design his own amplifier instead of just copying someone else's is liable to get into a daze. It was with the object of ameliorating his condition—and that of anyone else in trouble with negative feedback—that last month I expounded the Nyquist diagram as an aid to visualizing the workings of feedback circuits. There was only time then to apply it to very simple situations. So now I propose to go on to the more complicated cases where it really begins to pay.

But before doing so let us recapitulate. The basic idea of negative feedback is, as I said, so simple: some of the output voltage of the amplifier is put against the input voltage, so that to maintain the same level as without feedback the input voltage has to be increased until it is equal to the original input and the fed-back voltage combined. I say "combined," because although with perfectly negative feedback they would simply be added together, feedback can never be made perfectly negative at all frequencies simultaneously, and when the phase of the feedback is not exactly 180° simple addition fails. The thing can be dealt with by the usual methods for a.c., but a great help is a vector diagram, in which the original or net input voltage to the amplifier is shown as a fixed vector 1 unit long, at zero phase (denoted by pointing to 3 o'clock). The fed-back voltage is a vector that varies in length and phase with frequency, and the gross input required is equal to both together.



(a)

Fig. 1. (a) The essentials of a particular cathode-follower circuit, and (b) the diagram used for investigating its performance at high frequencies. f_m and f_t are particular values of f.



As an example, shall we take the one I gave last month to work out? It was a cathode follower, Fig. 1(a), in which the valve had a g_m of 6 mA/V and an r_a of 10 kΩ, R_L was 4 kΩ, and C was 0.002μF. C_s can be regarded as a short-circuit. The question was to find the "turning frequency" f_t (at which the total resistance and total reactance in the equivalent parallel or series circuit are equal) and the loss and phase shift caused by C at that particular frequency.

To facilitate comparison with last month's diagrams, I have used the same lettering. So *e* and *i* in Fig. 1(a) are the direct input terminals; and the unit signal voltage that is assumed to be maintained there, whatever the frequency, is represented in the vector diagram (b) by a line *ei* 1 unit long. The output terminals are *eo*, across which A units of signal appear, A denoting the voltage amplification. A fraction B of this output voltage is tapped off between terminals *ef* and fed back, this voltage being represented by vector *ef*. So the overall input terminals are *fi*. A special feature of the cathode follower is that *all* the output voltage is fed back (i.e., B = 1), so terminals *o* and *f* coincide.

Constructing the Vectors

The first stage of constructing the vector diagram in every case is to draw *ei* 1 unit long, pointing to 3 o'clock. The next is to calculate AB under perfect negative-feedback conditions and draw an *ef* vector that number of units long pointing in the opposite direction. In this case AB = A, and A can of course be calculated by the well-known formula derived from the valve equivalent voltage generator, which is expressed as follows:

$$A = \frac{-\mu R_L}{R_L + r_a}$$

The minus sign is to remind us that there is a phase reversal in the valve, if both output and input are reckoned from *e*. We were not told μ, but as it is equal to $g_m r_a$ it must be 60. So $A = -60 \times 4/(4 + 10) = -17.1$.

That would be the most likely method of calculation if C had not to be considered, but as it has we might as well adopt the equivalent *current* generator from the start, because being in parallel with the load it greatly simplifies calculation of parallel circuits. The reason I used the voltage equivalent just now is in case there are any doubters who need to be convinced that both equivalents give the same answer, and that it is purely a matter of convenience which is used. The current generated is $-g_m V_{ei}$, and as we have made $V_{ei} = 1$ it is equal to $-g_m$ in this case. The output voltage is developed by this current flowing through r_a and the load in parallel; and 10 $k\Omega$ and 4 $k\Omega$ in parallel is 2.85 $k\Omega$, denoted by R. Therefore A = − 6 × 2.85 = − 1.71 as before.

So to continue the diagram we draw *ef* 17.1 units to the left (that being the negative direction, in contrast to *ei*). To distinguish this particular *f*, corresponding to frequencies low enough for C to be ignored, let us call it f_m. The gain of the valve used as a cathode follower (i.e., with 100% negative feedback), denoted as usual by A′, is the ratio of output to gross input, so is represented on the diagram by the ratio of f_me to f_mi, or 17.1/18.1 = 0.945. Note that the output voltage is reckoned from terminal *f* in Fig. 1(a), that being the "earthy" output terminal of a cathode follower, so the output voltage is represented by *fe*, not *ef* as in anode-loaded amplifiers, and is positive. This corresponds to the well-known fact that in a cathode-follower stage there is no phase reversal, and illustrates how the lettered diagrams help one to take strict account of signs.

The same result can, of course, very easily be found by using the basic formula A′ = A/(1 − AB), in which A too must be reckoned as positive if *f* is the reference terminal:

$$17.1/(1 - [-17.1]) = 17.1/18.1 = 0.945.$$

Drawing the Diagram

Having got the position of f_m, we can draw the Nyquist diagram, because we found that for a single parallel combination of R and C it is a semicircle standing on f_me as diameter. We also found that the point representing the turning frequency f_t, at which the reactance of C is equal to R, is halfway along it, so that can be plotted and f_te and f_ti drawn in. Of course, the brighter boys wouldn't have bothered to draw ef_m or the semicircle at all; they would straightway have drawn ef_t at 45°, $A/\sqrt{2}$ long. All my rather lengthy rigmarole was for the benefit of any readers who were absent last month and started on this second article without a clue.

The actual value of f_t, for which you were asked, could have been worked out as soon as R, the resistance effectively in parallel with C, was found to be 2.85kΩ, for f_t is the frequency at which the reactance of C is equal to that; i.e., $1/(2\pi f_t \times 0.002 \times 10^{-6}) = 2{,}850$, from which $f_t = 27{,}900$ c/s. (The 10^{-6} is to bring 0.002μF to farads, as is necessary if f_t is to be in c/s rather than Mc/s; the bright boys would have left C in μF and R in kΩ and got f_t in kc/s.)

The last thing to be found was the phase shift and loss in A′ caused by C at frequency f_t. If there were no feedback, the phase shift would be 45° and A would drop from 17.1 to $17.1/\sqrt{2} = 12.1$; a loss of just on 30% or 3 dB. But in the cathode follower the phase difference between input and output is represented on the diagram by the angle between the corresponding vectors, marked ϕ'. When the diagram is drawn to scale (Fig. 1(b) is not) this angle turns out to be just over 3°—a remarkable improvement on 45°.

The new A′ is represented by f_te/f_ti of course, and you will have to draw the diagram on an enormous scale to detect any difference between it and the medium-frequency A′, given by f_me/f_mi. According

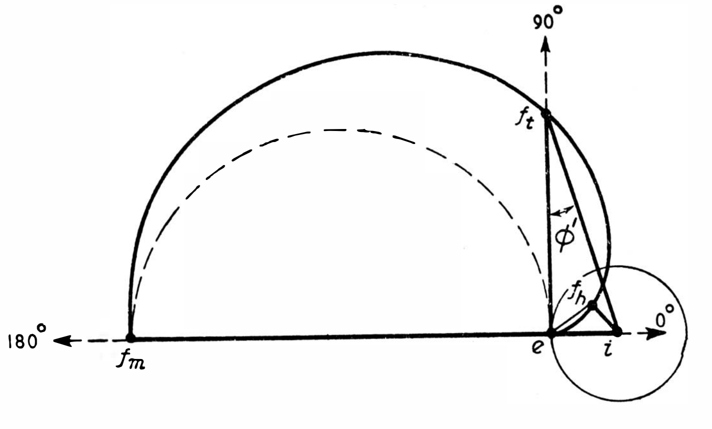


Fig. 2. Here the Nyquist diagram for a single CR circuit such as in Fig. 1. is repeated as a dotted semicircle, and the heart-shaped diagram for two such circuits is shown in full line. The small circle on the right marks the area within which feedback is positive.

to my rough calculation it is between 0.1% and 0.2% less, or say 0.01 dB—anyway, utterly negligible. This just shows why cathode followers are so popular in spite of their non-existent voltage amplification; a severe capacitive shunt across the load fails to pull down the output voltage appreciably, and has very little effect even on the phase angle. Lest I be accused of flattery, I should add that if the *gross* input voltage *fi* is kept up, instead of being allowed to drop from f_mi to f_ti, the signal current through the valve goes up accordingly and there may be a risk of overloading. This particularly applies to sudden pulses, containing very high frequencies, which can put cathode followers momentarily out of action (see W. T. Cocking in the March, 1946, issue.).

More Complicated Situation

That has been rather a long recap, even though some cathode-follower lore has been thrown in for good measure, so we must get on with the more complicated cases; in particular, feedback over more than one stage. The importance of this is that feedback over a single stage, while it may be delightfully simple to apply and effective in reducing distortion, does rather cripple the amplifier as an amplifier—as we have just seen. The effectiveness of feedback depends on the quantity 1 − AB, which also is the amount by which the original voltage amplification is divided. Now to be really worth having, −AB must be much larger than 1. One can then say that the effectiveness is approximately proportional to AB. As the books invariably point out, the basic formula A′ = A/(1 − AB) then becomes $A' \simeq 1/-B$, which means that roughly the amplification depends only on the fraction of output fed back, which can easily be made very constant. In other words, the voltage amplification is virtually independent of the amplifier itself, and of any changes therein caused by ageing valves or fluctuating supply voltage—always provided that its amplification remains high enough for $-AB \gg 1$. The consequences of this particular condition can be seen in the diagram by making *ei* comparatively very small.

In a single stage, applying such effective feedback destroys practically all its gain; but the same sacrifice in a three-stage amplifier still leaves the gain of two stages, which should be enough to reduce the input to a level at which feedback in preceding stages, if any, is unnecessary.

All right, then; what are we waiting for? Let's apply feedback to three stages! If I may be allowed to restrain the natural impatience a little longer, however, may I suggest that as a preliminary step we first draw a Nyquist diagram for two stages? To simplify the process let us assume that the stages are identical and that there are no couplings other than those deliberately provided.

Fig. 2 shows the now familiar dotted semicircle for one stage, from which the diagram for two can be derived. Take the turning-frequency point f_t, for instance. The second stage shifts the phase another 45°, making a full right angle, and it reduces the amplitude by another factor of $1/\sqrt{2}$, making it exactly half of the original ϵf_m. Filling in a sufficient number of such points to draw through, we get the full-line curve. Note that at f_t the phase angle ϕ' is still only a small fraction of the 90° lag that would be effective but for feedback. At a higher frequency still, f_h, we find that the input voltage $f_h i$ is actually less than it would be without feedback (ei). Consequently A′ is greater than A; that is to say the effect of feedback is to increase the amplification, which means that it is positive. At that frequency the gain curve will not only not fall off; it will rise to a peak. (Even so, note that the phase lag with feedback is less than it would be without.)

It is quite easy to mark on the diagram the boundary between positive and negative feedback. Positive means that fi must be less than ei; negative, that it must be greater. So the boundary is where fi and ei are equal, which is on the circumference of a circle with radius ei (which is 1 unit) and centre i. Clearly, feedback that starts off purely negative can never be made positive by a single CR circuit, but with two it is bound to be positive at all frequencies above a certain figure. If you try different amounts of feedback on paper, by varying the size of the "semi-heart" Nyquist trace, you will find that the greater it is the greater the phase lag (and therefore the higher the frequency) before feedback becomes positive. But when it does become positive, it does it more thoroughly.

At this stage it will be a good idea to draw some ordinary graphs of the magnitude and phase of the output against frequency, corresponding to the Nyquist diagrams we already have. In doing this we will follow what is now the standard practice with regard to the frequency scale: (a) using a logarithmic scale, so that 1 to 1 occupies the same distance as 0.1 to 1 and 10 to 100, and (b) making f_t the unit of frequency, so that the graph is of general application, and the scale readings only have to be multiplied by the particular value of f_t to adapt it to a particular case. (This practice is known as "normalizing" the scale.) Another advantage is that if the curves are turned around, left to right about $f/f_t = 1$, they apply in their entirety to the low-frequency cut-off caused by series coupling capacitors, where used. And as it is relative magnitude to output that matters, rather than actual voltage, we will show it in decibels. The result of this whole scale policy is that the shapes of the curves plotted will be standard. At least, that is so with no feedback. The shapes of the curves with feedback depend on how much is used.

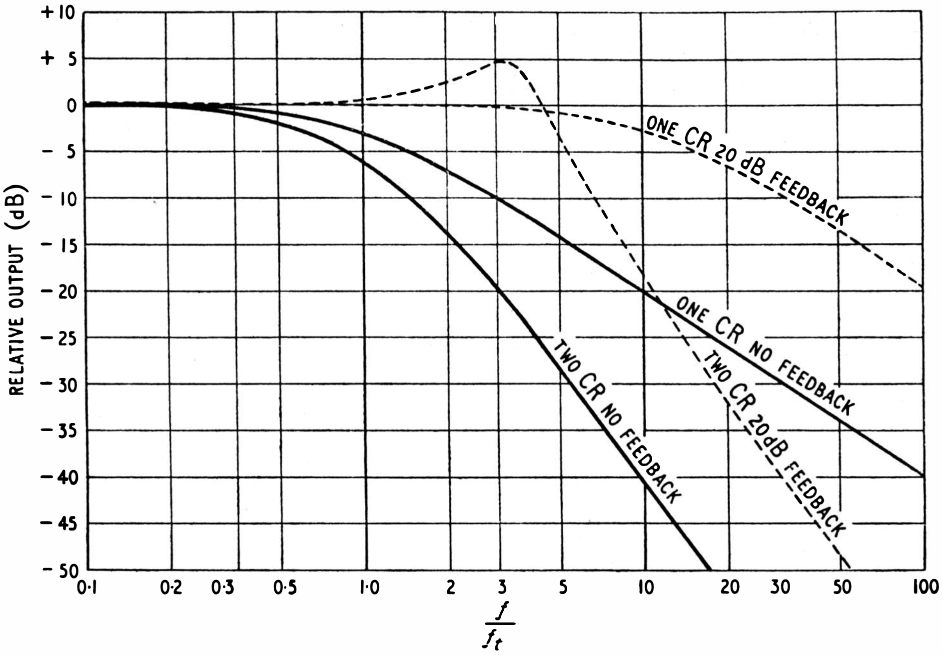


Fig. 3. Relative output plotted against frequency (relative to the turning frequency, f_t) for one and two CR circuits with and without 10:1 (= 20dB) feedback.

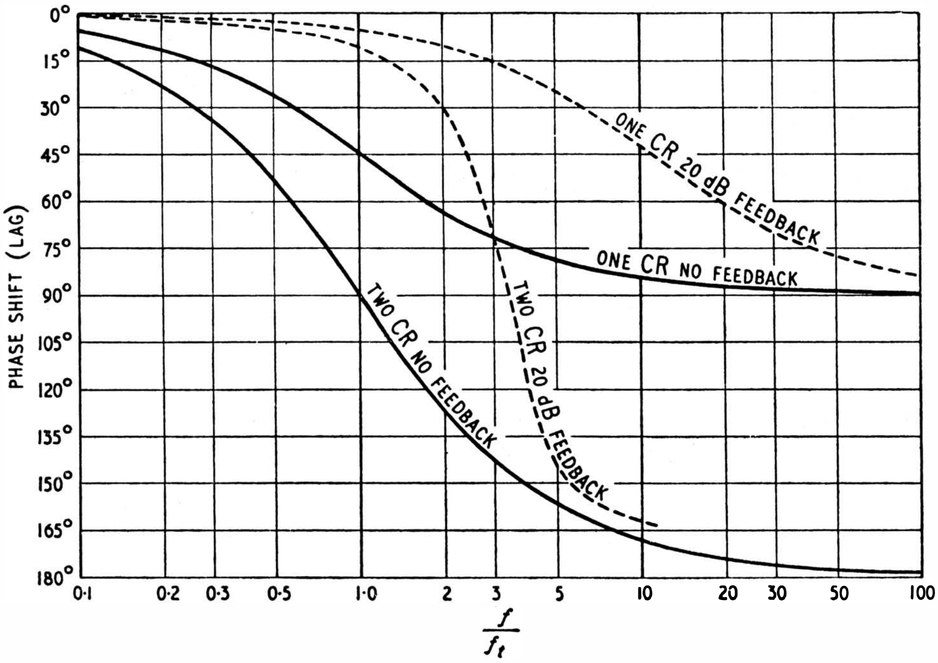


Fig. 4. Phase shift graphs corresponding to Fig. 3.

Fig. 3 shows in full line the relative-amplitude curves for one and two identical stages with capacitive top-frequency cut-off, and Fig. 4 the corresponding phase shift curves. (Note that if reversed to show low-frequency cut-off the phases would be *leading*, not lagging.) These curves show in a different way some of the things we already know; for instance, that at the turning frequency ($f/f_t = 1$) the loss is 3 dB for one stage and 6 dB for two, and that the corresponding phase lags are 45° and 90°. They also show that at very high frequency the lag approaches double these figures. More clearly than the Nyquist diagram, Fig. 3 shows that at high frequencies the loss tends to increase at a steady rate. This rate is 20 dB (1 : 10 ratio) per decade of frequency (1 : 10 ratio) for one stage, and 40 dB for two; but these rates are more often quoted as (almost exactly) 6 dB and 12 dB per octave (1 : 2 ratio).

Comparing Figs. 3 and 4 we see that these slopes are approached just as the 90° and 180° phase lags are approached. This is no accident; in fact it applies in the same proportion to any number of simple combinations of resistance and reactance or two opposite reactances (transmission lines and certain filters excluded). So if you look at a frequency characteristic curve of an amplifier (without feedback) in which the slopes are caused by such circuit combinations, and find that at some frequency the slope is at the rate of 12 dB per octave you are thereby provided with the important information that at that frequency the phase shift is 180°. If negative feedback were applied, it would at that frequency actually be positive, and if enough gain were left at the same frequency to make the fed-back voltage at least equal to the input voltage the amplifier would oscillate.

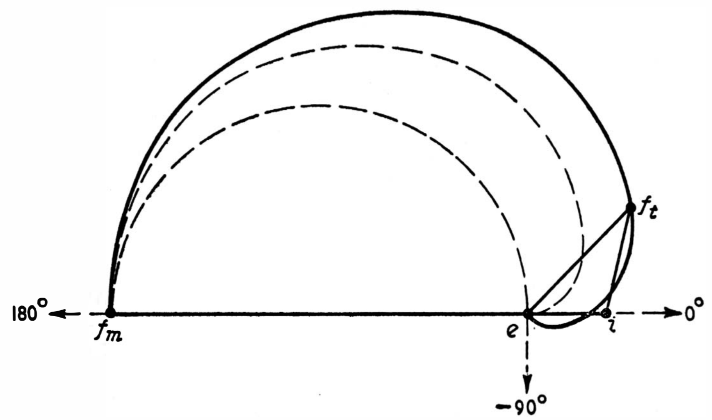


Fig. 5. Nyquist diagram for three CR stages, compared with those for one and two (dotted) repeated from Fig. 2.

There is no fear of that with one CR circuit; or even with two, for 180° shift is attained only at infinite frequency, at which the gain is zero.

Another fact is that at half the ultimate phase shift the dB curves have half their ultimate slope. It happens at $f/f_t = 1$, where the slope is 3 dB per octave with one CR and 6 with two. This might easily lead one to suppose that at one-third the phase shift the slopes would be 2 and 4 respectively, and so on, pro rata. I confess I thought so myself at one time, but on checking up mathematically found that this half-way proportionality was a fluke; the slope is not in fact proportional to the phase angle but to *n* times the square of the sine of one *n*th of that angle, *n* being the number of CR circuits.

However, you will not be so interested in how "C.R." came to see the light as in what happens when negative feedback is applied over the CR circuits. This is shown by the dotted lines in Figs. 3 and 4. They apply to 10 : 1 (= 20 dB) feedback; that is to say at f_m AB = 10, represented by making ef_m in Fig. 2 ten times *ei*. The dotted curves were derived from Fig. 2 (or rather a larger scale version of it) by measuring distances, but afterwards in another burst of enthusiasm I worked out formulae for them and plotted them again by computation. Fortunately the two lots agreed (when finally I got the formulae right!), but for initial study I unhesitatingly recommend the Nyquist diagram, even though it does mean a bit of work with drawing instruments. The procedure is of course the same as for the cathode-follower example. The phase angle with feedback, marked ϕ' in Figs. 1(b) and 2, is pretty obvious; but it may be as well to repeat that what are plotted in Fig. 3 (after conversion to dB) are the ratios of output/input ratio at the frequency in question to the same ratio at f_m. At f_t, for instance, it is represented by the ratio of f_te/f_ti to f_me/f_mi; viz., $(f_te \times f_mi)/(f_ti \times f_me)$.

It looks as if the dotted curves for one CR are the same as their full-line counterparts, except for being pushed higher in frequency. My original drawings help one to be more precise and suggest 11 times higher in frequency. This is 10 + 1, which leads one to guess that the use of *n*:1 feedback pushes the frequency characteristics $n + 1$ times higher in frequency. This time a mathematical check completely upholds the guesswork. It is a nice, simple thing to remember that feedback not only reduces gain $n + 1$ times but extends the frequency range (as regards cut-off and phase-shift) that number of times.

Rise in the Gain Curve

Unfortunately this simple rule applies only to one CR circuit, which is not very useful in practice except in connection with cathode followers. A glance at the two-CR curves shows that their relationships are decidedly less simple. The effect of feedback on the gain curve is to make it rise before plunging steeply—a characteristic that is quite useful if not carried too far. The rise is nothing to be surprised about, seeing we have already observed in Fig. 2 that two stages bring us within the positive-

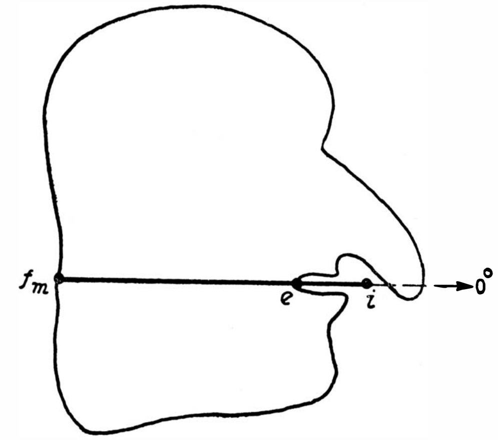


Fig. 6. This kind of Nyquist diagram, in which the oscillation point i is not enclosed, but which crosses the 0° axis beyond it indicates what is called conditional stability. Some Nyquist lines have very strange shapes.

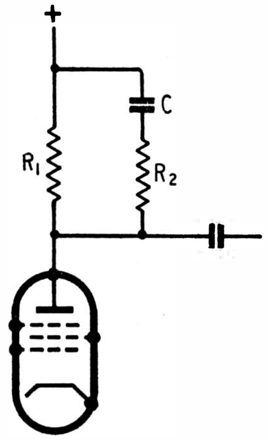
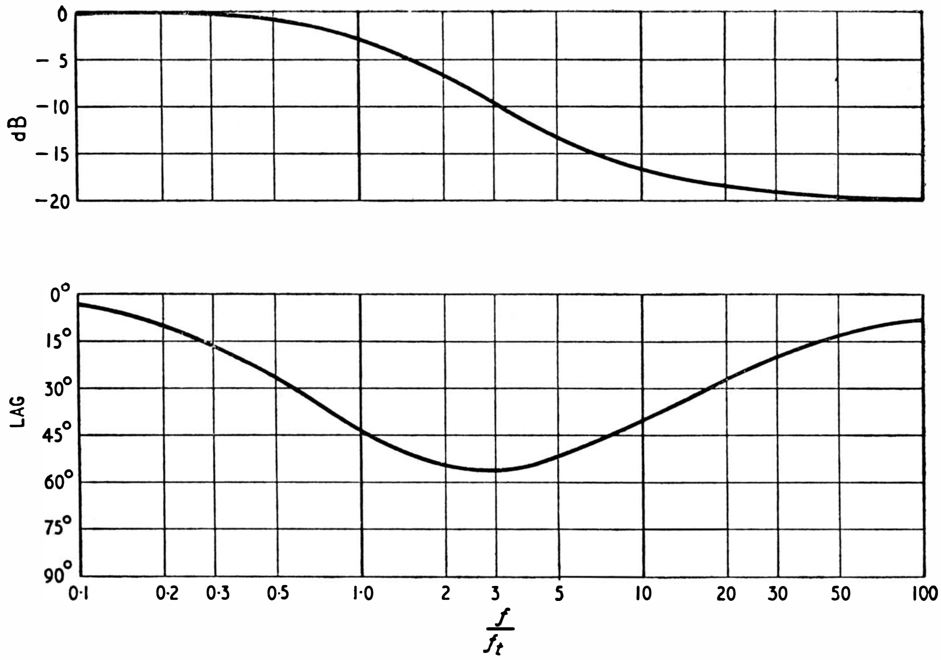


Fig. 7. One type of "step" circuit for cutting down gain without increasing high-frequency phase shift.



Right : Fig. 8. Typical characteristics of step circuit such as Fig. 7.

feedback circle. The more the feedback, the sharper the peak; but it can never go right through the roof and cause oscillation—with only two CR circuits. If this widening and peaking performance reminds us of the effect of over-coupling two resonant r.f. circuits, we may not be surprised to know that the mathematical formulae for the two things are somewhat similar in form.

As regards phase shift, we see that feedback postpones it to a higher frequency, but when the plunge comes it is all the steeper.

One could meditate still longer over Figs. 2-4, but must hurry on to the more practically important three-stage case. The Nyquist diagram (full-line in Fig. 5) can be derived from the two-stage in the same way as that was derived from the one-stage semicircle; both of those are shown dotted for comparison. The vitally unpleasant feature about the latest curve is that it passes through 180° phase shift (0° line) when it still has quite an appreciable fraction of the original (f_m) gain. It is an easy matter to calculate how much. When the total phase shift for three circuits is 180°, each (being identical) must be contributing 60°. The semicircle diagram, or Fig. 4 in relation to Fig. 3, show that at 60° the amplitude is halved; and halving three times leaves one eighth. So if as much as 8 : 1 (=18 dB) feedback is used over three CR circuits having the same f_t there will be oscillation. Such a situation is represented by the Nyquist curve passing through point *i*.

Double Crossing Curves

Last month I gave a rather qualified answer to the awkward gentleman I imagined to be asking what would happen if the curve passed through the 0° line *beyond i*—to its right. The reason for the slight hesitation was that some of the more complicated kinds of amplifiers are known to give Nyquist curves that cross the 0° line beyond *i*, *and then cross back again, also beyond it*, as in Fig. 6. The rule that Nyquist achieved fame by establishing is that if the whole curve is drawn, to include all frequencies from zero to infinity, and it *encloses* the point *i*, then oscillation is certain. The state of affairs represented by diagrams such as Fig. 6 is called conditional stability, which means that if the feedback is put into effect at the full force shown there will be no oscillation, but that if it grows gradually while heaters are warming up there probably will. It is unlikely that people who are reading this would find themselves keeping their amplifiers from oscillating by means of this sort of Nyquist curve, and if they did they would be well advised to think of some other way. For practical purposes we may regard the aim as being to keep the curve well to the left of *i* if it has to cross the 0° line at all. In other words, somehow we must increase the loss of the amplifier-feedback circuit—i.e., reduce AB—by the time the total phase shift amounts to 180°.

How to accomplish this aim is a big subject—too big a subject to start just now, and all I can do here is to refer readers to the practical procedure described in the March 1951 issue by Thomas Roddam.* Although something can be done by seeing that the stages do *not* all have the same turning frequency, the most useful weapon is the step circuit, which is a combination of a reactance with two resistances, as for example C, R_1 and R_2 in Fig. 7. The value of this device is that its amplitude curve doesn't continue to plunge for ever, like Fig. 3, but flattens out at a lower level. This reduction of slope is accompanied by a proportionate reduction of phase shift (Fig. 8). So what one gets at the high-frequency end is a substantial cut in gain without much phase shift. Which is just what one wants.

The need for such devices is all the greater because of the desirability of including the output transformer in the feedback loop. As regards high-frequency phase shift, a transformer is equivalent to two CR "stages," so even if there is only a single CR in addition it is enough to get one into difficulty.

Obviously, only just stopping an amplifier from oscillating isn't good enough; the slightest rise in mains voltage or change of load or valves or even a slight drift in component values might set it off again. Some margin is needed, and there should be a standard method of specifying how much.

One method gives it in the form of phase margin—the smallest angle between the Nyquist curve and *i*.

* A short summary of it is given in *Radio Designers Handbook*, 4th edn., pp. 363-365.