MORE DISTORTION

What Causes Musical Unpleasantness?

By "CATHODE RAY"

LAST month we dismissed frequency distortion as no longer a problem,* and concentrated on non-linearity distortion. The object was to decide, if possible, what the distortion figures given nowadays by makers of sound-reproducing equipment mean. They are usually "percentage harmonic distortion," but there is often a strong undercurrent of suggestion that they ought to be intermodulation. If they were, would we be any the wiser?

Well, after reviewing the elementary facts of harmonic production by non-linear equipment, I referred to an experiment I described in 1938 to demonstrate that the unpleasantness of non-linearity distortion is due not so much to the harmonics as to intermodulation products. These only occur when there are at least two frequencies present in the original signal, and the experiment was to apply two different frequencies and note that at an amplitude great enough for considerable harmonic distortion they sound quite clear when heard separately, but perfectly horrible when together, even if the total amplitude is then no greater. On the other hand they sound clear together if the amplitude is substantially reduced so that the distortion is slight. I mentioned that some doubt had been expressed whether it was safe to conclude from this one experiment that most of the unpleasantness of distortion is due to intermodulation. Even though at this much later date that is generally accepted, it seemed to me there would be no harm in looking into the matter more closely. And so (limiting our enquiry to musical programmes) we considered what it is that makes some combinations of sound frequencies blend smoothly and harmoniously and others harshly. Generally speaking, the smaller the numbers in which the frequency ratio can be expressed, the less conspicuous is the addition of the second frequency (assumed to be the higher one). The simplest of all (not counting 1 : 1) is of course the 2 : 1, or octave, and the higher frequency is then so concordant with the lower as to form a new starting point for the musical scale; for example, if the two frequencies are 100 c/s and 600 c/s (fundamental and sixth harmonic) the 600 can be reckoned in relation to the nearest octave above 100, namely 400, and the ratio of frequencies can be regarded as 600 : 400, or 3 : 2, a basic musical harmony. For this reason the even harmonics have to be higher than the odd before they are noticeably discordant; the lowest odd harmonic that sounds definitely discordant is the 7th, but the lowest discordant even harmonic is the 14th. After considering the relationship between the shapes of equipment transfer (input/output) characteristic curves and the resulting harmonics, we concluded that with properly designed and operated equipment, in which only second and/or third harmonics are appreciable, the harmonics alone wouldn't cause any harshness of tone, though they might perhaps shift the balance of tone upwards in frequency and also make it sound richer or thicker (according to personal reactions). In arriving at this conclusion we considered only the harmonics in relation to their own fundamentals. But how about the harmonic frequencies of different notes played at the same time? For instance, two of the notes in the common chord are in the frequency ratio 5 : 8 and the third harmonic of one and the second harmonic of the other are therefore in the ratio 15 : 16, roughly a semitone apart, and that is not a pleasant musical sound. But unless both second and third harmonics are comparable in strength with the fundamentals (which, if due to distortion, would *not* be typical of properly designed and operated equipment!) this discordant tone would be relatively very weak. I am told that musical composers are aware of the inadvisability of prescribing chords for strongly harmonic-producing instruments if they want to obtain a smooth-sounding result.

* Don't take that too literally, of course!

Experiment Repeated

And now we are ready to compare the results of purely harmonic distortion with what the same knowledge of musical harmony would lead us to expect the effects of intermodulation to be. Anybody who may have been so painstaking as to compare the account of my experiment given last month with the original in 1938 has no doubt been itching to accuse me of cheating. The original frequencies were given as 50 and 400; last month's, as 100 and 533. Well, perhaps I did cheat. Having recently repeated the experiment, I believe that if my original frequencies had been *exactly* as stated, in 8 : 1 ratio, they wouldn't have made such an unpleasant noise as they did. Using an exact frequency ratio, the two reproduced together by a distorting triode or pentode do not lose all trace of their individual character, as in the pre-war experiment, though they do sound much more distorted than simply their separately distorted selves added together. But if the ratio is not exact—say 50 c/s and 410 c/s—the result fully deserves my earlier description. As the upper frequency is varied, the unpleasantness goes through marked fluctuations, being sometimes very bad indeed and sometimes by comparison almost tolerable (though of course not by "hi fi" standards!)

This fits in perfectly with our musical ideas. With exactly 50 and 400 c/s, the second-order intermodulation products (as they are called), $f_1 \pm f_2$, are 350 and 450. These, of course, are the 7th and 9th harmonics of 50 c/s, and 400 c/s is the 8th, so the only difference as compared with harmonic distortion of 50 c/s alone is that these three harmonics are abnormally strong. In fact, this seems to be quite a good way of finding out what exaggerated upper-harmonic distortion sounds like. If the intermodulation were mainly third-order, $f_1 \pm 2f_2$, the frequencies created would be 300 and 500, the 6th and 10th harmonics, which ought to sound smoother than the musically discordant 7th and 9th. Fig. 1 shows the frequency pattern.

A critic complained that frequencies such as 50 c/s and 400 c/s are an unlikely basis for musical programmes. Had they been, say, 200 and 600 or even 150 and 400 the intermodulation products would have been the same frequencies as non-discordant harmonics. If, in order to demonstrate the objectionableness of intermodulation I deliberately chose frequencies such as 50 and 410, or 200 and 410, I would be wide open to the criticism that such ratios do not occur in music at all, except perhaps the kind of music in which the worst discords could pass unnoticed. So this time I chose 100 c/s and the rather odd figure of 533, because although these actual frequencies do not come on musical instruments with standard tuning, they are in the ratio (which is what mainly counts) of notes G and C, which very frequently do occur together in music, being the so-called dominant and tonic of the scale of C major. Unless both second and third harmonic distortions are grossly excessive, any jarring tone is almost or quite negligible. But the corresponding intermodulation product frequencies are 433 and 633, and 333 and 733, respectively (Fig. 2). These are out of tune with any notes on the musical scale, harmonious or discordant, so the unmusicalness of the sound is hardly surprising.

Here, then, we have two frequencies which are harmonious with one another and with one another's lower harmonics, but whose intermodulation frequencies are altogether unmusical by any standard. The listening test confirms these expectations. On the other hand frequencies could be chosen for the two input tones that would yield concordant intermodulation products, and this too is confirmed by one's ears. I don't know whether it would be practicable to compose music using only notes that could not, when sounded together, be distorted into discordant intermodulation tones, but I fancy composers would find it rather a serious restriction. And not only are the intermodulation tones introduced by distortion into typical musical programmes likely to be more discordant than the harmonics, but they are far more numerous. One has only to try to reckon the number superimposed on orchestral music to guess how the confused "muddy" sound of non-linear reproduction is caused. The doctrine that most of the audible unpleasantness of non-linearity distortion is due to intermodulation tones rather than harmonics is, I conclude, in general justified, at least for the lower-order distortion that is normal in reasonable apparatus.

Distorted Discords

One criticism that has been voiced is that modern composers like nothing better than a good hearty discord, and so discordant distortion products are not so serious as I made out. But (1) the amount of listening to music by that kind of composer is a small fraction of the whole, (2) even that kind of composer does not (except for a few obscure experimenters) write music for notes outside all recognized musical scales, and (3) in spite of what such music may sound like to some, the occurrence and nature of the discords is intended to be as composed and not as it may happen to result from chance distortion. A similar reply can be made to the criticism that intermodulation tones are generated in our ears because they are non-

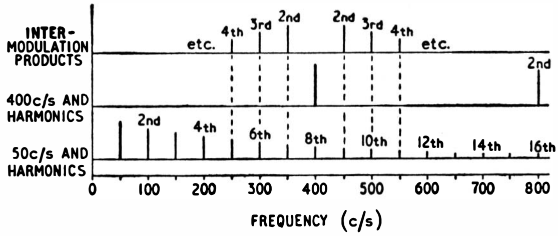


Fig. 1. This diagram shows, above a frequency scale, the harmonic frequencies of a 50-c/s signal, the same for a 400-c/s signal (only fundamental and second are within range), and the frequencies of the products of intermodulation between the two.

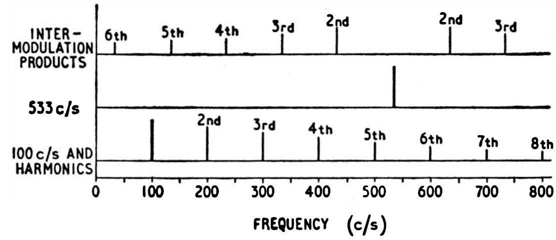


Fig. 2. Similar to Fig. 1, but with fundamental frequencies of 100 c/s and 533 c/s.

linear, and therefore distortion doesn't matter. But this ear distortion becomes prominent only when the sound is loud, so the distortion coming from reproducers, which doesn't disappear when we walk away and hear it more distantly, sounds unnatural.

Very well then, let us take the relative unpleasantness of intermodulation as established, and pass on to measurement of the distortion. And here there seems to be a tendency to argue that because intermodulation is the cause of the unpleasantness it is the thing that should be measured, rather than harmonics. It may quite possibly be true that it is better to measure intermodulation than harmonics, but this is not the argument to prove it. Remember, we can't measure unpleasantness as such; we can only look for something to which unpleasantness seems to be more or less proportional. If we find that unpleasantness is proportional to the percentage of intermodulation products, then it may seem natural to measure that. But it could be equally appropriate to measure percentage harmonics, even if they themselves contributed nothing to the unpleasantness, *provided that they were directly proportional to the intermodulation.* It is rather like voltage measurement. A difference of potential causes mutual electrostatic attraction, whereas it does not directly cause a magnetic field, but nevertheless voltmeters actuated by magnetic fields are far commoner than electrostatic voltmeters. The magnetic voltmeters are worked by current, which (according to Ohm's law) happens to be directly proportional to a voltage.

The relationship between harmonics and intermodulation is even closer than that between voltage and magnetic field, because harmonics are actually a particular kind of the same thing as intermodulation. This is a suitable moment for clearing up the numbering of these things. At one time it was quite usual to call the double-frequency harmonic the first harmonic. I believe musicians still do (they also often use the word "partial" for "harmonic.")

It was quite reasonable. But it was also rather awkward that the nth harmonic should be $n+1$ times the frequency, so to make the nth harmonic n times the fundamental frequency the fundamental is now reckoned as the first harmonic. Similarly the simple sum and difference intermodulation products, of frequency $f_1 \pm f_2$, were (and are) sometimes called the first-order intermodulation products; and this too was awkward because the kind of distortion causing them also caused what we now call *second* harmonic. So the rule is that the order number of the general intermodulation product $pf_1 \pm qf_2$ is $p+q$. With $f_1 \pm f_2$, p and q are both 1, so the order is 2. In this way the order of intermodulation is always the same as that of the harmonic produced by the same kind of distortion. If you didn't at first see my point about the vast number of intermodulation products compared with harmonics, it should be clearer now. Seventh-order distortion of two frequencies comprises only two seventh harmonics—$7f_1$ and $7f_2$—but all these intermodulation products: $6f_1+f_2$, $5f_1+2f_2$, $4f_1+3f_2$, $3f_1+4f_2$, $2f_1+5f_2$, f_1+6f_2, $6f_1-f_2$, $5f_1-2f_2$, $4f_1-3f_2$, $3f_1-4f_2$, $2f_1-5f_2$ and f_1-6f_2. Both mathematical calculation and practical test show that this distortion also produces fifth, third and first harmonics and intermodulation products. So imagine the result with a full orchestra playing!

Next, see what happens to the distortion products just listed when f_2 becomes equal to f_1. The only harmonic frequency, of course, is $7f_1$ (because $7f_2$ is the same). All the sum intermodulation products also boil down to $7f_1$. The difference products are $5f_1$, $3f_1$ and f_1, which also were there before. So harmonic distortion is not an entirely separate subject from intermodulation, but can be regarded as a special case of it. No wonder then if there is a close numerical relationship between figures for harmonic distortion and those for intermodulation.

It would take too long to go through all the calculations here and now to show what the relationship is, because it depends on the kind of distortion. But the data have been clearly tabulated in the article by Callendar and Matthews I mentioned last month. There are also some very interesting comparisons between calculations and experimental results in a paper by W. J. Warren and W. R. Hewlett*. All I can do in the space left is to outline some of the main principles.

We have already seen that the relative strength of each harmonic produced by distortion depends on the shape of the transfer characteristic of whatever is causing the distortion. The same goes for intermodulation products. And I have mentioned that the shape that generates, say, second harmonics, is also the shape that generates second-order intermodulation. Conveniently enough for the memory, it is the second-power (or square-law) shape. What does that mean? Well, suppose we take first of all a linear device, say a resistor. The equation stating the relationship between the voltage applied and the current flowing through it is commonly known as Ohm's law: I=E/R. In algebra, however, it is a custom to use small letters for variables and capitals for constants. The whole meaning of Ohm's law is that however the current and voltage may vary, the ratio of the two—the resistance—is constant. So we can write the same thing

$$i=\frac{1}{R}e$$

and because 1/R is the conductance, for which the usual symbol is G, we can make a neater job:

$$i=Ge$$

If we plot a graph of i against e, by choosing some fixed value of G and then choosing various values for e to give corresponding values for i, giving points to join up into a line, we find that the line is always a straight one. That is what we mean when we say that the resistor is *linear*. We can alter the slope of the line by choosing a different value for G; that would mean a different, but still linear, resistor. We could also shift the line bodily (which would be useful for approximately imitating the nearly-linear part of a valve characteristic) by adding another constant, say I_o, to stand for the current flowing when there is no voltage:

$$i=I_o+Ge$$

Our e stands for any value of input voltage varying in any way at all, but supposing we use a definite kind of input voltage, with a sine waveform, we can substitute for e the equation of that waveform, usually written $e=E\sin\omega t$, where E is the peak voltage and ω is 2π times the frequency. The result of the substitution is

$$i=I_o+GE\sin\omega t$$

from which we see that the current also has the same sine wave form and frequency. What we have done is to prove that a linear device—resistor, valve, amplifier or what not—is distortionless (as if we didn't know!).

To study non-linear devices we try to find an equation which, when graphed, closely imitates the characteristic curve of the device. One of the commonest shapes, especially where valves are used, is the one that bends increasingly in one direction, as in Fig. 3. This can be imitated by adding a square or second-power term to the equation, with its own constant to decide the amount of curvature:

$$i=I_o+G_1e+G_2e^2$$

When the signal waveform is substituted for e the new term becomes $G_2\,(E\sin\omega t)^2$, and this is equal to $\frac{1}{2}G_2\,(1-\cos 2\omega t)$, which shows that a signal of twice the frequency (i.e., the second harmonic) is produced. To imitate the device's curve more accurately it is usually necessary to add some higher even-number terms, and each brings in its own harmonic and also harmonics of all the lower even numbers.

If the curve bends over equally at both ends it

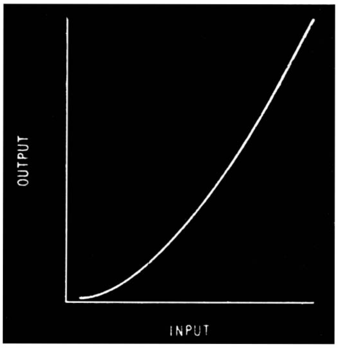


Fig. 3. Type of transfer characteristic giving rise to even-number distortion.

* An Analysis of the Intermodulation Method of Distortion Measurement," *Proc. I.R.E.*, April 1948, p. 457.

can be shown in a similar way that odd-number terms are needed in the equation, and odd harmonics are produced.

The same procedure is adopted in studying intermodulation, except that e must be (at least) two sine (or cosine) waves of different frequencies. The algebra and trigonometry needed to reckon up all the frequencies in the output, and the amplitudes of each, becomes really formidable, and that is why it was very kind of Messrs. Callendar and Matthews to go through it all and present the results in convenient tables. They show that the relationship between the powers of e in the characteristic equation and the harmonic frequencies produced by the corresponding distortion holds good for intermodulation products—that an even power causes intermodulation products of that order and all lower even orders, and similarly for odd powers.

Distortion Measurement

The fact I have been leading up to in all this is that if the equation of a distorting device's transfer characteristic is known, the amplitude of every harmonic and intermodulation product follows (provided, of course, that we have the skill and patience to deal with all the necessary calculation!). So there is, corresponding to any combination of harmonics resulting from a given combination of input signals, one particular combination of intermodulation products. And vice versa. Theoretically at least, if either harmonics *or* intermodulation are known, both are known. So theoretically at least it doesn't matter which is measured. There is a fixed rate of equivalence between the two.

But that doesn't mean that for every 1% harmonic distortion the intermodulation distortion is some fixed number of %. It isn't nearly as simple as that. In general, there is a different ratio between harmonics and intermodulation for every order (second, third, etc.), and that number is not fixed but depends on the respective amplitudes of the two or more input frequencies, and on the amount of distortion of other orders. The reason for this last is that the amount of second-order distortion (say) depends not only on the second-power term in the equation but also all higher even-power terms. This complication drops out if the distortion is exclusively second or third, as approximately it often is. Another complication can be avoided by always using the same ratio of signal amplitudes for intermodulation testing; a commonly-used ratio is 4 : 1. If the single signal used for harmonic testing has the same peak value as these two combined (i.e., 5 times the amplitude of the weaker) then with second-order distortion alone each of the two intermodulation products, reckoned as a percentage of the weaker signal forming its " carrier wave," is 1.6 times the percentage harmonic distortion. With third-order distortion alone, the corresponding ratio is 1.92. And if both " sidebands " are counted, these two figures are doubled. Fortunately these ratios are not very much affected by reasonable amounts of higher-order distortion, and practical tests with the 5 : 4 : 1 signal ratio show that the intermodulation product percentage of any order is usually 1.5-2 times the same-numbered harmonic percentage. Because the carrier wave is only one-fifth of the amplitude used for harmonic testing, however, the intermodulation product itself is smaller than the corresponding harmonic, so it is not really correct to say (as American writers do) that intermodulation measurement is more sensitive.

All this is on the assumption that there is no frequency distortion. Of course if the various frequencies are amplified by different amounts in the " device," that upsets the calculations accordingly.

For the sake of simplicity, everybody wants to sum up the distortion in a single number. But looking at Figs. 1 and 2 again we may well ask how this can be done. Even single-signal harmonic measurement is liable to produce a considerable number of harmonics of assorted amplitudes, and intermodulation measurement yields vastly more. Is there any way of combining those groups of percentages into one, in such a way that it gives a fair indication of the unpleasantness of the distortion?

It would be very nice if there were, and several ways have been proposed, but I am afraid that the answer is, if not an outright negative, at least doubtful. One of the most popular schemes of measurement is to apply a single tone at the input, measure the total output (fundamental plus harmonics due to distortion), and then insert a bridge filter between output and meter to stop the fundamental completely, so that what is measured is the total harmonics. The ratio of 100 times the second reading to the first is " percentage total harmonics." This scheme is popular because it can be worked with comparatively simple apparatus and gives a single figure. But unfortunately that figure is not a *fair* measure of unpleasantness. Although the subject is full of controversy, one thing universally agreed is that a given amount of third harmonic distortion is worse than the same amount of second, and that the high harmonics are worse still. To make the " total " figure take this into account it was proposed in 1936 that harmonics should be measured separately and each multiplied by $n/2$ before being combined. For the second harmonic n is 2, so its reading is unaffected; the third is multiplied by 3/2; the fourth 2; and so on. By the way, whether the individual harmonics are " weighted " like this or not, they must not be just added together to give the total; as I explained in " Total Power " (March, 1952) when adding up a number of simultaneous voltages or currents it is necessary to square each, add them all together, and take the square root of the result.

According to D.E.L. Shorter of the B.B.C.*, this system still doesn't give enough weight to the unpleasantness of the high-order distortion, and he reckons that multiplying each harmonic reading by $n^2/4$ lines up better with listening tests. You can see, of course, how difficult it is to discover exactly how much worse one kind of distortion sounds than another; for one thing it probably depends a good deal on the kind of programme being heard. So any weighting system is rather arbitrary. I doubt whether anyone would be prepared to swear that fourth harmonic is either 4/3 or 16/9 times as bad as third, or even that it is equally bad. And besides the extra calculation, measuring all the harmonics separately necessitates much more expensive apparatus, especially for the Shorter weighting, in which the very high harmonics are multiplied so much that one has to be able to measure accurately very small percentages of them.

How about intermodulation measurements? They

* " The Influence of High-Order Products in Non-Linear Distortion," *Electronic Engineering*, April 1950, p. 152.

are even more controversial. The most popular method (again, because it requires simple apparatus and gives a single reading) applies a strong low-frequency signal and a quarter-strength high frequency signal, and measures the total of the "sidebands" around the latter; e.g., those shown on the top line in Fig. 1. The procedure has been described in *Wireless World* by Thomas Roddam (April 1950) and E. W. Berth-Jones (June 1951). It comes under the same criticism as the total harmonic distortion method, over which it seems to have no very obvious advantages.

Another system, called the C.C.I.F. method, varies the frequencies of both input signals in such a way that one signal is always a certain number of c/s (say 1,000) more than the other. The frequency of the second-order intermodulation product $f_1 - f_2$ is therefore constant and hence relatively easily measured. This method is very highly spoken of in some circles, but since it indicates only second-order distortion, it presumably pronounces a push-pull amplifier having strong third-order distortion as absolutely perfect. To my mind this is a fatal objection.

The simpler methods have their uses (e.g., production tests of units having possibly varying amounts of similar distortion), so long as one doesn't regard them as unpleasantness meters. For thorough investigation it seems to be necessary to have a wave analyser for separately measuring every distortion product, and preferably to supplement it by visual examination of the transfer characteristic and of the output when the fundamental has been removed. For most purposes I should say that harmonics are enough, but there is an exception if one wants to know what the distortion is like near the upper frequency limit, because then the harmonics are all "off the map," but two signals inside the limit can still intermodulate to give a distortion product right inside the audible range.

Nobody would be more pleased than I to be able to hand out a simple cut-and-dried solution to this problem of distortion measurement. Perhaps some painstaking and well-provided organization will give a team of research workers a year or two to find out what reasonable conditions and method of test take into fair account every cause of unpleasantness of distortion.

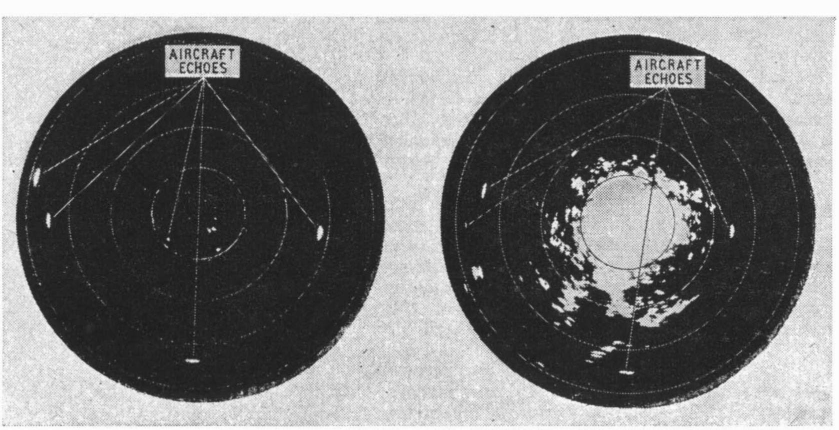
AIRFIELD RADAR DEVELOPMENTS

Crystal-controlled and with Permanent-Echo Suppression

THE clutter of permanent echoes (p.e.'s) familiar to all operators of radar equipment, and which is particularly troublesome on airfield radar screens, can now be successfully eliminated by an ingenious cancellation system embodied in the latest Type S232 airfield radar introduced by Marconi Wireless Telegraph Company. Known as "Moving Target Indicator" (MTI) it provides permanent-echo suppression better than 46 db. Another unusual feature is that it is crystal-controlled throughout, which to a large extent accounts for the good p.e.-suppression.

Briefly, the operation of the equipment is as follows:—the output from a crystal-controlled reference oscillator on 5.625 Mc/s is mixed with an harmonic of another crystal-controlled oscillator and the beat frequency amplified and multiplied to give the final output frequency, which in this case is in the frequency range 500 to 610 Mc/s (50 cms). The output power is between 50 and 60 kW at a pulse length of 2 to 4 μsec as required and at a pulse repetition frequency of 500 to 800 c/s.

The received (echo) signals after conversion to an intermediate frequency of 45 Mc/s together with the eighth harmonic of the 5.625-Mc/s reference oscillator (also 45 Mc/s), are fed to a homodyne detector. The output from this detector is therefore proportional to the difference in phase of the two input signals. As the phase of the reference oscillator is fixed, echoes from stationary objects will have the same phase difference on all successive echoes, but those from a moving target will have a continuously changing phase. It is only necessary to compare the homodyne output produced by successive echoes in order to determine whether an echo is moving or not. A special liquid delay line is used for this purpose and in this device identical signals resulting from permanent echoes cancel out and only those whose phases have changed between successive echoes appear in the output circuit. Here they are rectified and fed through a video amplifier to separate cathode followers and thence by coaxial cables to p.p.i. display consoles. Up to eight p.p.i. display units can be used with one aerial head so that the equipment can be used for long-range, short-range or segmental viewing simultaneously in several different places. As demonstrated by an experimental equipment installed at London Airport, this radar is capable of detecting aircraft at ranges of from ¾ to 100 miles.



P.p.i. displays showing permanent-echo suppression (MTI system) with the Marconi Type S232 airfield radar. On the left MTI switched on, on the right, switched off. Range markers at intervals of 5 nautical miles.