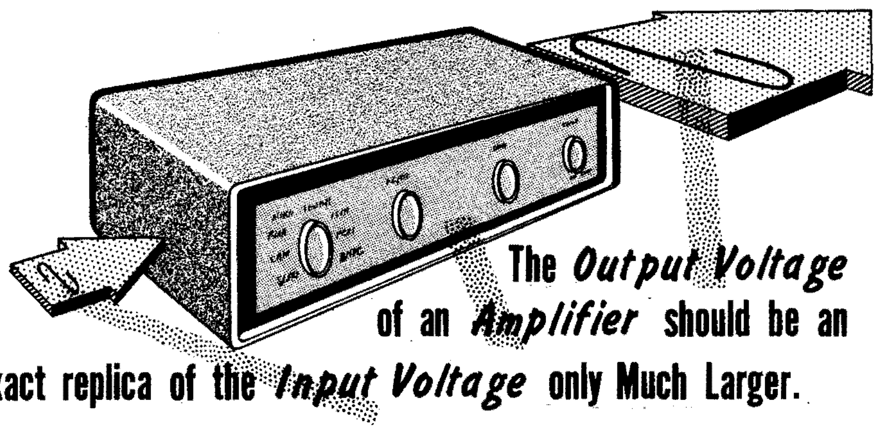
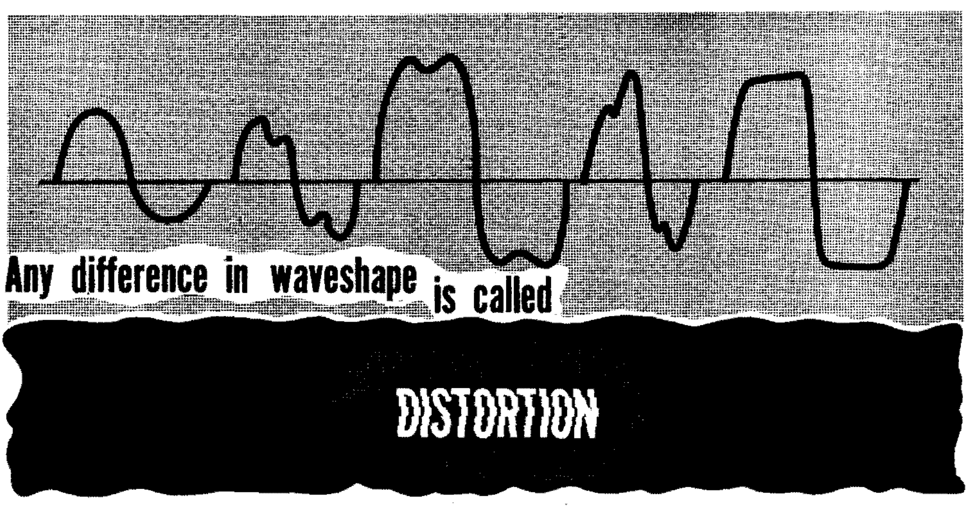
Harmonic Distortion

The basic purpose of an amplifier is to amplify the audio input voltages without distorting them in any way. The output audio voltage should be an *exact* replica of the input voltage, except that it is very much larger—1000 or even a greater number of times. Practical amplifiers never *completely* achieve this exactness, although they may get very close to it. There is always some distortion that makes the output waveform a trifle different from the input waveform.

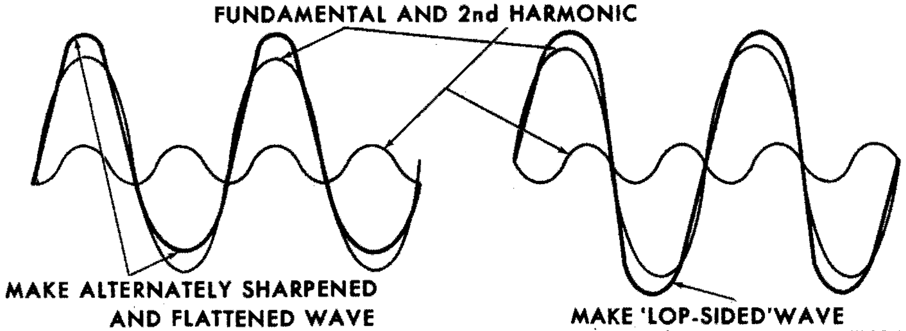
When the word distortion is used without qualification, it is usually taken to mean the kind of distortion due to curved or nonlinear characteristics in the amplifier. The fact that a change in audio voltage at the input is not accompanied by an exactly corresponding change at the output, at different points on the waveform is a form of distortion.

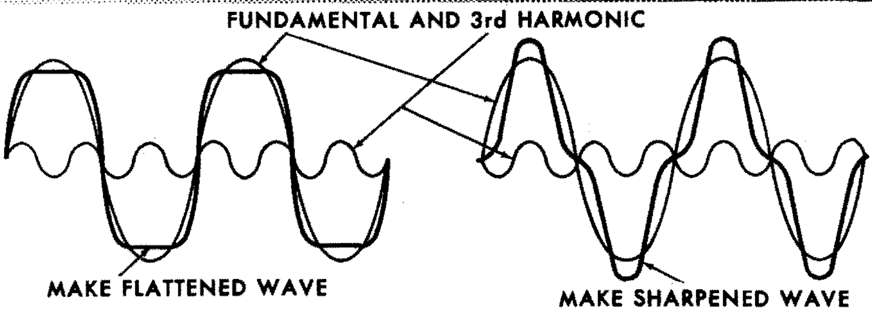




Harmonic Distortion (contd.)

Distortion due to 2nd or even-numbered harmonics





Distortion due to 3rd or odd-numbered harmonics

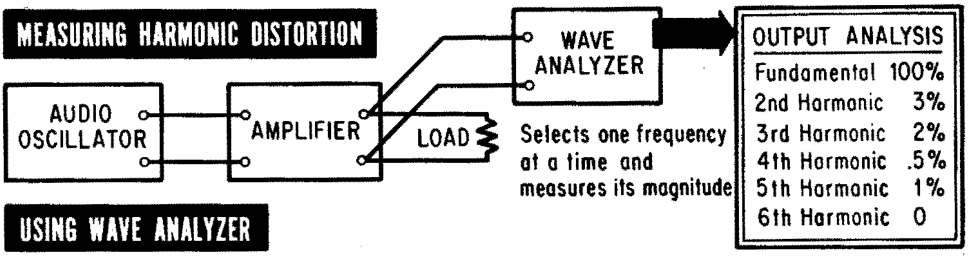
If we consider what this kind of curvature does to the amplification of a wave of single frequency, we can see how it introduces harmonic distortion —the presence of overtones of the fundamental frequency that are not present at the input.

A sharpening or flattening of both tops and bottoms of the waves is equivalent to the addition of third or other odd-numbered harmonics of the original frequency. A flattening at one peak and a sharpening at the other is equivalent to the addition of second and other even-numbered harmonics of the original frequency. If the waveform goes lopsided, that is, the upward slope is different from the downward slope, this is also due to second or other even-numbered harmonics added to the original frequency.

These are the principal kinds of harmonic distortion. Any real example will usually consist of one or a combination of two or more of them.

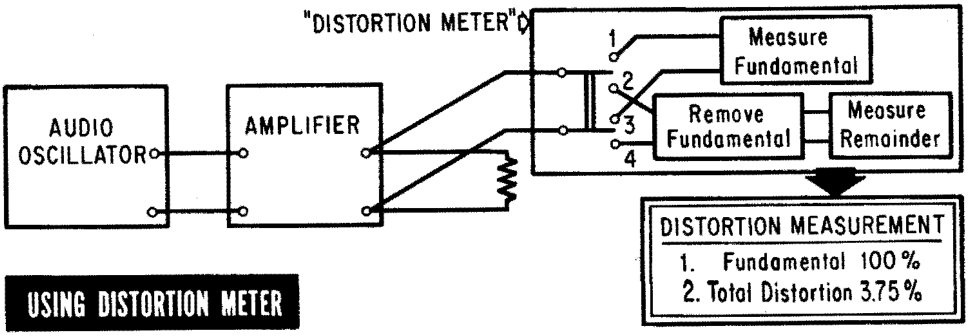
Measurement of Harmonic Distortion

The presence of these harmonics can be measured by using a *wave analyzer*. It is quite a complicated and expensive piece of measuring equipment that has a frequency selective amplifier that permits it to measure the amplitude of any particular frequency in a composite output waveform. By setting the frequency dial first to the fundamental frequency of a pure sinusoidal input and then to successive harmonics, the component of each harmonic in the output waveform can be measured to find out how much total distortion is produced.



The use of the wave analyzer is rather a long-winded method, so a simple distortion measuring set is usually used to give the answer quite quickly. This method uses another kind of frequency-selective filter to eliminate the fundamental. Two positions are provided on the switch: one for measuring the amount of fundamental and the other for measuring the total amount of audio after the fundamental has been removed. This gives a quick and ready means of measuring the total distortion.

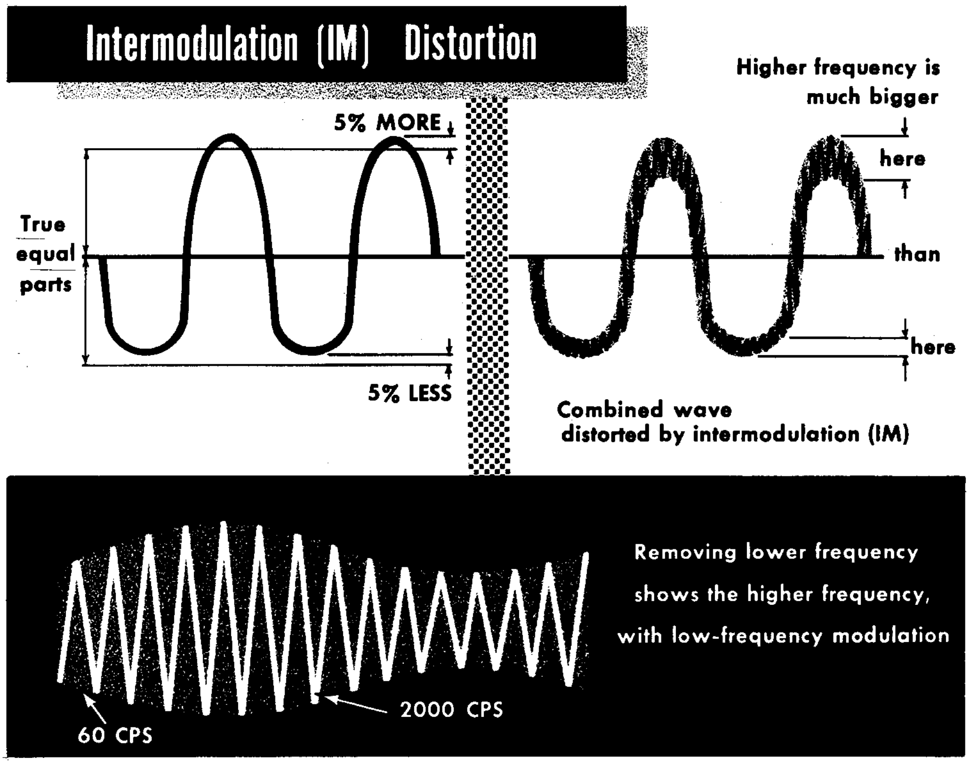
In the early days of audio amplifiers 5% harmonic was considered a good figure of distortion. At that time, tests were made which showed that human hearing could barely detect 5% of second harmonic. If the distortion is third harmonic, about 1.5% is just audible. At higher harmonics lower percentages become audible. Modern amplifiers produce harmonic distortion figures that are a fraction of 1%. According to the tests just described, this distortion should be completely inaudible.



Intermodulation Distortion

Low percentages of harmonic distortion may be inaudible, as such, but the same curvature in the amplifier characteristics causes another kind of distortion, called *intermodulation distortion* (IM for short). The effects of this kind of distortion can be audible when the harmonic distortion is not. There are two basic kinds of intermodulation distortion.

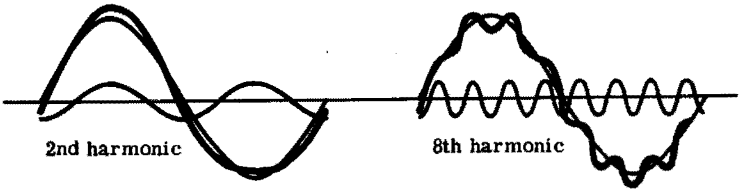
The first kind occurs because the amplification changes during a wave as well as introducing harmonics of this wave. This change in amplification will modulate or change the amplification of higher frequencies present in the same composite audio wave and this modulation of the higher frequencies is what becomes audible.



Suppose that a 60-cycle wave has 5% of second harmonic. This will mean one-half of the wave will get amplified by 5% more, while the other is amplified by 5% less. If the amplifier is also called upon to handle a 2000-cycle wave of much smaller magnitude than the 60-cycle wave, this will also get amplified by 5% more on one peak of the 60-cycle wave than it does on the other peak of the 60-cycle wave. Thus the 2000-cycle wave will be fluctuating in amplitude at the rate of 60 cycles.

Intermodulation Distortion (contd.)

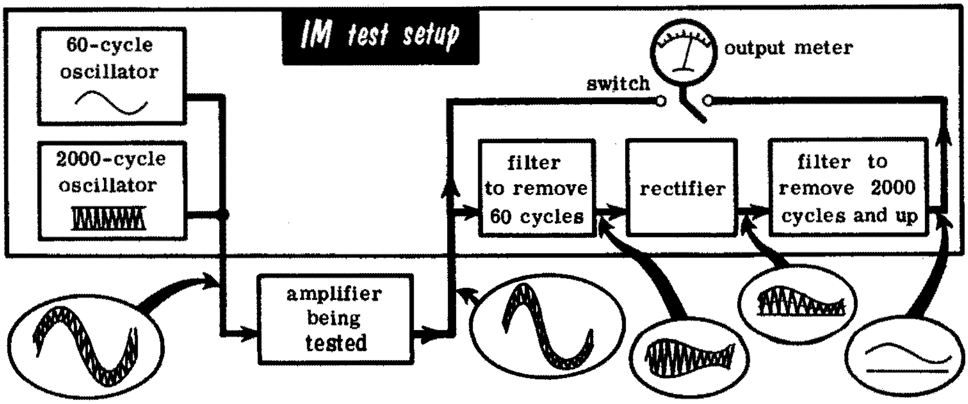
Measuring intermodulation



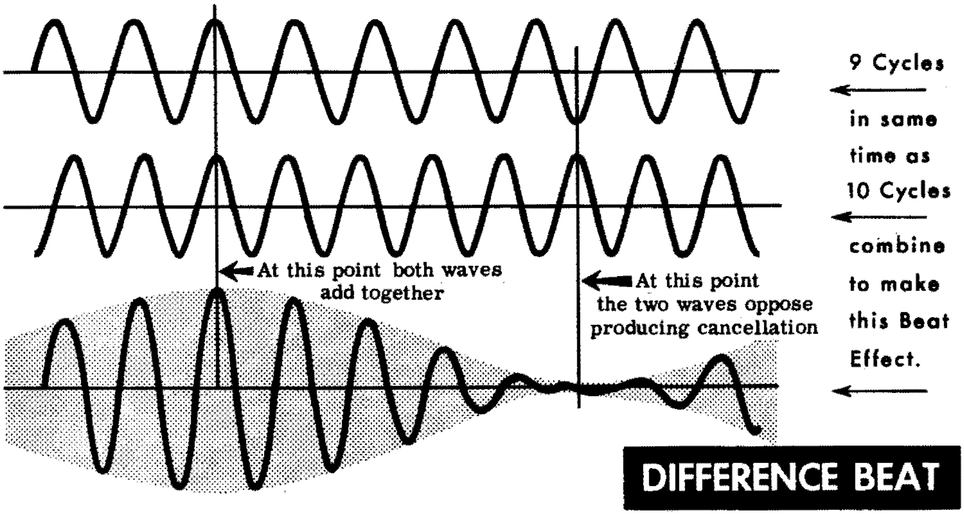
Both harmonics are same magnitude, but the higher numbered one has much bigger effect on the shape of the combined waveform.

This effect on the 2000-cycle tone is quite audible as a dithery modulation of the tone. It is often noticed in organ music accompanying deep bass tones. If the curvature in the amplifier is of a kind that produces higher numbered harmonics than second or third, it will also produce increasing amounts of intermodulation because the smaller amount of higher frequencies added to the basic fundamental tone produce more noticeable changes in the waveform.

This kind of intermodulation distortion is measured by using two tones, usually a combined audio voltage at two frequencies, such as 60 and 2000 cycles, with the voltage at 60 cycles 4 times that at 2000 cycles. The combined waveform is fed into the amplifier and a special distortion measuring set applied to the output waveform. First the waveform is fed through a filter that removes the 60-cycle component completely. This leaves the 2000-cycle component which fluctuates in amplitude if intermodulation is present. This 2000-cycle waveform is then rectified, which gives a d-c output with the fluctuation riding on it. The d-c component can now be readily removed by passing the wave through a blocking capacitor, and the fluctuation is measured as an audio voltage. By careful calibration of the whole setup the amount of fluctuation at the output can be measured as a percentage of the total output waveform.

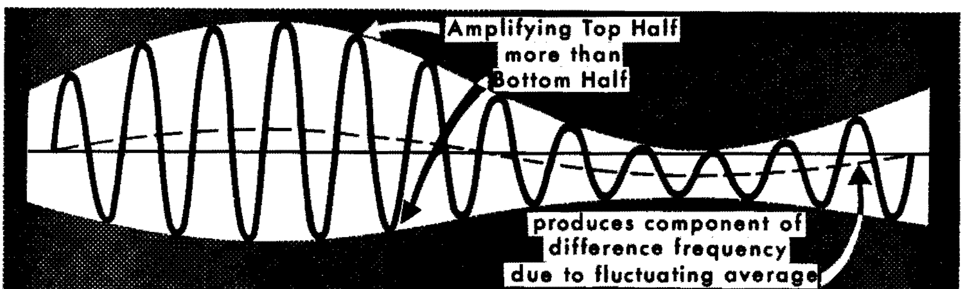


Intermodulation Distortion (contd.)



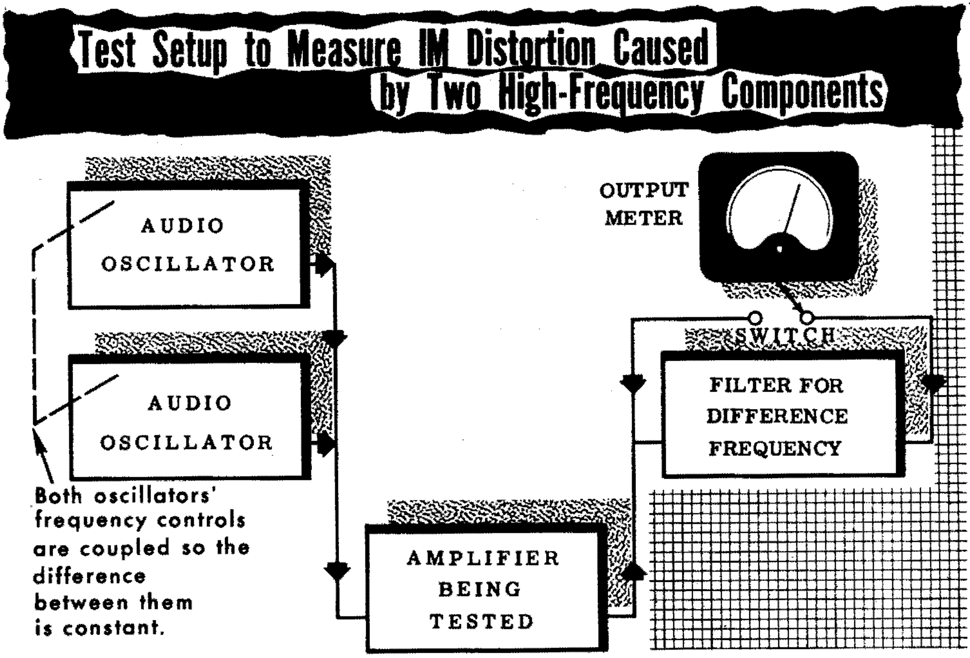
The second kind of intermodulation distortion is caused by two relatively high frequencies producing a combined tone at a lower frequency. If two frequencies are very nearly the same, the combined waveform will gradually move in and out of phase at a rate dependent on the difference between the two frequencies. At one point, the two frequencies will add, producing a double amplitude, while at a point a little later, the two frequencies will subtract, giving an amplitude which is the difference between the individual amplitudes.

If this combination is applied to an amplifier without distortion, the waveform will be faithfully reproduced as would any other waveform. If the amplifier introduces any asymmetrical distortion, however, the upper part, at the peak in the combined waveform due to addition of the two components, will be amplified more than the lower part of the same peaks. This is equivalent to adding a component of the low frequency corresponding to the difference between the individual test frequencies.



Intermodulation Distortion (contd.)

Suppose one of the frequencies is 4000 cycles and the other one 4200 cycles; the difference frequency is 200 cycles, which is quite clearly audible and of a frequency so different from the original frequencies that quite a small percentage of distortion becomes audible.

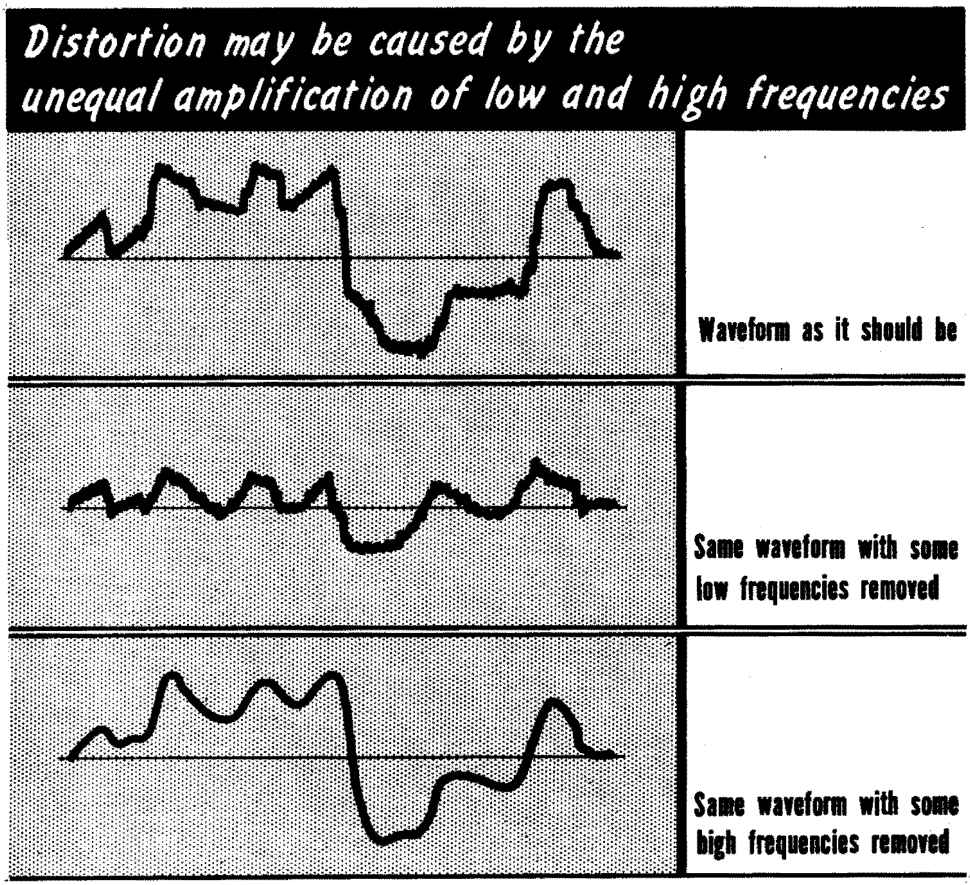


The method of testing for this kind of distortion is to use two oscillators that generate audio frequencies differing by a fixed amount. For example, if we decide to use the 200-cycle difference frequncy, we would arrange that one oscillator give 4000 cycles and the other one 4200 cycles. Or, to test at a higher frequency, when one oscillator gives 8000 cycles the other must give 8200 cycles. The output from the amplifier is passed through a filter that rejects the high frequencies and picks out any component at 200 cycles.

The problem with this method of measurement is that it only discovers whether there is any distortion due to the curvature that would cause *second* harmonic distortion. Other kinds of curvature will also produce distortion, but will not result in a simple 200-cycle difference tone. Rather, they will cause all sorts of other unwanted tones. For this reason and others too complicated to give a complete explanation, the results of the two methods of intermodulation test and harmonic measurement are not consistent, but they depend on the amount of different kinds of curvature in the amplification characteristic of the amplifier.

Frequency Response

Amplifiers produce another kind of distortion because they do not amplify all frequencies uniformly (by the same amount). Low frequencies are reduced in amplitude by the effect of coupling capacitors. Various stray capacitances and the leakage inductance in the output transformer result in loss of high frequencies. Thus no amplifier amplifies all frequencies absolutely uniformly.

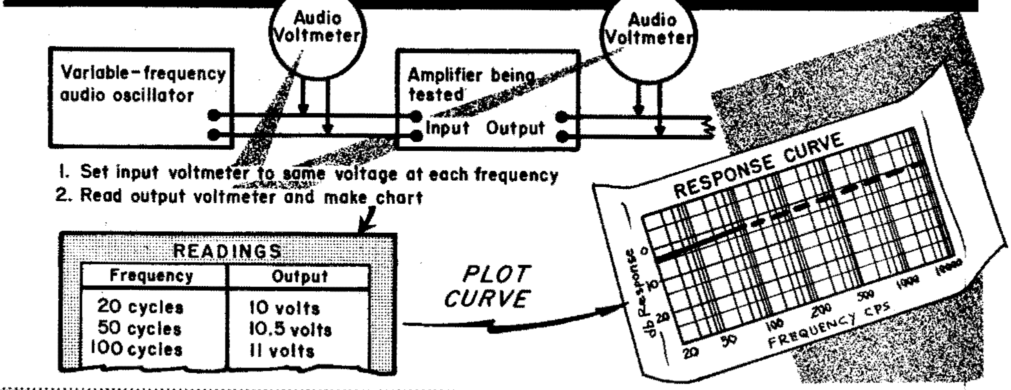


While this non-uniformity does not result in the introduction of any spurious or unwanted frequencies, it does result in the change of the relative magnitude of different frequencies in a composite audio waveform. The frequencies in the middle of the band will usually be amplified more than the extremely low or extremely high frequencies and this will cause a change in the resultant waveform. Fortunately the difference in amplification over the audio range in modern amplifiers is extremely small.

Frequency Response (contd.)

This form of distortion can readily be measured by taking frequency-response measurements of the amplifier. Audio voltages are fed into the amplifier at different frequencies, from the lowest to the highest, and the output voltage is carefully measured to see how closely it corresponds with the input voltage. If an input of 1 millivolt at 1000 cycles produces an output of 10 volts, then 1 millivolt is applied to the amplifier at all frequencies from 20 cycles up to 20,000 cycles and the output voltage is also measured. This will deviate up or down from 10 volts, according to the frequency response of the amplifier. The measurements are usually converted into db according to the ratio of the actual output voltage to the 10-volt output that should be there.

Taking Amplifier Frequency Response



Calculating Response in db

Output at 1000 cycles = 12V.
Output at 100 cycles = 11 V. Ratio = $\frac{12}{11}$ = 1.091

DECIBEL TABLE

Voltage Ratio	Decibels
1.0	0.
1.091	.75
1.1	.828
1.2	1.584
1.3	2.279

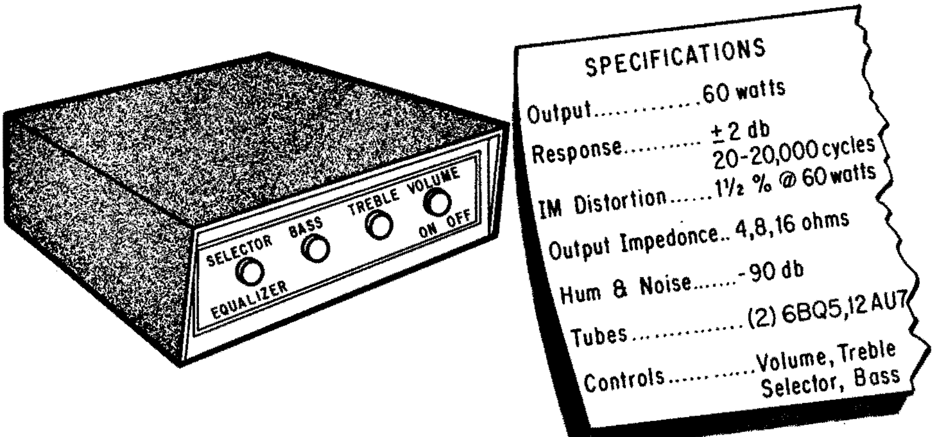
Interpolate from table

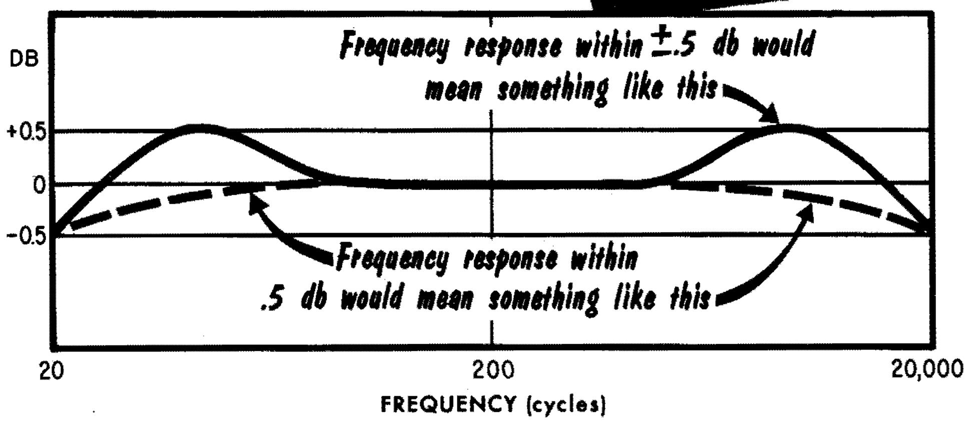
Therefore response is down .75 db at 100 cycles.

Amplifier Specifications

A good amplifier specification will give information about all the types of distortion that we have discussed in order to show how good the amplifier is. The frequency response figure will indicate how closely the amplifier adheres to the same amplification at all frequencies. Sometimes a complete response curve is given and sometimes the specification merely states that the response is within 0.5 db from 20 cycles to 20,000 cycles (or some similar figure). A difference of 0.5 db corresponds with a voltage change of almost 6%, so this means that the amplification will be within 6% of constant through this frequency range.

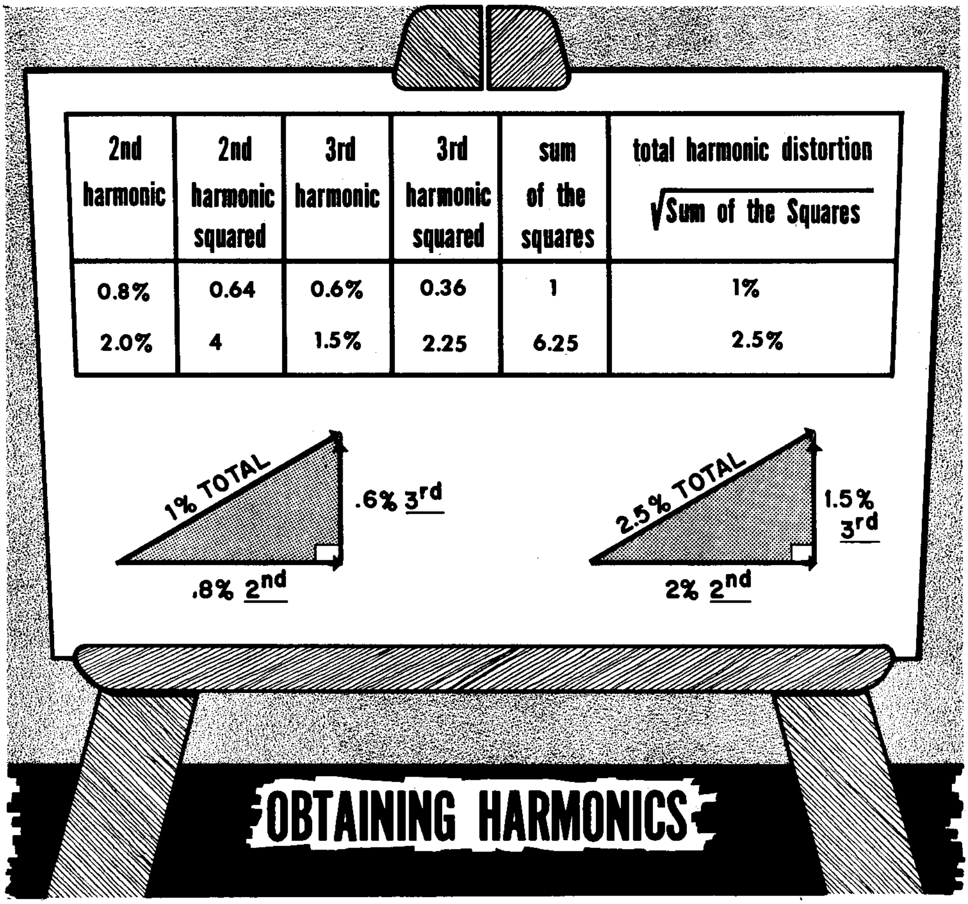
A distortion figure is also given and, unless otherwise specified, this indicates the amount of harmonic distortion. Unfortunately, it is not usual or convenient to specify what kind of harmonic distortion the figure given may be—entirely second harmonic, third harmonic, or it may be a composite of higher harmonics. This is an unfortunate deficiency of this method of specification.





Amplifier Specifications (contd.)

A good modern amplifier might specify a maximum harmonic distortion of 1%. This means that the total of all the harmonic components produced in the amplification of a single sine wave will be less than 1% of the fundamental, and when all of these voltages are squared, added together, and the square root taken, this square root will still not be more than 1% of the fundamental voltage.

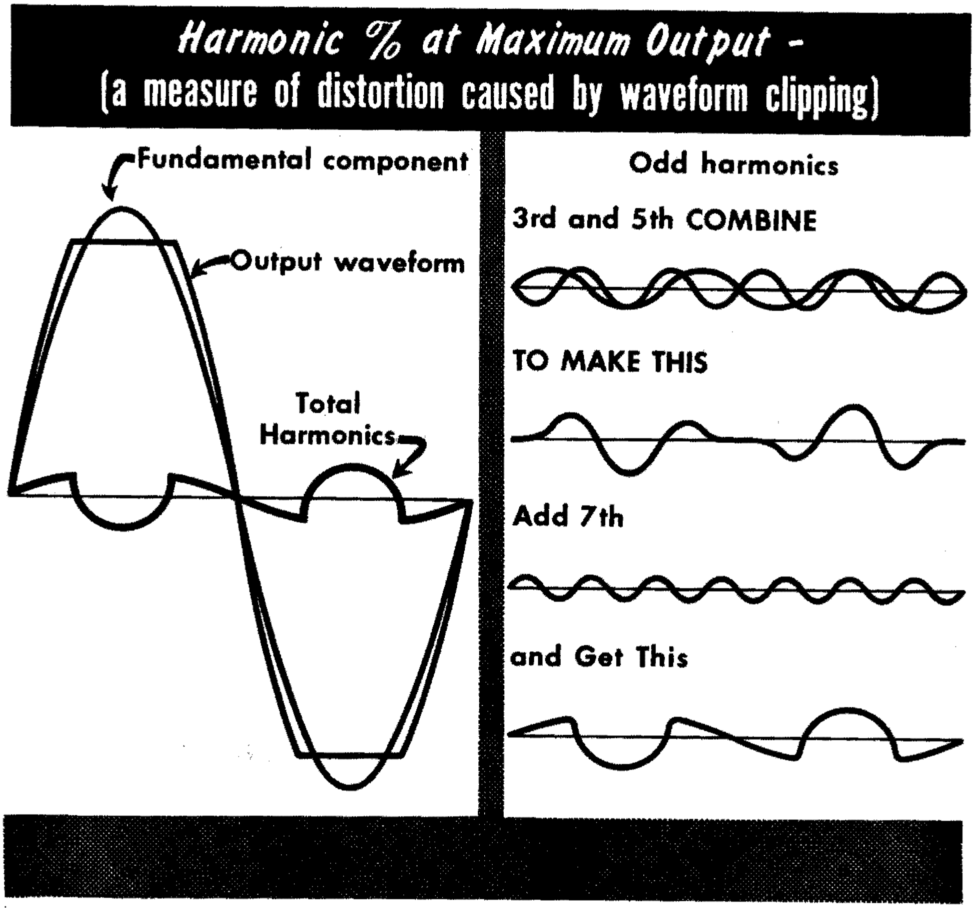


2nd harmonic	2nd harmonic squared	3rd harmonic	3rd harmonic squared	sum of the squares	total harmonic distortion $\sqrt{\text{Sum of the Squares}}$
0.8%	0.64	0.6%	0.36	1	1%
2.0%	4	1.5%	2.25	6.25	2.5%

It is important that, to obtain the total harmonic distortion, one must obtain the square root of the sum of squares of the individual harmonics. The method of combining harmonics is the same regardless of which components are combined. The different components could be second and fifth or any other combination, or a combination of more than just two individual components.

Amplifier Specifications (contd.)

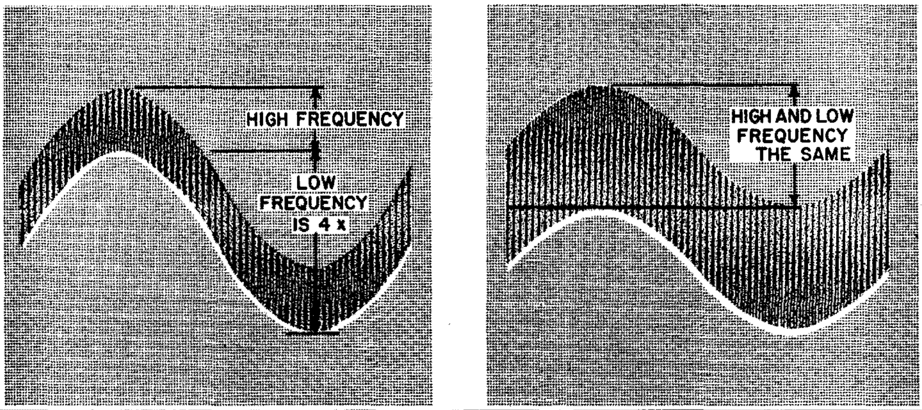
In modern amplifiers in which the harmonic percentage is specified at maximum output, the distortion usually takes the form of clipping on the tops of the waveform, due to the beginning of grid current.



The sharp-peaked distortion component can be analyzed into a series of odd-numbered harmonics, third, fifth, seventh, and so on up the scale. The magnitude of any individual component is quite small compared to the combined peak, as is also the *measured* combined value. The audible effects as well as the visible one seen on an oscilloscope, however, can still be quite noticeable. It has the sound of a knocking at the fundamental frequency, as when the voice coil of the loudspeaker knocks against its end stops. Even a measured 0.5% distortion of this kind is quite readily audible, as well as visible on the waveform displayed by the oscilloscope.

Amplifier Specifications (contd.)

DISTORTION MAY BE SPECIFIED AS INTERMODULATION (IM)

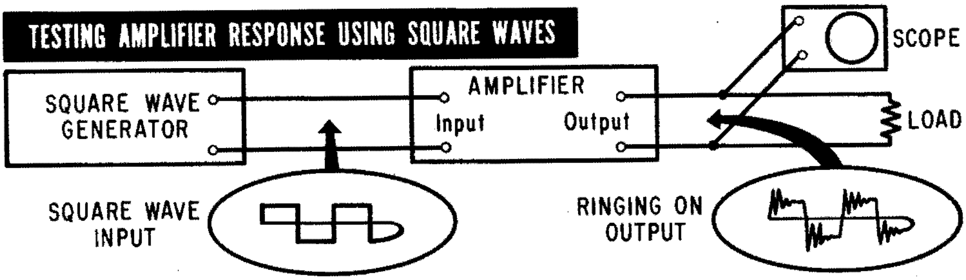


Different possible combination of Amplitude	4:1 or 1:1
and	
Different combinations of Frequency can be used	40 cps -- 3000 cps 100 cps -- 2000 cps 40 cps -- 2000 cps 60 cps -- 7000 cps, etc

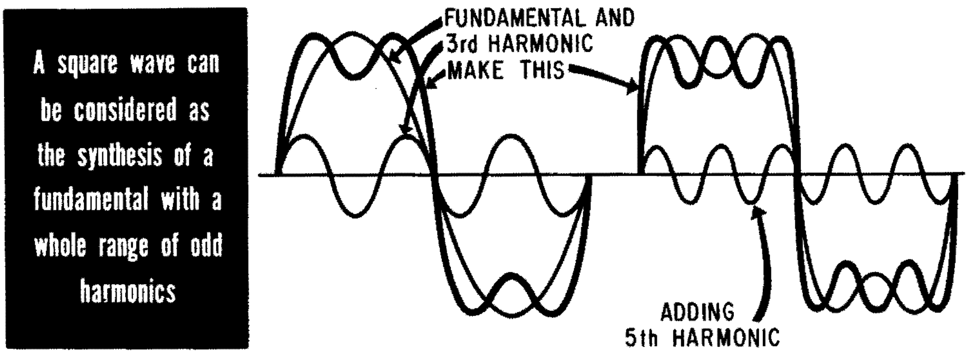
Sometimes the distortion is specified as IM (intermodulation). Unless further information is given about the frequencies used for the test, such a specification is valueless. If the first method of IM test is used, the low frequency (its actual value in cycles per second) is important, as well as the ratio between the *magnitudes* at the two frequencies. The low frequency may be 40 cycles, 60 cycles, or even 100 cycles. Because the handling capacity of an amplifier may be quite different at these three different frequencies, a specification of IM without stating which low frequency was used for the test conveys no real comparison of the performance of different amplifiers.

The peak-to-peak waveform which the amplifier has to handle is the low-frequency voltage *plus* the high-frequency voltage, because one rides atop the other. Consequently the amount of distortion produced will depend upon the ratio between the two voltages, whether this is 4:1, or as is sometimes used, 1:1. Because the results obtained will depend on the precise nature of the curvature or distortion causing them, there is no ready means of converting figures obtained by one test arrangement into figures that would be obtained using the other test arrangement. Consequently the only safe basis for making comparisons between the performance of different amplifiers is to insure that the same combination is used for both tests.

Transient Distortion



This is still another kind of distortion which can be subdivided into further groups. It is not indicated in the majority of amplifier specifications. An amplifier may have a perfectly flat frequency response throughout the audio range, which should indicate that it will give a faithful reproduction of the input waveform at the output. It may show quite low harmonic distortion and yet when a square wave is amplified, the wave may become considerably distorted by a ringing at the corners of the square.



A square wave can be considered as a synthesis or combination of fundamental with a whole range of odd-numbered harmonics. If all these are amplified uniformly, surely the output waveform should still be square? This is true, but it does not take into account possible effects due to time delay in the amplifier, which may not be uniform at all frequencies. Every bit of stray capacitance from a plate or other electrode to ground at different points in the amplifier causes a slight time delay in the amplified audio. This adds up on the way through the amplifier.

If the time delay to all components of the audio waveform is the *same*, the output waveform will still be a square wave, but if it is different for the higher frequency components than it is for the low-frequency fundamental and its lower harmonics, the waveform gets altered. This is one way in which ringing occurs.

Transient Distortion (contd.)

Another kind of transient distortion occurs when the amplitude of the audio suddenly changes. Suppose that a sine wave is amplified, but is stepped up and down at intervals. This will cause the output waveform to step up and down at intervals, which will cause the output voltage to step up and down. If the amplitude of the sine wave is stepped up and down in such a way that the outline (or *envelope*) follows a square waveform, then the output should faithfully reproduce this.

Many amplifiers are not satisfactory in this regard. When a larger audio voltage is being amplified, the output tubes draw more current, which may alter the bias condition. This means that the supply voltages at different points in the circuit will change. The change will take place according to the time constants of the resistances and capacitances in the supply unit, which may not be the same for all the changing voltages. Consequently the gain of the amplifier may go up and down again or down and up again after a sudden change in the amplitude of the audio. This results in an envelope at the output that is different from the envelope at the input. Unfortunately these effects *can* prove quite severe, even with an amplifier whose specification, using the other methods of test, tells of quite good performance—extremely low distortion and very good frequency response.

Thus amplification is far from being the simple matter we started out by supposing. There are many ways in which an amplifier can distort a composite audio program. Whichever method of specifying these is used, it becomes quite an involved matter to give a statement that is completely satisfactory for comparison purposes.