This book is dedicated to
J. N. A. Hawkins, who in
1938 whetted my interest
and stimulated my research
in this interesting subject.

SECTION I

Harmonic and Intermodulation Distortion

It is mandatory that distortion of program material be reduced to a negligible amount in modern wide-range monophonic and stereophonic sound systems, as well as in the transmission of frequency-modulated radio. This is especially true when the frequency range covers the same bandwidth as that of the human ear. Unless the distortion of a wide-range recording and reproducing system is reduced to a negligible value, better listening quality is generally afforded by a system employing a reduced bandwidth.

1.0 GENERAL

Nonlinear distortion in a sound system affects the reproduction by introducing frequency components that were not present in the original program material. These added components are an annoyance to the listener due to the masking and interference effects superimposed on the original components of the signal.

In the most commonly accepted method of measuring harmonic distortion, a single sine-wave frequency is applied to the device under test and the internally generated harmonics are measured at the output. A distortion factor meter or a harmonic wave analyzer may be used, and the harmonic distortion read in percent of the fundamental frequency. Although this single-frequency method of measuring harmonic distortion has been in use for many years, it is an acknowledged fact that such tests are inadequate because they do not present a complete picture of the distortion characteristics of a given device. It is not uncommon for a listener to observe an objectionable quality in the sound reproduction of a transmission system in which a low percentage of harmonic distortion is measured. Amplifiers of the same design, manufacture, and harmonic distortion characteristics often present different qualities in reproduction. This variation in identical amplifiers is usually due to the percent of intermodulation distortion generated within the amplifiers.

Intermodulation distortion may be defined as "the production, in a nonlinear circuit element, of frequencies corresponding to the sum and differences of the fundamentals and harmonics of the two or more frequencies that are transmitted through the element." Thus, when two or more frequencies are simultaneously transmitted through an amplifier, sum and difference frequencies that are not always related to the fundamental frequencies will appear in the output signal as intermodulation distortion. Measuring the distortion characteristics by the intermodulation method more than by any other form of measurement approaches the manner in which the human ear responds. It not only presents a more realistic determination of distortion characteristics, but it is also several times more sensitive than the conventional single-frequency harmonic measurement.

One of the virtues of an intermodulation measurement is that it can be made in the presence of considerable noise and flutter. Noise is excluded in the intermodulation analyzer to a greater degree than in the conventional single-frequency distortion set. Furthermore, distortion measurements by the intermodulation method permit higher frequencies to be measured in a limited-bandwidth system such as photographic film recording systems where the upper frequency is generally limited to 8,000 cps. In such a system, measuring the harmonic distortion at 5,000 cps would be meaningless since the second harmonic occurring at 10,000 cps. would be unmeasurable due to the 8,000-cps cutoff. However, an intermodulation test signal, consisting of 40 and 7,000 cps will often reveal distortion which would seriously affect the reproduction but would not be apparent using other methods of measurement. It is equally important that the lower frequencies of the program material be transmitted through a system without distortion. Both the low and high frequencies are generated simultaneously in an orchestra, and additional low frequencies, such as gun shots, explosions, thunder, etc., may be added.

All of these frequencies, plus the sum and difference frequencies, must be passed by the transmission system without appreciable distortion.

In a complete recording and reproducing system distortion may be caused by amplifiers, filters, equalizers, compressors, cutting heads, light modulators, magnetic recorders or reproducers, monitor speakers, or a combination within the system. Distortion within the amplifiers may be caused by operating tubes on the nonlinear portion of the plate-current characteristic curve, driving the control grid into the region of grid-current flow, improper load termination, poor power-supply regulation, unbalance in the push-pull stages, saturation at the low frequencies of the transformer core materials, and faulty components. Many times after an overall distortion measurement is made, the difficulty can be traced to a *single* component in the system.

Transformers and other components of a device that will pass a 40-cps intermodulated test signal with a low percent of intermodulation distortion will have superior performance over those of low harmonic distortion where a single test frequency is employed. While 40 cps is generally used as the low-frequency component for transmission systems using a bandwidth of 40 to 8,000 cps, an intermodulation test signal of 60 to 100 cps in combination with 2,000 cps can suffice. Distortion of a transmission system can also be observed on a cathode-ray oscilloscope. However, it is rather difficult to determine small percentages of waveform distortion, even with a 5-inch screen. As a rule, 2 or 3% distortion can be observed for the higher order harmonics.

To acquaint the reader with the numerous methods of measuring distortion, several single-frequency as well as intermodulation measurements will be discussed.

1.1 GENERATION OF HARMONIC DISTORTION

Harmonic distortion in an amplifier stage is caused by the nonlinear action of the vacuum tube. If a pure sine wave (a single frequency containing *no* harmonics) is applied to the control grid of a vacuum tube operating with correct voltages and load impedance and the amplitude of the sine wave is such that it does not *overdrive* the control grid, a pure sine wave can be expected at the output of the tube. The only departure from the original waveform of the input will be an increase in its amplitude and a phase reversal of 180° between the input and output circuits;

this is normal for a vacuum tube. To illustrate how nonlinear distortion is generated within a vacuum tube, a plot of the control-grid voltage versus the plate current for a general-purpose triode biased for Class-A operation is shown in Fig. 1-1. The tube selected for this illustration normally requires a negative grid bias of 6 volts. This point is indicated as the normal operating point on the plate-current characteristic curve.

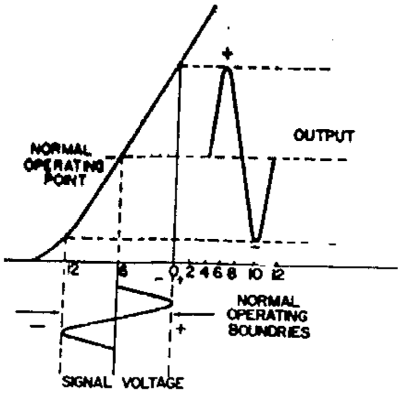

Fig. 1-1. Plate-current characteristic of a medium-mu triode, biased for Class-A operation.

Operating under the aforementioned conditions if a sine wave is applied to the control grid, a sine wave will be reproduced in the plate circuit. However, this will be true only if the signal voltage at the control grid does not exceed the normal operating boundaries. These boundaries are the straight-line portion of the curve, starting at the minus 12-volt bias point and continuing up to the zero-voltage bias point. If the peak signal voltage does not exceed these limits, a sine wave will be developed in the plate circuit, and distortion of the input waveform will not occur. The peak value is determined by multiplying the DC bias voltage by 1.414; hence, in this example, it is 8.5 volts.

Fig. 1-2 shows what happens when the bias is too large. Here, the negative bias has been increased until the operating point (value of plate current) has shifted downward toward the toe or bend in the lower portion of the plate current characteristic at the minus 12-volt point. Under this condition if a sine wave is applied to the control grid, a greater amount of plate current will flow on the positive half of the grid swing than for the negative half and an unsymmetrical waveform will result in the plate circuit. Likewise, if the bias voltage is reduced to a value of minus 2 volts, as shown in Fig. 1-3, the operating point is shifted upward. If the control grid swings positive, the amplitude of the positive half of the plate-current waveform will be less than the negative half, since the tube is being driven into the saturation region of the plate-current characteristic.

When the control grid is driven above the zero point, grid current will start to flow because the control grid is now positive with respect to the cathode. The flow of the grid current will cause a voltage drop across the load in the control-grid circuit which will be in opposition to the signal voltage on the control grid, and the amplitude of the signal-voltage positive-going peaks will be reduced. This region is indicated by the dotted line in the positive half of the signal voltage at the control grid (Fig. 1-3). Under the conditions described, second-order harmonics will be generated in the plate circuit. Thus, it becomes apparent that the bias voltage on the control grid and the amplitude of the applied signal voltage are the controlling factors in the generation of harmonic distortion within a vacuum tube.

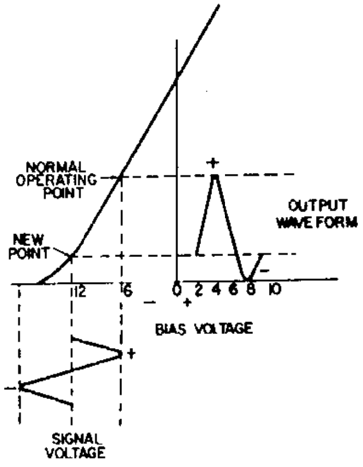

Fig. 1-2. Distortion of plate-circuit waveform caused by overbiasing.

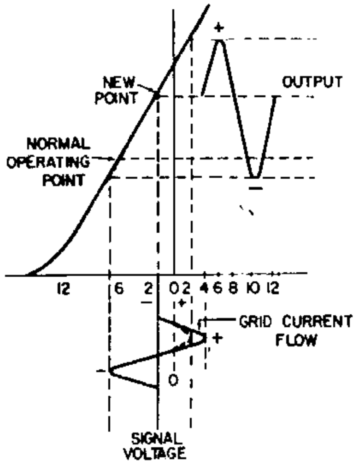

Fig. 1-3. Distortion of plate-current waveform caused by underbiasing.

Harmonic distortion can also be generated within a vacuum tube that is correctly biased (Fig. 1-4). It should be noted that the signal voltage applied to the control grid is large enough to extend into both the toe and knee (saturation) areas of the plate-current characteristic. In the saturation region the peak of the output waveform is flattened since the tube cannot supply further plate current. In the negative region the flow of the plate current is reduced to a near cutoff value. Thus, both the positive and negative peaks of the plate current waveform are distorted, and both odd and even harmonics are generated. The generation of harmonics within a vacuum tube is an inherent characteristic, since the plate resistance is not always uniform and the so-called straight-line portion of the plate current characteristic is not actually straight—it has some curvature.

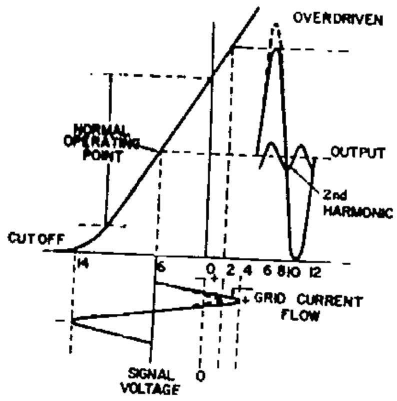

Fig. 1-4. Harmonic distortion generated in a vacuum tube that is correctly biased, but overdriven by too large a signal voltage.

To the human ear the most objectionable harmonics are the second, third, fifth, and seventh; they are usually of higher amplitude and fall within the band of audible frequencies. To further illustrate the generation of harmonics, three superimposed sine waves are shown in Fig. 1-5. F_1 is the fundamental frequency, F_2 is the second harmonic, and F_3 is the third harmonic. Adding the second harmonic to the fundamental frequency algebraically (Fig. 1-6) results in a new frequency, indicated as the resultant. This new frequency is the same as the fundamental frequency, except that the amplitude of the negative peak is increased while the amplitude of the positive peak is reduced. The second harmonic will predominate and is similar to a condition in Fig. 1-3 where the bias voltage was too low. Fig. 1-7 illustrates the fundamental frequency and the second harmonic in a different phase relationship. In this illustration the positive peak of the resultant waveform is of greater amplitude than the negative peak. It is interesting to note that in the condition in Fig. 1-7, the generated second harmonic is similar to the condition in Fig. 1-2, where too great a grid-bias voltage was applied.

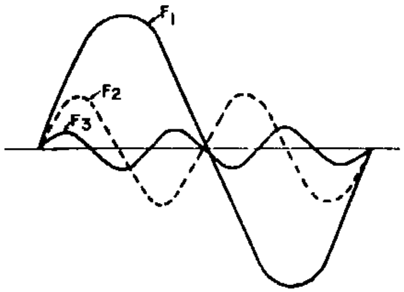

Fig. 1-5. A fundamental frequency (f_1) and its second and third harmonics (f_2 and f_3).

In Figs. 1-6 and 1-7 a horizontal line has been added (in dashed lines) to indicate the shifted operating point. This line represents a DC component, caused by self-rectification within the tube. Self rectification shifts the operating point in the plate-current characteristic. Depending on the circumstances under which the tube is functioning at the moment, it may move either up or down. The operating point in Fig. 1-7 has shifted upward because of the increase of the positive half cycle and the decrease of the negative half cycle, causing the introduction of a second harmonic. When a large value of second harmonic is present, the plate current will increase with an increase of signal voltage. The reverse is true for Fig. 1-6.

When distortion results in the flattening-off of only one peak in the output waveform, the principal distortion is second harmonic. This is accompanied by a change in the average plate current and is referred to as self rectification. Whenever both peaks of the output waveform are flattened, third-harmonic distortion will predominate. If considerable flattening occurs, a fourth-order harmonic will also be generated, particularly if the control grid is driven positive. When the control grid is driven positive, grid current will flow and a voltage drop will be created across the

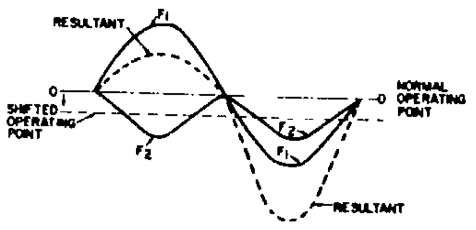

Fig. 1-6. Adding the fundamental frequency (f_1) and its second harmonic to produce a new frequency, called the resultant.

DC resistance of the control-grid load (usually a transformer, choke, or resistor). The voltage drop created by the grid-current flow flattens off the peaks of the input waveform. This flattening and, in turn, the distortion caused by driving the tube into saturation are reflected in the output.

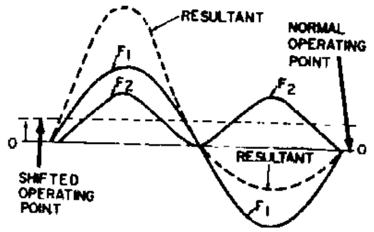

Fig. 1-7. Resultant waveform when operating point is shifted by self rectification.

If the voltage drop caused by grid-current flow in the grid circuit is in the order of 20% of the peak signal voltage, a second-order harmonic of 5% will be generated. As a rule, when a second harmonic is generated, third-harmonic distortion is also generated at approximately the same amplitude.

In the design of a Class-B amplifier (in which grid-current flows for an appreciable part of the input voltage cycle) the DC resistance of the control-grid circuit must be kept quite low to prevent an excessive voltage drop across the load element. If this precaution is not observed, the flow of grid current will increase the control-grid bias voltage, making it more negative than normal. This is why input transformers designed for Class-B operation require a low DC resistance secondary winding.

When a triode tube is overloaded, second-order harmonics are created, while a pentode generates third-order harmonics. Even-order harmonics may be eliminated or reduced to a negligible amplitude by the use of push-pull circuits. Odd-order harmonics generated by a pentode are reduced by the use of push-pull circuitry and the selection of the proper load impedance. Negative feedback will also reduce odd harmonics to a negligible value. However, negative feedback should not be added to an amplifier to reduce its harmonic distortion until the distortion in the amplifier without negative feedback has been reduced to its lowest point. As a rule, pentodes are operated into load impedances of one-fifth to one-sixth their plate resistance. Since the value of load impedance is rather critical, the manufacturer's recommendations should be followed closely for a given set of operating conditions.

The dynamic characteristics of a vacuum tube reveals that at the peak of the applied signal voltage, the instantaneous grid voltage should not drive the grid more negative than a value corresponding to the instantaneous plate current approaching zero. To accomplish this the value of the signal potential will be governed by the value of the grid bias voltage. In a triode it will also be affected by the value of the load resistance. The positive peak of the signal voltage is limited by the flow of the grid current. Often the current is required to be zero at all times. Under these circumstances the instantaneous grid potential must not exceed 1 or 2 volts negative. When a small amount of grid current is permissible, the maximum that the grid may be driven positive is limited by the value of the impedance in the grid circuit, the permissible distortion of the driving voltage and the tube characteristics. An additional limitation appears when using a pentode because when the instantaneous plate voltage is less than a given value, a virtual cathode is formed in the vicinity of the suppressor grid, and the plate voltage is no longer independent of the plate potential. Therefore the operating conditions change, particularly when the load impedance is high.