50

Harmonics

50–1 Musical tones

Pythagoras is said to have discovered the fact that two similar strings under the same tension and differing only in length, when sounded together give an effect that is pleasant to the ear *if* the lengths of the strings are in the ratio of two small integers. If the lengths are as one is to two, they then correspond to the octave in music. If the lengths are as two is to three, they correspond to the interval between C and G, which is called a fifth. These intervals are generally accepted as "pleasant" sounding chords.

Pythagoras was so impressed by this discovery that he made it the basis of a school—Pythagoreans they were called—which held mystic beliefs in the great powers of numbers. It was believed that something similar would be found out about the planets—or "spheres." We sometimes hear the expression: "the music of the spheres." The idea was that there would be some numerical relationships between the orbits of the planets or between other things in nature. People usually think that this is just a kind of superstition held by the Greeks. But is it so different from our own scientific interest in quantitative relationships? Pythagoras' discovery was the first example, outside geometry, of any numerical relationship in nature. It must have been very surprising to suddenly discover that there was a *fact* of nature that involved a simple numerical relationship. Simple measurements of lengths gave a prediction about something which had no apparent connection to geometry—the production of pleasant sounds. This discovery led to the extension that perhaps a good tool for understanding nature would be arithmetic and mathematical analysis. The results of modern science justify that point of view.

Pythagoras could only have made his discovery by making an experimental observation. Yet this important aspect does not seem to have impressed him. If it had, physics might have had a much earlier start. (It is always easy to look back at what someone else has done and to decide what he *should* have done!)

We might remark on a third aspect of this very interesting discovery: that the discovery had to do with two notes that *sound pleasant* to the ear. We may question whether *we* are any better off than Pythagoras in understanding *why* only certain sounds are pleasant to our ear. The general theory of aesthetics is probably no further advanced now than in the time of Pythagoras. In this one discovery of the Greeks, there are the three aspects: experiment, mathematical relationships, and aesthetics. Physics has made great progress on only the first two parts. This chapter will deal with our present-day understanding of the discovery of Pythagoras.

Among the sounds that we hear, there is one kind that we call *noise*. Noise corresponds to a sort of irregular vibration of the eardrum that is produced by the irregular vibration of some object in the neighborhood. If we make a diagram to indicate the pressure of the air on the eardrum (and, therefore, the displacement of the drum) as a function of time, the graph which corresponds to a noise might look like that shown in Fig. 50–1(a). (Such a noise might correspond roughly to the sound of a stamped foot.) The sound of *music* has a different character. Music is characterized by the presence of more-or-less *sustained tones*—or musical "notes." (Musical instruments may make noises as well!) The tone may last for a relatively short time, as when a key is pressed on a piano, or it may be sustained almost indefinitely, as when a flute player holds a long note.

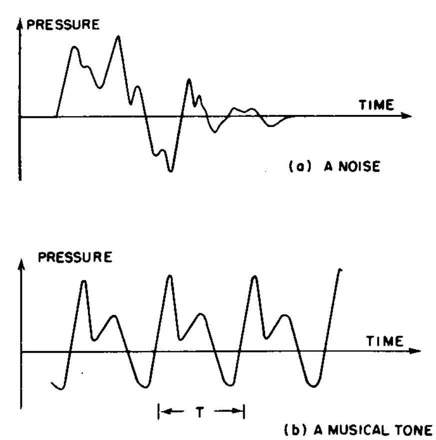

Fig. 50–1. Pressure as a function of time for (a) a noise, and (b) a musical tone.

What is the special character of a musical note from the point of view of the pressure in the air? A musical note differs from a noise in that there is a periodicity in its graph. There is some uneven shape to the variation of the air pressure with time, and the shape repeats itself over and over again. An example of a pressure-time function that would correspond to a musical note is shown in Fig. 50–1(b).

Musicians will usually speak of a musical tone in terms of three characteristics: the loudness, the pitch, and the "quality." The "loudness" is found to correspond to the magnitude of the pressure changes. The "pitch" corresponds to the period of time for one repetition of the basic pressure function. ("Low" notes have longer periods than "high" notes.) The "quality" of a tone has to do with the differences we may still be able to hear between two notes of the same loudness and pitch. An oboe, a violin, or a soprano are still distinguishable even when they sound notes of the same pitch. The quality has to do with the structure of the repeating pattern.

Let us consider, for a moment, the sound produced by a vibrating string. If we pluck the string, by pulling it to one side and releasing it, the subsequent motion will be determined by the motions of the waves we have produced. We know that these waves will travel in both directions, and will be reflected at the ends. They will slosh back and forth for a long time. No matter how complicated the wave is, however, it will repeat itself. The period of repetition is just the time T required for the wave to travel two full lengths of the string. For that is just the time required for any wave, once started, to reflect off each end and return to its starting position, and be proceeding in the original direction. The time is the same for waves which start out in either direction. Each point on the string will, then, return to its starting position after one period, and again one period later, etc. The sound wave produced must also have the same repetition. We see why a plucked string produces a musical tone.

50–2 The Fourier series

We have discussed in the preceding chapter another way of looking at the motion of a vibrating system. We have seen that a string has various natural modes of oscillation, and that any particular kind of vibration that may be set up by the starting conditions can be thought of as a combination—in suitable proportions—of several of the natural modes, oscillating together. For a string we found that the normal modes of oscillation had the frequencies ω_0, $2\omega_0$, $3\omega_0, \ldots$ The most general motion of a plucked string, therefore, is composed of the sum of a sinusoidal oscillation at the fundamental frequency ω_0, another at the second harmonic frequency $2\omega_0$, another at the third harmonic $3\omega_0$, etc. Now the fundamental mode repeats itself every period $T_1 = 2\pi/\omega_0$. The second harmonic mode repeats itself every $T_2 = 2\pi/2\omega_0$. It *also* repeats itself every $T_1 = 2T_2$, after *two* of its periods. Similarly, the third harmonic mode repeats itself after a time T_1 which is 3 of its periods. We see again why a plucked string repeats its whole pattern with a periodicity of T_1. It produces a musical tone.

We have been talking about the motion of the string. But the *sound*, which is the motion of the air, is produced by the motion of the string, so its vibrations too must be composed of the same harmonics—though we are no longer thinking about the normal modes of the air. Also, the relative strength of the harmonics may be different in the air than in the string, particularly if the string is "coupled" to the air via a sounding board. The efficiency of the coupling to the air is different for different harmonics.

If we let $f(t)$ represent the air pressure as a function of time for a musical tone [such as that in Fig. 50–1(b)], then we expect that $f(t)$ can be written as the sum of a number of simple harmonic functions of time—like $\cos \omega t$—for each of the various harmonic frequencies. If the period of the vibration is T, the fundamental angular frequency will be $\omega = 2\pi/T$, and the harmonics will be 2ω, 3ω, etc.

There is one slight complication. For each frequency we may expect that the starting phases will not necessarily be the same for all frequencies. We should, therefore, use functions like $\cos(\omega t + \phi)$. It is, however, simpler to use instead both the sine and cosine functions for *each* frequency. We recall that

$$\cos(\omega t + \phi) = (\cos\phi \cos\omega t - \sin\phi \sin\omega t) \tag{50.1}$$

and since ϕ is a constant, *any* sinusoidal oscillation at the frequency ω can be written as the sum of a term with $\cos \omega t$ and another term with $\sin \omega t$.

We conclude, then, that *any* function $f(t)$ that is periodic with the period T can be written mathematically as

$$\begin{aligned} f(t) = a_0 & \\ &+ a_1 \cos \omega t + b_1 \sin \omega t \\ &+ a_2 \cos 2\omega t + b_2 \sin 2\omega t \\ &+ a_3 \cos 3\omega t + b_3 \sin 3\omega t \\ &+ \dots \qquad + \dots \end{aligned} \tag{50.2}$$

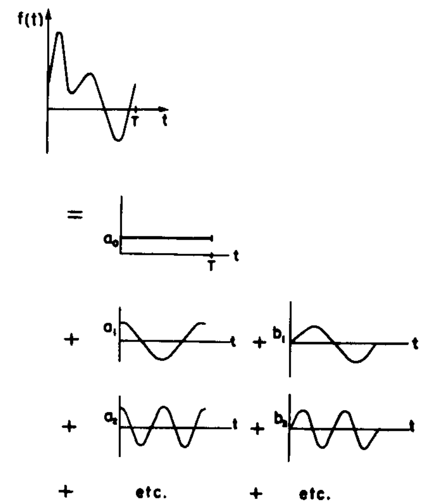

Fig. 50–2. Any periodic function $f(t)$ is equal to a sum of simple harmonic functions.

where $\omega = 2\pi/T$ and the a's and b's are numerical constants which tell us how much of each component oscillation is present in the oscillation $f(t)$. We have added the "zero-frequency" term a_0 so that our formula will be completely general, although it is usually zero for a musical tone. It represents a shift of the average value (that is, the "zero" level) of the sound pressure. With it our formula can take care of any case. The equality of Eq. (50.2) is represented schematically in Fig. 50–2. (The amplitudes, a_n and b_n, of the harmonic functions must be suitably chosen. They are shown schematically and without any particular scale in the figure.) The series (50.2) is called the *Fourier series* for $f(t)$.

We have said that *any* periodic function can be made up in this way. We should correct that and say that any sound wave, or any function we ordinarily encounter in physics, can be made up of such a sum. The mathematicians can invent functions which cannot be made up of simple harmonic functions—for instance, a function that has a "reverse twist" so that it has two values for some values of t! We need not worry about such functions here.

50–3 Quality and consonance

Now we are able to describe what it is that determines the "quality" of a musical tone. It is the relative amounts of the various harmonics—the values of the a's and b's. A tone with only the first harmonic is a "pure" tone. A tone with many strong harmonics is a "rich" tone. A violin produces a different proportion of harmonics than does an oboe.

We can "manufacture" various musical tones if we connect several "oscillators" to a loudspeaker. (An oscillator usually produces a nearly pure simple harmonic function.) We should choose the frequencies of the oscillators to be ω, 2ω, 3ω, etc. Then by adjusting the volume control on each oscillator, we can add in any amount we wish of each harmonic—thereby producing tones of different quality. An electric organ works in much this way. The "keys" select the frequency of the fundamental oscillator and the "stops" are switches that control the relative proportions of the harmonics. By throwing these switches, the organ can be made to sound like a flute, or an oboe, or a violin.

It is interesting that to produce such "artificial" tones we need only one oscillator for each frequency—we do not need separate oscillators for the sine and cosine components. The ear is not very sensitive to the relative phases of the harmonics. It pays attention mainly to the *total* of the sine and cosine parts of each frequency. Our analysis is more accurate than is necessary to explain the *subjective* aspect of music. The response of a microphone or other physical instrument does depend on the phases, however, and our complete analysis may be needed to treat such cases.

The "quality" of a spoken sound also determines the vowel sounds that we recognize in speech. The shape of the mouth determines the frequencies of the natural modes of vibration of the air in the mouth. Some of these modes are set into vibration by the sound waves from the vocal chords. In this way, the amplitudes of some of the harmonics of the sound are increased with respect to others. When we change the shape of our mouth, harmonics of different frequencies are given preference. These effects account for the difference between an "e–e–e" sound and an "a–a–a" sound.

We all know that a particular vowel sound—say "e-e-e"—still "sounds like" the same vowel whether we say (or sing) it at a high or a low pitch. From the mechanism we describe, we would expect that *particular* frequencies are emphasized when we shape our mouth for an "e-e-e," and that they *do not* change as we change the pitch of our voice. So the relation of the important harmonics to the fundamental—that is, the "quality"—changes as we change pitch. Apparently the mechanism by which we recognize speech is not based on specific harmonic relationships.

What should we say now about Pythagoras' discovery? We understand that two similar strings with lengths in the ratio of 2 to 3 will have fundamental frequencies in the ratio 3 to 2. But why should they "sound pleasant" together? Perhaps we should take our clue from the frequencies of the harmonics. The second harmonic of the lower shorter string will have the *same* frequency as the third harmonic of the longer string. (It is easy to show—or to believe—that a plucked string produces strongly the several lowest harmonics.)

Perhaps we should make the following rules. Notes sound consonant when they have harmonics with the same frequency. Notes sound dissonant if their upper harmonics have frequencies near to each other but far enough apart that there are rapid beats between the two. Why beats do not sound pleasant, and why unison of the upper harmonics does sound pleasant, is something that we do not know how to define or describe. We cannot say from this knowledge of what *sounds* good, what ought, for example, to *smell* good. In other words, our understanding of it is not anything more general than the statement that when they are in unison they sound good. It does not permit us to deduce anything more than the properties of concordance in music.

It is easy to check on the harmonic relationships we have described by some simple experiments with a piano. Let us label the 3 successive C's near the middle of the keyboard by C, C′, and C″, and the G's just above by G, G′, and G″. Then the fundamentals will have relative frequencies as follows:

C –2	G – 3
C′ –4	G′ – 6
C″–8	G″–12

These harmonic relationships can be demonstrated in the following way: Suppose we press C′ *slowly*—so that it does not sound but we cause the damper to be lifted. If we then sound C, it will produce its own fundamental *and* some second harmonic. The second harmonic will set the strings of C′ into vibration. If we now release C (keeping C′ pressed) the damper will stop the vibration of the C strings, and we can hear (softly) the note C′ as it dies away. In a similar way, the third harmonic of C can cause a vibration of G′. Or the sixth of C (now getting much weaker) can set up a vibration in the fundamental of G″.

A somewhat different result is obtained if we press G quietly and then sound C′. The third harmonic of C′ will correspond to the fourth harmonic of G, so *only* the fourth harmonic of G will be excited. We can hear (if we listen closely) the sound of G″, which is two octaves above the G we have pressed! It is easy to think up many more combinations for this game.

We may remark in passing that the major scale can be defined just by the condition that the three major chords (F–A–C); (C–E–G); and (G–B–D) *each* represent tone sequences with the frequency ratio (4: 5: 6). These ratios—plus the fact that an octave (C–C′, B–B′, etc.) has the ratio 1: 2—determine the whole scale for the "ideal" case, or for what is called "just intonation." Keyboard instruments like the piano are *not* usually tuned in this manner, but a little "fudging" is done so that the frequencies are *approximately* correct for all possible starting tones. For this tuning, which is called "tempered," the octave (still 1: 2) is divided into 12 equal intervals for which the frequency ratio is $(2)^{1/12}$. A fifth no longer has the frequency ratio 3/2, but $2^{7/12} = 1.499$, which is apparently close enough for most ears.

We have stated a rule for consonance in terms of the coincidence of harmonics. Is this coincidence perhaps the *reason* that two notes are consonant? One worker has claimed that two *pure* tones—tones carefully manufactured to be free of harmonics—do not give the *sensations* of consonance or dissonance as the relative frequencies are placed at or near the expected ratios. (Such experiments are difficult because it is difficult to manufacture pure tones, for reasons that we shall see later.) We cannot still be certain whether the ear is matching harmonics or doing arithmetic when we decide that we like a sound.

50–4 The Fourier coefficients

Let us return now to the idea that any note—that is, a *periodic* sound—can be represented by a suitable combination of harmonics. We would like to show how we can find out what amount of each harmonic is required. It is, of course, easy to compute $f(t)$, using Eq. (50.2), if we are *given* all the coefficients a and b. The question now is, if we are given $f(t)$ how can we know what the coefficients of the various harmonic terms should be? (It is easy to make a cake from a recipe; but can we write down the recipe if we are given a cake?)

Fourier discovered that it was not really very difficult. The term a_0 is certainly easy. We have already said that it is just the average value of $f(t)$ over one period (from $t = 0$ to $t = T$). We can easily see that this is indeed so. The average value of a sine or cosine function over one period is zero. Over two, or three, or any whole number of periods, it is also zero. So the average value of all of the terms on the right-hand side of Eq. (50.2) is zero, except for a_0. (Recall that we must choose $\omega = 2\pi/T$.)

Now the average of a sum is the sum of the averages. So the average of $f(t)$ is just the average of a_0. But a_0 is a *constant*, so its average is just the same as its value. Recalling the definition of an average, we have

$$a_0 = \frac{1}{T}\int_0^T f(t)\,dt. \tag{50.3}$$

The other coefficients are only a little more difficult. To find them we can use a trick discovered by Fourier. Suppose we multiply both sides of Eq. (50.2) by some harmonic function—say by $\cos 7\omega t$. We have then

$$\begin{aligned} f(t)\cdot\cos 7\omega t = {} & a_0\cdot\cos 7\omega t \\ & + a_1\cos\omega t\cdot\cos 7\omega t + b_1\sin\omega t\cdot\cos 7\omega t \\ & + a_2\cos 2\omega t\cdot\cos 7\omega t + b_2\sin 2\omega t\cdot\cos 7\omega t \\ & + \cdots \qquad\qquad\qquad\quad + \cdots \\ & + a_7\cos 7\omega t\cdot\cos 7\omega t + b_7\sin 7\omega t\cdot\cos 7\omega t \\ & + \cdots \qquad\qquad\qquad\quad + \cdots \end{aligned} \tag{50.4}$$

Now let us average both sides. The average of $a_0 \cos 7\omega t$ over the time T is proportional to the average of a cosine over 7 whole periods. But that is just zero. The average of *almost all* of the rest of the terms is *also* zero. Let us look at the a_1 term. We know, in general, that

$$\cos A\cos B = \tfrac{1}{2}\cos(A + B) + \tfrac{1}{2}\cos(A - B). \tag{50.5}$$

The a_1 term becomes

$$\tfrac{1}{2}a_1(\cos 8\omega t + \cos 6\omega t). \tag{50.6}$$

We thus have two cosine terms, one with 8 full periods in T and the other with 6. *They both average to zero.* The average of the a_1 term is therefore zero.

For the a_2 term, we would find $a_2 \cos 9\omega t$ and $a_2 \cos 5\omega t$, each of which also averages to zero. For the a_9 term, we would find $\cos 16\omega t$ and $\cos(-2\omega t)$. But $\cos(-2\omega t)$ is the same as $\cos 2\omega t$, so both of these have zero averages. It is clear that *all* of the a terms will have a zero average *except* one. And that one is the a_7 term. For this one we have

$$\tfrac{1}{2}a_7(\cos 14\omega t + \cos 0). \tag{50.7}$$

The cosine of zero is one, and its average, of course, is one. So we have the result that the average of all of the a terms of Eq. (50.4) equals $\frac{1}{2}a_7$.

The b terms are even easier. When we multiply by any cosine term like $\cos n\omega t$, we can show by the same method that *all* of the b terms have the average value zero.

We see that Fourier's "trick" has acted like a sieve. When we multiply by $\cos 7\omega t$ and average, all terms drop out except a_7, and we find that

$$\text{Average}\,[f(t)\cdot\cos 7\omega t] = a_7/2, \tag{50.8}$$

or

$$a_7 = \frac{2}{T}\int_0^T f(t)\cdot\cos 7\omega t\,dt. \tag{50.9}$$

We shall leave it for the reader to show that the coefficient b_7 can be obtained by multiplying Eq. (50.2) by $\sin 7\omega t$ and averaging both sides. The result is

$$b_7 = \frac{2}{T}\int_0^T f(t)\cdot\sin 7\omega t\,dt. \tag{50.10}$$

Now what is true for 7 we expect is true for any integer. So we can summarize our proof and result in the following more elegant mathematical form. If m and n are integers other than zero, and if $\omega = 2\pi/T$, then

$$\text{I.}\quad \int_0^T \sin n\omega t\cos m\omega t\,dt = 0. \tag{50.11}$$

$$\left.\begin{aligned}\text{II.}\quad &\int_0^T \cos n\omega t\cos m\omega t\,dt = \\ \text{III.}\quad &\int_0^T \sin n\omega t\sin m\omega t\,dt = \end{aligned}\right\}\begin{cases}0 & \text{if } n \neq m.\\ T/2 & \text{if } n = m.\end{cases} \tag{50.12}$$

$$\text{IV.}\quad f(t) = a_0 + \sum_{n=1}^{\infty} a_n\cos n\omega t + \sum_{n=1}^{\infty} b_n\sin n\omega t. \tag{50.13}$$

$$\text{V.}\quad a_0 = \frac{1}{T}\int_0^T f(t)\cdot dt. \tag{50.14}$$

$$a_n = \frac{2}{T}\int_0^T f(t)\cdot\cos n\omega t\,dt. \tag{50.15}$$

$$b_n = \frac{2}{T}\int_0^T f(t)\cdot\sin n\omega t\,dt. \tag{50.16}$$

In earlier chapters it was convenient to use the exponential notation for representing simple harmonic motion. Instead of $\cos\omega t$ we used $\text{Re}\, e^{i\omega t}$, the real part of the exponential function. We have used cosine and sine functions in this chapter because it made the derivations perhaps a little clearer. Our final result of Eq. (50.13) can, however, be written in the compact form

$$f(t) = \text{Re}\sum_{n=0}^{\infty}\hat{a}_n e^{in\omega t}, \tag{50.17}$$

where $\hat{a}_n$ is the complex number $a_n - ib_n$ (with $b_0 = 0$). If we wish to use the same notation throughout, we can write also

$$\hat{a}_n = \frac{2}{T}\int_0^T f(t)e^{-in\omega t}\,dt \qquad (n \geq 1). \tag{50.18}$$

We now know how to "analyze" a periodic wave into its harmonic components. The procedure is called *Fourier analysis*, and the separate terms are called Fourier components. We have *not* shown, however, that once we find all of the Fourier components and add them together, we do indeed get back our $f(t)$. The mathematicians have shown, for a wide class of functions, in fact for all that are of interest to physicists, that if we can do the integrals we will get back $f(t)$. There is one minor exception. If the function $f(t)$ is discontinuous, i.e., if it jumps suddenly from one value to another, the Fourier sum will give a value at the breakpoint halfway between the upper and lower values at the discontinuity. So if we have the strange function $f(t) = 0, 0 \leq t < t_0$, and $f(t) = 1$ for $t_0 \leq t \leq T$, the Fourier sum will give the right value everywhere *except at* t_0, where it will have the value $\frac{1}{2}$ instead of 1. It is rather unphysical anyway to insist that a function should be zero *up to* t_0, but 1 *right at* t_0. So perhaps we should make the "rule" for physicists that any discontinuous function (which can only be a simplification of a *real* physical function) should be defined with halfway values at the discontinuities. Then any such function—with any finite number of such jumps—as well as all other physically interesting functions, are given correctly by the Fourier sum.

As an exercise, we suggest that the reader determine the Fourier series for the function shown in Fig. 50–3. Since the function cannot be written in an explicit algebraic form, you will not be able to do the integrals from zero to T in the usual way. The integrals are easy, however, if we separate them into two parts: the integral from zero to $T/2$ (over which $f(t) = 1$) and the integral from $T/2$ to T (over which $f(t) = -1$). The result should be

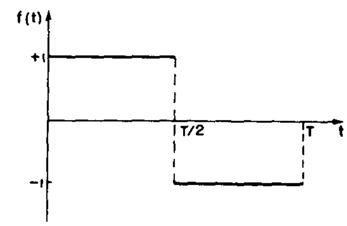

Fig. 50–3. Square-wave function. $f(t) = +1$ for $0 < t < t/2$, $f(t) = -1$ for $T/2 < t < T$.

$$f(t) = \frac{4}{\pi}\left(\sin \omega t + \tfrac{1}{3} \sin 3\omega t + \tfrac{1}{5} \sin 5\omega t + \cdots\right), \tag{50.19}$$

where $\omega = 2\pi/T$. We thus find that our square wave (with the particular phase chosen) has only odd harmonics, and their amplitudes are in inverse proportion to their frequencies.

Let us check that Eq. (50.19) does indeed give us back $f(t)$ for some value of t. Let us choose $t = T/4$, or $\omega t = \pi/2$. We have

$$f(t) = \frac{4}{\pi}\left(\sin \frac{\pi}{2} + \frac{1}{3} \sin \frac{3\pi}{2} \frac{1}{5} \sin \frac{5\pi}{2} + \cdots\right) \tag{50.20}$$

$$= \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right) \tag{50.21}$$

The series* has the value $\pi/4$, and we find that $f(t) = 1$.

50–5 The energy theorem

The energy in a wave is proportional to the square of its amplitude. For a wave of complex shape, the energy in one period will be proportional to $\int_0^T f^2(t)\,dt$. We can also relate this energy to the Fourier coefficients. We write

$$\int_0^T f^2(t)\,dt = \int_0^T \left[a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega t + \sum_{n=1}^{\infty} b_n \sin n\omega t\right]^2 dt. \tag{50.22}$$

When we expand the square of the bracketed term we will get all possible cross terms, such as $a_5 \cos 5\omega t \cdot b_7 \cos 7\omega t$. We have shown above, however, [Eqs. (50.11) and (50.12)] that the integrals of all such terms over one period is zero.

---

* The series can be evaluated in the following way. First we remark that $\int_0^x [dx/(1 + x^2)] = \tan^{-1} x$. Second, we expand the integrand in a series $1/(1 + x^2) = 1 - x^2 + x^4 - x^6 + \ldots$ We integrate the series term by term (from zero to x) to obtain $\tan^{-1} x = 1 - x^3/3 + x^5/5 - x^7/7 + \ldots$ Setting $x = 1$, we have the stated result, since $\tan^{-1} 1 = \pi/4$.

We have left only the square terms like $a_5^2 \cos^2 5\omega t$. The integral of any cosine squared or sine squared over one period is equal to $T/2$, so we get

$$\int_0^T f^2(t)\,dt = Ta_0^2 + \frac{T}{2}(a_1^2 + a_2^2 + \cdots + b_1^2 + b_2^2 + \cdots)$$
$$= Ta_0^2 + \frac{T}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2). \tag{50.23}$$

This equation is called the "energy theorem," and says that the total energy in a wave is just the sum of the energies in all of the Fourier components. For example, applying this theorem to the series (50.19), since $[f(t)]^2 = 1$ we get

$$T = \frac{T}{2}\cdot\left(\frac{4}{\pi}\right)^2\left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2}\cdots\right),$$

so we learn that the sum of the squares of the reciprocals of the odd integers is $\pi^2/8$. In a similar way, by first obtaining the Fourier series for the function and using the energy theorem, we can prove that $1 + 1/2^4 + 1/3^4 + \cdots$ is $\pi^4/90$, a result we needed in Chapter 45.

50–6 Nonlinear responses

Finally, in the theory of harmonics there is an important phenomenon which should be remarked upon because of its practical importance—that of nonlinear effects. In all the systems that we have been considering so far, we have supposed that everything was linear, that the responses to forces, say the displacements or the accelerations, were always proportional to the forces. Or that the currents in the circuits were proportional to the voltages, and so on. We now wish to consider cases where there is not a strict proportionality. We think, at the moment, of some device in which the response, which we will call x_{out} at the time t, is determined by the input x_{in} at the time t. For example, x_{in} might be the force and x_{out} might be the displacement. Or x_{in} might be the current and x_{out} the voltage. If the device is linear, we would have

$$x_{\text{out}}(t) = Kx_{\text{in}}(t), \tag{50.24}$$

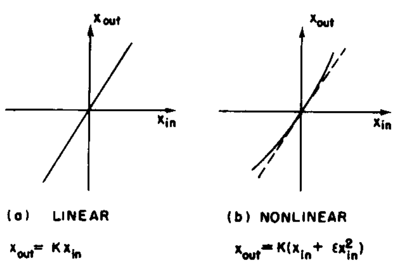

Fig. 50–4. Linear and nonlinear responses.

where K is a constant independent of t and of x_{in}. Suppose, however, that the device is nearly, but not exactly, linear, so that we can write

$$x_{\text{out}}(t) = K[x_{\text{in}}(t) + \epsilon x_{\text{in}}^2(t)], \tag{50.25}$$

where ϵ is small in comparison with unity. Such linear and nonlinear responses are shown in the graphs of Fig. 50–4.

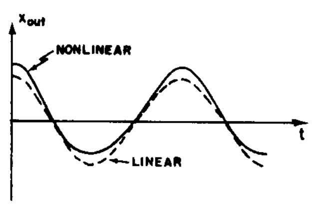

Fig. 50–5. The response of a nonlinear device to the input $\cos \omega t$. A linear response is shown for comparison.

Nonlinear responses have several important practical consequences. We shall discuss some of them now. First we consider what happens if we apply a pure tone at the input. We let $x_{\text{in}} = \cos \omega t$. If we plot x_{out} as a function of time we get the solid curve shown in Fig. 50–5. The dashed curve gives, for comparison, the response of a linear system. We see that the output is no longer a cosine function. It is more peaked at the top and flatter at the bottom. We say that the output is *distorted*. We know, however, that such a wave is no longer a pure tone, that it will have harmonics. We can find what the harmonics are. Using $x_{\text{in}} = \cos \omega t$ with Eq. (50.25), we have

$$x_{\text{out}} = K(\cos \omega t + \epsilon \cos^2 \omega t). \tag{50.26}$$

From the equality $\cos^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$, we have

$$x_{\text{out}} = K\left(\cos \omega t + \frac{\epsilon}{2} - \frac{\epsilon}{2}\cos 2\omega t\right)\cdot \tag{50.27}$$

The output has not only a component at the fundamental frequency, that was present at the input, but also has some of its second harmonic. There has also appeared at the output a constant term $K(\epsilon/2)$, which corresponds to the shift of the average value, shown in Fig. 50–5. The process of producing a shift of the average value is called *rectification.*

A nonlinear response will rectify and will produce harmonics of the frequencies at its input. Although the nonlinearity we assumed produced only second harmonics, nonlinearities of higher order—those which have terms like x_{in}^3 and x_{in}^4, for example—will produce harmonics higher than the second.

Another effect which results from a nonlinear response is *modulation.* If our input function contains two (or more) pure tones, the output will have not only their harmonics, but still other frequency components. Let $x_{\text{in}} = A\cos\omega_1 t + B\cos\omega_2 t$, where now ω_1 and ω_2 are *not* intended to be in a harmonic relation. In addition to the linear term (which is K times the input) we shall have a component in the output given by

$$x_{\text{out}} = K\epsilon(A\cos\omega_1 t + B\cos\omega_2 t)^2 \tag{50.28}$$

$$= K\epsilon(A^2\cos^2\omega_1 t + B^2\cos^2\omega_2 t + 2AB\cos\omega_1 t\cos\omega_2 t). \tag{50.29}$$

The first two terms in the parentheses of Eq. (50.29) are just those which gave the constant terms and second harmonic terms we found above. The last term is new.

We can look at this new "cross term" $AB\cos\omega_1 t\cos\omega_2 t$ in two ways. First, if the two frequencies are widely different (for example, if ω_1 is much greater than ω_2) we can consider that the cross term represents a cosine oscillation of varying amplitude. That is, we can think of the factors in this way:

$$AB\cos\omega_1 t\cos\omega_2 t = C(t)\cos\omega_1 t, \tag{50.30}$$

with

$$C(t) = AB\cos\omega_2 t. \tag{50.31}$$

We say that the amplitude of $\cos\omega_1$ is *modulated* with the frequency ω_2.

Alternatively, we can write the cross term in another way:

$$AB\cos\omega_1 t\cos\omega_2 t = \frac{AB}{2}[\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t]. \tag{50.32}$$

We would now say that two *new* components have been produced, one at the *sum* frequency $(\omega_1 + \omega_2)$, another at the *difference* frequency $(\omega_1 - \omega_2)$.

We have two different, but equivalent, ways of looking at the same result. In the special case that $\omega_1 \gg \omega_2$, we can relate these two different views by remarking that since $(\omega_1 + \omega_2)$ and $(\omega_1 - \omega_2)$ are near to each other we would expect to observe beats between them. But these beats have just the effect of *modulating* the amplitude of the *average* frequency ω_1 by one-half the difference frequency $2\omega_2$. We see, then, why the two descriptions are equivalent.

In summary, we have found that a nonlinear response produces several effects: rectification, generation of harmonics, and modulation, or the generation of components with sum and difference frequencies.

We should notice that all these effects (Eq. 50.29) are proportional not only to the nonlinearity coefficient ϵ, but also to the product of two amplitudes—either A^2, B^2, or AB. We expect these effects to be much more important for *strong* signals than for weak ones.

The effects we have been describing have many practical applications. First, with regard to sound, it is believed that the ear is nonlinear. This is believed to account for the fact that with loud sounds we have the sensation that we *hear* harmonics and also sum and difference frequencies even if the sound waves contain only pure tones.

The components which are used in sound-reproducing equipment—amplifiers, loudspeakers, etc.—always have some nonlinearity. They produce distortions in the sound—they generate harmonics, etc.—which were not present in the original sound. These new components are heard by the ear and are apparently objectionable. It is for this reason that "Hi-Fi" equipment is designed to be as linear as possible. (Why the nonlinearities of the *ear* are *not* "objectionable" in the same way, or how we even know that the nonlinearity is in the *loudspeaker* rather than in the *ear* is not clear!)

Nonlinearities are quite *necessary*, and are, in fact, intentionally made large in certain parts of radio transmitting and receiving equipment. In an AM transmitter the "voice" signal (with frequencies of some kilocycles per second) is combined with the "carrier" signal (with a frequency of some megacycles per second) in a nonlinear circuit called a *modulator*, to produce the modulated oscillation that is transmitted. In the receiver, the components of the received signal are fed to a nonlinear circuit which combines the sum and difference frequencies of the modulated carrier to generate again the voice signal.

When we discussed the transmission of light, we assumed that the induced oscillations of charges were proportional to the electric field of the light—that the response was linear. That is indeed a very good approximation. It is only within the last few years that light sources have been devised (lasers) which produce an intensity of light strong enough so that nonlinear effects can be observed. It is now possible to generate harmonics of light frequencies. When a strong red light passes through a piece of glass, a little bit of blue light—second harmonic—comes out!