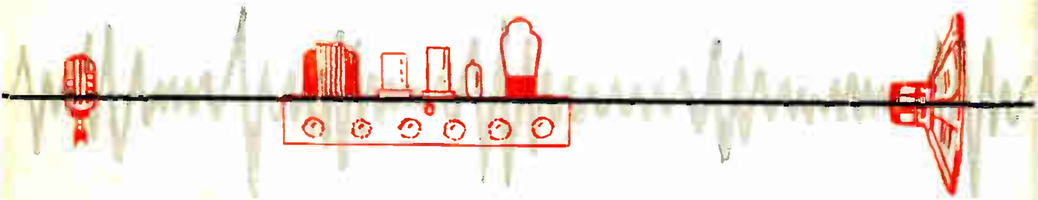
basic audio course

by Donald Carl Hoefler



GERNSBACK LIBRARY NO. 66

$2.75

Phillips H. Grant

basic audio course

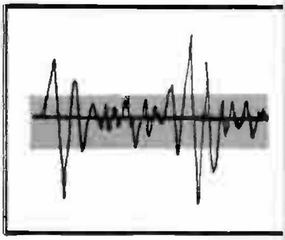
by Donald Carl Hoefler

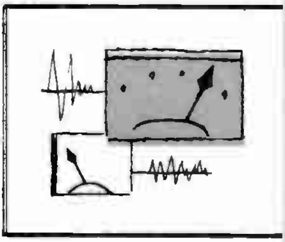
published by gernsback library, inc.
new york, new york

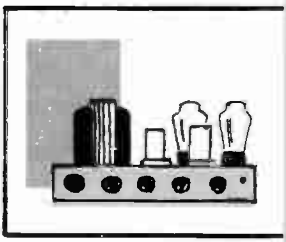
First Printing — December, 1955
Second Printing — August, 1957
Third Printing — June, 1961

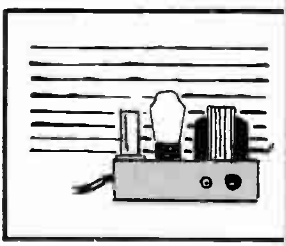
Library of Congress Catalog Card No. 55-11945

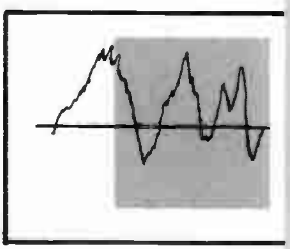
contents

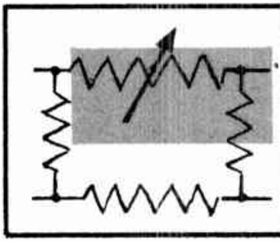


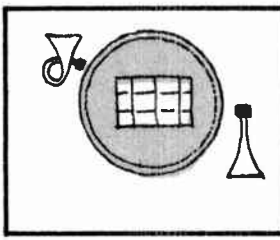


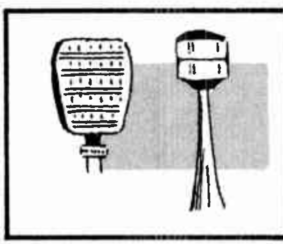


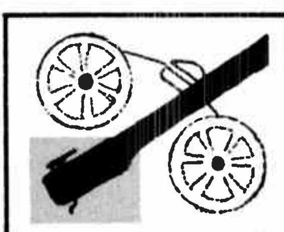


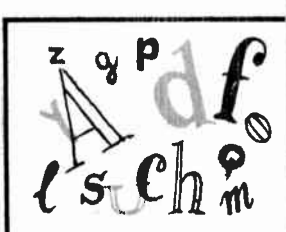












introduction

The art and science of audio engineering has come far since the days of Edison's tinfoil phonograph and Bell's battery telephone. Today the phonograph and telephone industries are just two of dozens which employ sound equipment. Radio would not exist and television would be mute without it. Motion pictures would be in the silent days. Business dictation would be slowed down to a walk. Speakers at a political rally wouldn't be able to project their voices beyond the first ten rows. None of these things is true today, however, because the audio engineer has made the aural communication of ideas and entertainment a commonplace thing.

But, as so often happens in an industry which is quickly called upon to fill so many basic needs, those who deal in sound have not been able to keep pace with their own requirements for fully-qualified personnel. As a result, it has been necessary to borrow trained men from related fields and "retread" them to fill a niche in the audio business. Telegraph operators have become broadcast engineers; radio service technicians have become public-address operators; telephone engineers have become motion-picture sound men; machinists have become phonograph record engineers; the pattern is the same throughout every branch of the industry.

Such a roundabout approach as this is costly in time and money for employers, and for the individuals involved as well. But it does bring into sharp focus the very real need which exists right now in every industry which uses sound, for good men who have a solid theoretical background in audio. The dynamic growth of television, the tremendous promise of tape recording, the awakening interest in high-fidelity reproduction: these are just a few of the encouraging signposts which should guide the man of intelligence and ambition. Nowhere is there a calling which offers greater opportunities for professional growth than in the field of audio.

There is increasing evidence, furthermore, that many persons who have no professional objectives in this lively art nevertheless want to know something about it. Audio experimenters and music lovers have long since learned that a high-quality audio system, like a fine car or a precision watch, requires loving care. And such care requires knowledge.

Whatever the reader's personal aims may be I can think of no better way to begin than by the careful study of Don Hoefler's BASIC AUDIO COURSE. This is a book which should have been written a long time ago, and now that it is here I am happy to note that it is both accurate and readable. Since this is a basic text, the author has carefully avoided higher mathematics and abstract theories, while at the same time presenting a thorough grounding in each facet of the art. Some of the basic laws of sound are proven by a simple mathematical development, but even this can be skipped without undue harm to the reader who is more interested in the "how" than the "why."

I have been personally acquainted with the author for some time, and regard him as exceptionally well qualified to present this subject. He knows the field and he knows how to explain its intricacies. I commend him and his BASIC AUDIO COURSE to you wholeheartedly.

William H. Miltenburg
Manager of Recording
RCA Victor Record Division

The author and publishers acknowledge with thanks the co-operation of the following companies during the preparation of this book: Altec Lansing Corp., Cinema Engineering Co., Electro-Voice, Inc., Jensen Mfg. Co., Karlson Associates, Inc., Klipsch and Associates, Radio Corporation of America, Weston Electrical Instrument Corp.

chapter 1

the nature of sound

Audio, a term dating back to one of our earliest civilizations, originally meant "I hear." The human ear is still the converging point of all of the efforts of the audio art, but the term today implies a wealth of human experience undreamed of even a century ago.

The serious audio student must now appreciate how and why sounds are produced in the first place; how they are propagated; how they may be collected, stored, amplified, transmitted and reproduced and, finally, how they affect the ears and intellect of the listener.

The psychologist defines sound simply as anything which is heard, while the physicist says that it is the vibrational energy which causes the sensation of hearing. But to the audio engineer, both the cause and the effect are equally important.

Nevertheless, right up to the point where sound is perceived by the brain—and perhaps even beyond—it may be assigned a simple one-word definition: *vibration*. In the beginning sound is created by causing an object to vibrate. It may be a bowed or plucked string, the human vocal cords, or even a window pane struck by a baseball. Whatever their source, these vibrations are immediately transferred to the surrounding medium, usually air.

The behavior of sound

While light cannot penetrate opaque objects, but is transmitted through a vacuum, sound behaves in somewhat the opposite fashion. Vibrations set up within a perfect vacuum cannot be heard at all, while steel is a much better transmitter of sound than is air. Consider the velocity of sound in some of the more common materials:

air (at sea level)	1,087	ft./sec.
water	4,708	" "
brick	12,000	" "
glass	16,410	" "
steel	16,500	" "

When sound waves traveling in air encounter another medium, such as the walls of a recording or broadcasting studio, part of the sound is reflected into the air while the remainder is transmitted and absorbed by the wall. Provided the wall surface is at least several wavelengths as compared with the sound, the law of light reflection will be effective, i.e., *the angle of incidence is equal to the angle of reflection.* This is illustrated in Figure 101.

The velocity of sound is determined by the elasticity and density of the medium, and has the following relationship:

$$C = \sqrt{E/d} \qquad (1)$$

where C is velocity; E is elasticity and d is density.

This explains why the part of our sound wave which entered the wall is at that point deflected somewhat off course. If the velocity of sound in the wall is less than in the air, the front of the wave slows up as it enters the wall, and an angle of refraction results which is less than the original angle of incidence.

If the window in a room located on a busy street is opened only slightly, the outside noises will immediately appear much louder and fill *all* parts of the room. But if the shade is drawn so that the sunlight is admitted through only the open part of the window, the light will continue in the form of a beam to the opposite wall and the rest of the room will remain in relative darkness. This ability of sound waves to fan out and bend around objects is known as *diffraction.* We ordinarily pay no attention to the fact that we are constantly hearing sounds from sources which we cannot see, but to the audio man this phenomenon means that he must consider every factor within the acoustical environment surrounding his desired sound source. On a movie set, for example, a group of stagehands might be engaged in a poker game just out of range of the camera shooting an ardent love scene, and the picture would be in no way affected. But if these same men begin to talk, the sound track will be ruined. Unfortunately we have not yet discovered any way around the laws of physics which will enable us to frame in our sound as we can a picture.

When two or more sound waves meet, the alternating pressures of each will sometimes add and at other times subtract from one another and cancel. This phenomenon is analogous to heterodyning in radio-frequency circuits, and the resultant sound waves

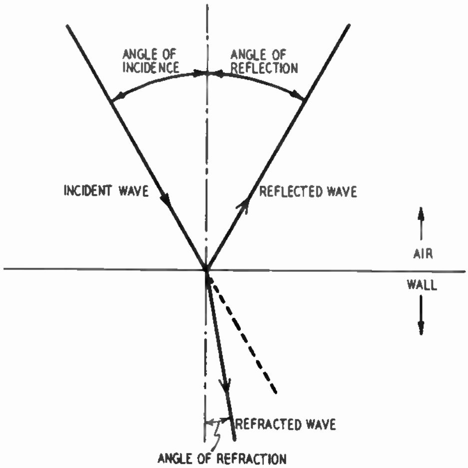


Fig. 101. *Sound, striking a wall, is reflected by it. Some of the sound is absorbed by the wall and a portion of the sound is transmitted by it.*

will be reinforced at some points and partially destroyed at others. These *interference* effects can be caused by waves arriving from several sources or by the combination of direct and reflected waves from a single source. The beats and dead spots produced by interference can be largely overcome by proper acoustical design, coupled with judicious placement of the sound sources.

When the single tone of an audio oscillator is reproduced by a loudspeaker, the moving cone of the speaker sets up variations in the air pressure surrounding it. When the cone moves out, the molecules of air are compressed and the pressure is increased above normal. When the cone moves in, the molecules are rarefied and the pressure reduced. If we could measure the atmospheric pressure at many points along a line in the path of the sound waves, a graphical representation of our readings would provide a curve or waveform such as that in Fig. 102.

Forms of sound

It is seldom that a simple pure tone is encountered in practice, most sounds being waves of considerably more complexity. As a rule sounds will readily fall into one of the three general categories known as *noise, speech* and *music*. Sounds such as the clap of thunder, the ringing of a doorbell or the tinkle of broken glass are called noise, simply because it is impossible to assign any of them a definite place in the musical scale. They are really very complex waves consisting of a large number of pure tones having no harmonic relationship and none of which is predominant enough to give the sound a definite pitch. Because of their complexity, they are useful as a test of audio response.

Speech sounds are simply those which are formed by the human speaking apparatus. The human voice-producing system consists of the lungs, air passages, vocal cords, nasal and throat cavities, tongue, teeth and lips. Linguists say that there are six basic classes of speech sounds, depending upon the way in which the various parts of the vocal system are used. These are: (1) the pure vowels (long and short); (2) diphthongs; (3) transitionals; (4) semivowels; (5) fricative consonants (voiced and unvoiced) and (6) stop consonants (voiced and unvoiced).

Whatever the method of production of speech sounds—through the vocal cords, the lips and teeth or a combination of both—the sounds produced have a more or less definite pitch, being most definite in the case of singing. Since the adult male vocal cords are thicker than those in women and children, they vibrate more slowly and thus produce the lower tones characteristic of men's voices. The average male voice has a fundamental pitch around 128 cycles per second, while that of the average female is around twice that frequency.

Since the human ear is more sensitive to the tones produced by the male voice than those of the female, a woman's speech is generally somewhat more difficult to understand than a man's. This fact imposes somewhat more stringent requirements upon the fidelity of an audio system, but by far the most difficult sounds to handle are those encountered in music.

Characteristics of music

The sensitive listener would not regard as musical any sounds which lacked a definite structure and significance or which failed to fall pleasantly on the ear. The audio engineer, however, not being concerned primarily with esthetics, consequently accepts a much broader definition. In the technical sense a musical sound is any which emanates from a source of *regular* vibration, that is, one which has a definite *pitch* or *frequency*. This is measured in terms of the number of cycles per second, often abbreviated as c.p.s. or sometimes simply "cycles."

A cycle is one complete set of pressure changes through which a sound wave repeatedly passes. Thus in Fig. 102 the cycle comprises an increase in pressure from normal or rest position at 1 to a maximum at 2, followed by a decrease to normal at 3,

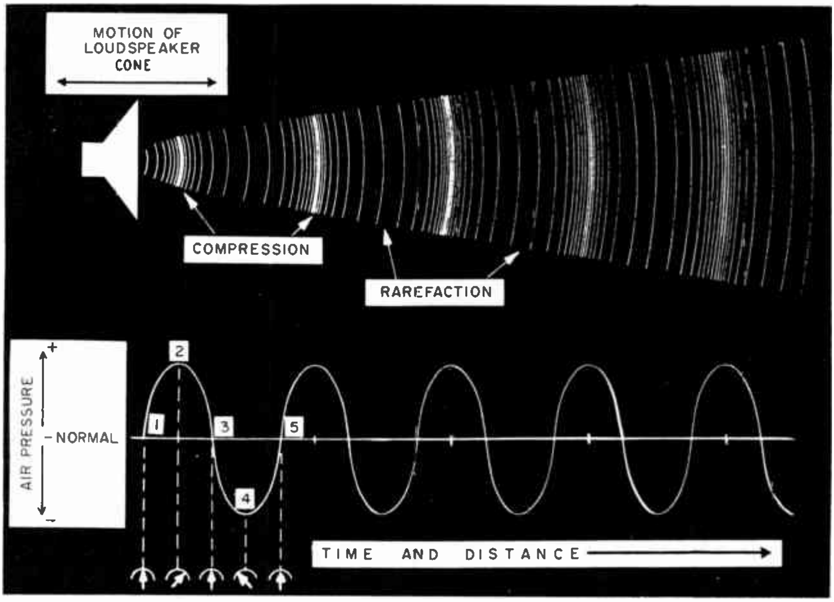


Fig. 102. *The motion of the cone of a loudspeaker produces changes in the pressure of the air surrounding the cone.*

with a further decrease below normal to maximum at 4, and finally a return to normal at 5. This entire series of changes through 1 to 5 is called one cycle. If one second of time is required for this excursion, the frequency of the wave is one cycle per second. If a thousand of these cycles can be completed in a second, the frequency is then 1,000 c.p.s. The fundamental frequencies produced by all of the notes on a standard 88-note piano keyboard are shown in Fig. 103. At first glance it would seem that an audio system which could pass all of these frequencies without alteration would be entirely adequate. But the problem is not quite that simple, as we shall see.

Another important charactertistic of all sound is its loudness: its relative intensity as perceived by the human hearing system.

The higher the pressures built up (i.e., the greater the height of peaks 2 and 4 in Fig. 102), the greater will be the amount of sound power produced. But the human ear is a peculiar device which does not respond in direct proportion to changes in pressure, which means that doubling the power will not double the

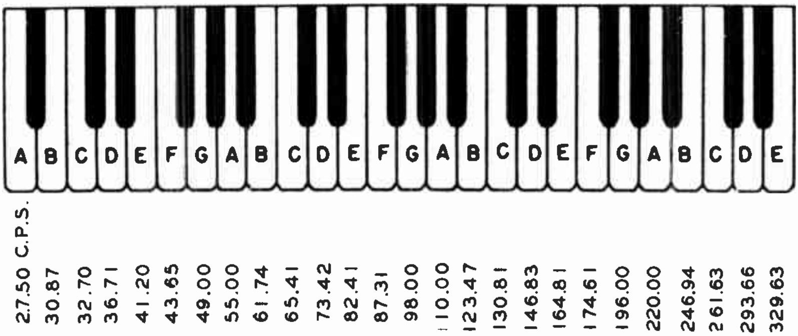


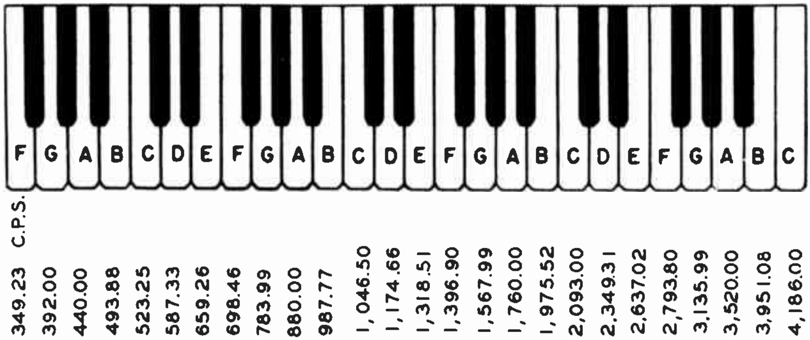
Fig. 103. *This illustration shows a standard 88-note piano keyboard. The figures placed below each key indicate the frequency in cycles per second of each corresponding note. Middle A on the piano has been fairly well standardized at 440 cycles per second (c.p.s.). The piano, in common with all other musical instruments, produces complex tones. These tones consist of sine waves plus a number*

loudness. It is extremely important, then, that we examine these subtle characteristics of hearing after we consider a final important characteristic of the sound itself.

We all know that the middle A (440 c.p.s.) sounded by a piano is quite different from the same note when played by an oboe, violin or any other instrument. Each member of the various families of musical instruments has its own particular tone *quality* or *timbre*. The reason for this is that all instruments produce what are known as *complex tones*. Up to now we have considered only simple tones consisting of a single fundamental frequency but, in fact, most sounds comprise a number of other frequency components known as partials, overtones, harmonics, subharmonics and sum and difference tones.

The most important of these are the harmonics, which are integral (whole number) multiples of the fundamental. Thus any musical instrument playing an A-440 will also produce simultaneously other frequencies including 880, 1,760, 3,520 c.p.s. and so on up to infinity. It is the relative intensities of these various harmonics combining with the fundamental which determine the particular timbre of a given tone. Typical waveshapes of some musical instruments are shown in Fig. 104.

Sometimes subharmonic frequencies are also present, such as one-half, one-third and one-quarter of the fundamental. Some observers feel that many of these components beat against one another to produce sum and difference combinations of themselves. It is therefore clearly obvious that the ideal audio system must be capable of handling a wide band of frequencies of very great complexity, subject only to the limitations of our ears.



of harmonics. The fundamental sine wave combines with the harmonics to produce complex waveforms. It is the presence of these harmonics which gives the piano its particular tone. Frequencies other than harmonics are produced when a key is struck, but the harmonics are the most important. This means that an audio system must be able to pass a wide range of frequencies.

Characteristics of hearing

The ear is a quite remarkable device about which much is yet to be learned. But while we do not know all the physiological factors, we do have rather complete data on the results of its performance.

We know for example that the ear is able to perceive sounds within the range of around 20 to 20,000 c.p.s. Since each doubling of the frequency is denoted as a *musical octave,* the ear has a total response of about 10 octaves. And since the highest tone produced by any present musical instrument is about 5,000 c.p.s., most ears can hear at least the fourth harmonic (four times the fundamental) of any musical tone.

But the ear is not equally sensitive to all sound frequencies throughout its range (see Fig. 105 on page 15) especially when the sound intensity level is low. This is especially significant when you reproduce sound at a level different than that at which it was originally produced. Particularly important at low levels, when low frequencies seem to disappear, this can be corrected by the use of a loudness control. The converse of this is that when two sounds of unlike frequencies are made to sound equally loud, the powers generated by those sounds will be quite different. These facts are very important in the design of audio systems, for they show that even the perfect speaker will require vastly greater amounts of power from the amplifier in order to reproduce bass tones equal in intensity to those in the middle range. They also explain why music which is reproduced at a much lower level than it was originally performed will appear to be very deficient in both the high and low frequencies, with nothing remaining but the middle tones.

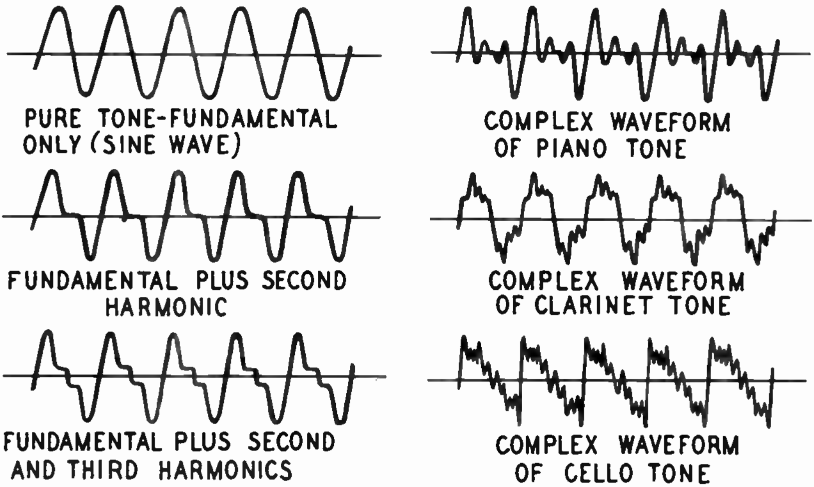


Fig. 104. *The three curves at the left show the formation of a complex waveform through the combination of a pure tone and its harmonics. Representative waveforms of various instruments are shown at the right.*

When the ear is subjected to sounds of rather large amplitude, it will distort in much the same manner as an overloaded amplifier. The result under these circumstances is the production of spurious harmonics and sum and difference tones. Although not present in the original sound, they are actually generated in the hearing organs and are perceived by the brain. Known as *subjective tones*, these sounds can actually appear as re-created fundamental tones when a system of limited range cuts off those fundamentals but still transmits their harmonics. This explains the condition of "false bass," wherein there are regenerated as subjective tones the low frequencies which the equipment itself is incapable of reproducing.

Masking

Another closely related characteristic of the human ear is the phenomenon known as *masking*, which appears as a deafening to high frequencies when in the presence of very loud low-pitched sounds. This is due to the fact that the ear subjectively produces very strong harmonics of the low frequencies, which then interfere with the perception of the higher-pitched sounds. This is why one must shout when attempting to carry on a conversation in a noisy location. But it is always the low-frequency components of the sound which cause the trouble, high-frequency

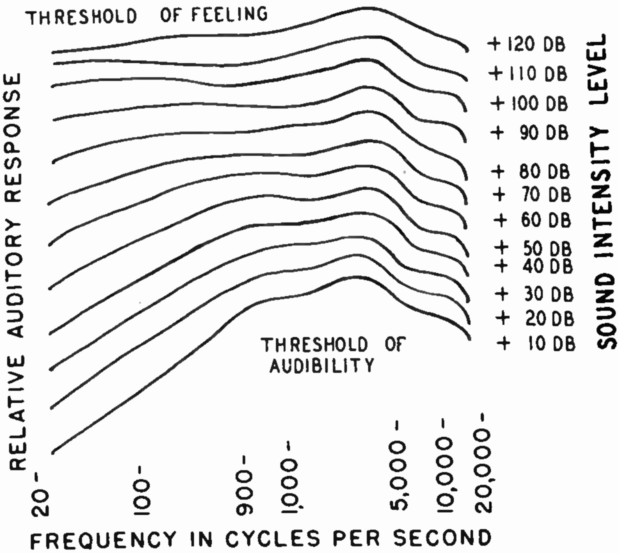


Fig. 105. *Our ears do not respond equally well at all sound intensity levels.*

sounds apparently not masking the lower tones to any degree. Since the subjective tones of the higher frequencies, if indeed they are produced at all, are mostly outside the range of human hearing, there is nothing with which they can interfere. And since low-frequency sounds are relatively so much more powerful, any subjective difference tones would have very little effect.

The physics of music

The science of sound and the study of music have been intimately related for centuries. Many of our fundamental concepts date back to work done about 550 B.C. by the well known Greek philosopher and mathematician, Pythagoras. Some of his work had to do with vibrating strings, the basic principle of an entire family of instruments of the orchestra.

It is well known that if a string is plucked, it will be set into vibration and will in turn set the surrounding air into vibration and thus produce sound. If the string is twanged precisely in the middle, it will vibrate as a whole and produce its fundamental tone. But if the same string is plucked, struck or bowed near one end, it may vibrate in several parts, with nodes (*points of zero vibration*) between the segments. Furthermore it is possible for the string to vibrate simultaneously as a whole and in segments. These conditions are shown in Fig. 106.

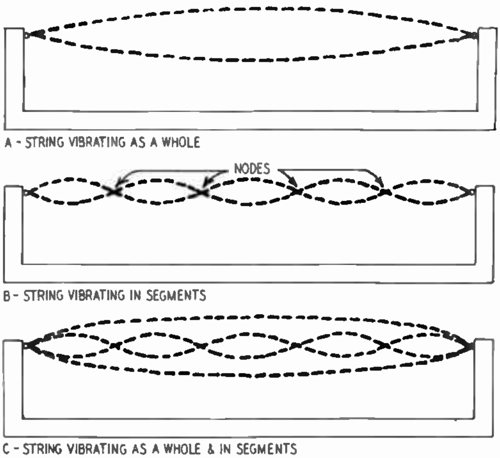


Fig. 106. *This illustration shows the characteristics of a vibrating string.*

Pythagoras discovered that when two strings are under equal tension, but one is exactly half the length of the other, then the shorter one will sound precisely an octave higher than the longer one. As the result of his studies he formulated a law which stated: *The simpler the ratio of the two parts into which a vibrating string is divided, the more perfect is the consonance of the two sounds.* In modern times this is not held to be strictly true, but we have found that any pair of tones whose frequencies are proportional to any pair of the integers 1, 2, 3, 4, 5 or 6 will produce a pleasing and harmonious effect on the ear. Thus the octave ratio (2 to 1) produces a pleasing effect. Similarly, referring to Fig. 103, middle C (261.63) and middle E (329.63) will simultaneously produce a harmonious sound, for the ratio is five to four (329.63 : 261.63 = 5:4). If we add a middle G to the C and E, we will then have a chord with a 6 : 5 : 4 relationship, one of the most common in all music. A chord with this harmonic relationship is known as a *major triad.*

A closer examination of Fig. 103 will indicate, however, that this ratio is never precisely attained. It is always only a close approximation. If we were to begin to construct a musical scale beginning with the note of C, it would have ideally the mathematical relationship shown in Fig. 107. But if we were then to build another scale based upon the key of D and employing the same set of ratios, we would discover some discrepancy in the existing notes, especially in the frequencies of F and C. And we find that what is needed here is a tone approximately midway between F and G, and C and D. These new tones are designated F# and C#. Continuing in a similar manner to build major

C	D	E	F	G	A	B	C	LETTER NAME
DO	RE	MI	FA	SOL	LA	TI	DO	SYLLABLE NAME
1	9/8	5/4	4/3	3/2	5/3	15/8	2	FREQUENCY RATIO

Fig. 107. *In this drawing we see the mathematical relationship between the notes of the musical scale.*

scales based upon each key, we find that it is necessary to add 5 new *semitone* intervals. This then will give us a *chromatic* scale of 12 tones.

But we are still troubled by the fact that the mathematical relationship between each of these intervals is only an approximation of the desired ratios and developed solely for the sake of simplicity. If the complete set of scales were to be built employing the ideal relationships, the result would be a chromatic scale consisting of about *70* notes per octave. Thus our 12-tone chromatic scale is really a compromise which is known as a modified or *tempered scale.* The precise mathematical relationship between each interval is based upon the twelfth root of 2, which is 1.05946. In the tempered scale the frequency rate of any tone can be found by multiplying the frequency of the note immediately below by this factor.

The starting point of all this is now almost universally accepted as a middle A of 440 c.p.s. This is largely an arbitrary designation, of course, as any other pitch in the same region could be used as well. Many other standard pitches have been employed in the history of music, in fact, but A-440 is now generally used as the reference. There still remain a few nonconformists, however, such as the Boston Symphony Orchestra, which generally tunes to an A pitched at 444 c.p.s.

A secondary set of scales, known as the *minor scales,* is constructed upon similar principles, the difference being that the frequency ratios of the minor triads are 10 : 12 : 15. Compositions in minor keys have a somewhat more plaintive character than those based upon the major scales. Any musical instrument playing chromatic scales based upon the twelfth-root-of-2 system can play all of the minor as well as the major scales.

The instruments of the orchestra

The families of musical instruments are usually designated in terms of the methods employed to produce a tone. Stringed instruments which are bowed include the violin, cello and contra bass. Bowing is accomplished by drawing across the strings a bow strung with rosined horsehair. This sets up vibrations because the bow hairs repeatedly stick to the string and then release it. Plucked instruments, such as the guitar, banjo and mandolin, have their strings set into motion by snapping them with a plectrum, which may be either a small pick or the thumb and fingers of the player. Pitch is altered in stringed instruments by pressing down on the strings at predetermined points and thus shortening their length. In every case some form of sounding board is employed which sets into motion a much larger mass of air.

The reed instruments employ one or two pieces of flat thin cane reed, set into motion by blowing. Examples of this group include the clarinet, saxophone, oboe and bassoon. Pitch is altered by varying the length of the air column by opening or closing holes in the side of the instrument.

In the brass instruments, such as the cornet, trumpet, trombone and tuba, the vibrating lips of the player act as the reed. Length of the air column in most of these instruments is varied by means of valves, but in the trombone is changed by means of a sliding tube which telescopes around a fixed tube.

Percussion instruments are those in which a sound is produced by an object being struck. Even the piano falls among the instruments of this group which have a definite pitch, although it uses the string as the vibrating body. Other percussion instruments of definite pitch include the xylophone, bells and chimes. Pitch is determined by the fixed size of the bar or tube being struck, and very little variation in tonal quality is possible. Instruments of this family include the balance of the drummer's gear, such as the drums, tom-toms, triangle, temple blocks and cymbals.

chapter 2

the measurement of sound

Sound, as we know, results only after a physical object has been set into motion. This immediately gives us the idea of work or energy and, going back a step further, for energy to have been expended, force must have been applied. The fundamental unit of force employed by physicists and engineers is the *dyne*.

The classical definition of the dyne, derived from Newton's second law of motion, is the force which, acting upon a mass of one gram for a period of one second, produces a change in the velocity of the mass of one centimeter per second. In practical terms the dyne is so small that it is about the equivalent of the force exerted by a postage stamp lying on one's hand.

In audio work we are more concerned with sound pressure, which physicists tell us consists of force within a given area. In keeping with the centimeter-gram-second system (c.g.s.), we have adopted as the basic unit of sound pressure the *dyne per square centimeter*. At the threshold of hearing (refer again to Fig. 105 in Chapter 1) the sound pressure is about .000204 dyne per square centimeter, while the boom near the muzzle of a 5-inch cannon is around 100,000,000 dynes per square centimeter.

To correlate the values of sound in air with sound in the form of electric power, we often refer to the *intensity* of sound rather than its pressure. The unit then employed is the *watt per square centimeter*.

Amplifiers and loudspeakers are often rated in terms of their power-handling capabilities, for we are primarily interested in the work they can do in setting into motion large quantities of air, thus reproducing sound. For this purpose we employ the familiar unit of electrical power, the watt.

When measuring the performance of an amplifier under steady-state conditions, we commonly employ a voltmeter and express the results in terms of volts or millivolts. This is largely a matter of convenience, since the voltmeter is one of the likeliest instruments to be found in the radioman's kit. Furthermore, the better voltmeters, especially those of the vacuum-tube variety, are quite accurate at audio frequencies.

Logarithmic units of measurement

Up to this point we have been discussing absolute units of sound measurement under certain fixed conditions. But we know that sound waves have a high degree of complexity and that the human ear has its own peculiarities in the perception of sound. Since the ear is the ultimate target of all of our audio efforts, it is necessary for us to consider methods of sound measurement which take into account the nonlinear characteristics of the ear.

Suppose we were to listen to a constant tone at a comfortable level on a pair of headphones and then were to turn up the volume until the change in intensity became just barely apparent. With a typical pair of ears it might be necessary to *double* the power before the change in volume would become noticeable. It would also be necessary to double the power once again to cause another perceptible change and a like amount for each additional change. From this we can infer that the smallest possible change in the amplitude of a sound which the ear can detect is approximately a constant proportion of the previous intensity. And this is the basis of the very important concept that *the loudness of sound as perceived by the human ear is proportional to the logarithm of the sound intensity.*

To understand the basic units of sound measurement employed every day by the practicing audio engineer, it is therefore essential that we have an understanding of the concept of logarithms. The principles are really quite simple, and in practice they are indispensable to the proficient audioman.

Principles of logarithms

The idea of logarithms is simply a special application of the *exponent,* which in turn is a simple bit of mathematical shorthand indicating how many times a number is to be multiplied by itself. For example, $2 \times 2 = 4$ can be shortened to $2^2 = 4$, and similarly $2 \times 2 \times 2 = 8$ becomes $2^3 = 8$. We could then draw a table of the powers of 2, as shown in Fig. 201. From this table we see that 2 to the 7th power equals 128 ($2^7 = 128$), or that $2^{12} = 4{,}096$.

Now let us consider briefly the surprising ways that this table can be used to solve readily some fairly difficult problems:

1. Problem: Multiply 32 by 128.
   Solution: $32 \times 128 = 2^5 \times 2^7 = 2^{12} = 4{,}096$
2. Problem: Divide 1,024 by 128.
   Solution: $1{,}024 \div 128 = 2^{10} \div 2^7 = 2^3 = 8$

Persons unaccustomed to working with exponents and having had some unfortunate classroom experiences with mathematics might very well try to avoid working with numbers that exist in this fashion. And yet exponents are to large numbers what power tools are to hand tools. You can build a house with a handsaw. A power saw is much easier, better, faster. You can multiply, divide or perform any arithmetical operation you want on very large numbers, using many large sheets of paper to do so or you can convert them to exponent form, make the work much easier, and also reduce the chance for error. Just regard exponents as another way of writing the numbers with which you are already familiar. For example, the number 16 can be written as: 16; 2×8; 4×4; 2^4. Quite obviously, there isn't much point to breaking the number 16 down into its factors, nor is there much sense to using exponents as a substitute for such a small number. Exponents are of greatest value when working with large numbers. Even if you do not become familiar enough with exponents to handle them with ease, the word itself crops up so often in audio literature that for this reason alone you should try to make exponents more than just another mathematical term.

There are a few simple rules involving exponents, just as we have rules for all mathematical operations. Thus, when multiplying numbers you add exponents. This sounds like an odd rule, but it works and works well. Consider the first problem in an earlier paragraph: $32 \times 128 = 2^5 \times 2^7 = 2^{12} = 4{,}096$. Let's take this step by step. First we converted 32 and 128 to their exponent form by looking in Fig. 201. When we did this, we had $2^5 \times 2^7$. The next thing we did was to *add* the exponents. This resulted in 2^{12}. Note that we did not multiply the base number 2 but just carried it along. Finally, we looked in Fig. 201 to find the number corresponding to 2^{12} and ended up with

4,096. Of course you could have multiplied 32×128 much faster without exponents, but small numbers were deliberately selected to make the explanation easier.

By now you probably have the general idea for the solution of the second problem—the division of 1,024 by 128. When dividing numbers subtract the exponents. Now we can go a little faster. Find the exponents corresponding to these numbers in Fig. 201, subtract the exponents, look up the resulting exponent in Fig. 201 and you have your answer.

The root of an exponent is the exponent divided by the root. Thus, if you wanted to find the third root of 2^9, just divide the exponent by 3.

For example: $\sqrt[3]{512} = \sqrt[3]{2^9} = 2^3 = 8.$

Similarly, complex problems involving any combination of these operations may be solved, *provided* all of the values could be expressed in terms of the power of 2 and our table was complete enough to cover all the figures involved. However, since 10 is the magic number occurring throughout our decimal system of computation, it would seem to be more convenient to express our numbers in terms of the powers of 10 and then to solve in the manner just shown. The section labeled "Exponents" in Fig. 201 shows the number 2 raised to various powers, while directly below are listed the equivalent values. A set of tables based upon the powers of 10 is known as *common logarithms* or logarithms to the *base 10*, since they are based upon the common number 10. Thus we can construct a table of the common powers of 10 in the same manner that we did for the table of the number 2. This table is shown as Fig. 202, and we can immediately see that the resultant values are much simpler. For example:

$10^0 = 1$	$\log 1 = 0$
$10^1 = 10$	$\log 10 = 1$
$10^2 = 100$	$\log 100 = 2$
$10^3 = 1{,}000$	$\log 1{,}000 = 3$

Since we have jumped rather rapidly from exponents to logarithms (abbreviated as log) a word or two of explanation is in order. Just as a simple example, consider a number such as 10^3. This number really consists of two parts – the number 10 and the number 3. As individuals in their own right, they are entitled to names so that we can identify them. The number 10 is called the *base* and the superscript number sitting up in the air is called the *exponent*. Now by looking at Fig. 202 we can easily see that $10^3 = 1{,}000$ so we do have two different ways for writing this number.

It is often necessary for us to jump back and forth between numbers (such as 1,000) and their exponent form (such as 10^3). When working with numbers you must always be ready with two items: you must first of all have the number, and then you must

EXPONENTS	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}
EQUIVALENT VALUE	2	4	8	16	32	64	128	256	512	1024	2048	4096

Fig. 201. *In this illustration we have the number 2 using various exponents. Exponents are particularly valuable when dealing with very large numbers.*

be prepared to issue instructions. By itself the number 1,000 means nothing. If you hand out that number, you must be ready to say what you want done with it. Should you multiply it by some other number? Extract the square root? Divide it, subtract it—whatever the operation, you must be specific.

LOG_{10}	0	1	2	3	4	5	6	7	8	9	10	11	12
NUMBER	1	10	100	1000	10,000	100,000	1000,000	10,000,000	100,000,000	1,000,000,000	10,000,000,000	100,000,000,000	1,000,000,000,000
EXPONENT	10^0	10^1	10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}	10^{11}	10^{12}

Fig. 202. *Logarithms are simply a short-cut method for doing arithmetic that would otherwise be tedious and difficult.*

Our particular problem is to issue instructions for the conversion of the number 1,000 to its exponential form. The way in which we do this is to put the word logarithm (or log) in front of the number. Thus,

$$\log 1{,}000 = 3$$

Note that we have omitted the base number 10 and have just written the exponent. The base number is understood to exist and would be needless repetition. For example, the number 6 is actually 6×1, but all of us take the number 1 fairly much for granted, omit it whenever possible and just write the number 6.

Going further, we can deduce that a number lying between 100 and 1,000 would have a logarithm (log) somewhere between 2 and 3; that is, the log of a number between 100 and 1,000 would have a *fractional exponent* of something greater than 2, but less than 3. The log of 700, for example, is 2.8451, and we can go

No.	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5083
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396

Fig. 203. *This is a typical table of mantissa. A mantissa is the decimal portion of a common logarithm.*

even a step further and prove that the log of 70 is 1.8451 and log 7 = 0.8451. The part of the logarithm following the decimal point is identical in each case, because the numbers differ only in the amount of zeros following the significant figures.

It is convenient, therefore, to divide a logarithm into two parts, the part following the decimal point (8451 in each of the illustrations given) being called the *mantissa* and the part of the logarithm preceding the decimal point (0, 1 and 2 in the examples given) being known as the *characteristic.* The latter is the part which will tell us the number of decimal places in the answer. Thus in the examples it was the mantissa 8451 which indicated that the corresponding number, called the *antilogarithm,* was 7 with an unspecified number of zeros either preceding or following. Then, when the characteristic was added, it indicated whether the number was 7.0, 70.0 or 700.0.

The characteristic of any logarithm can be conveniently determined by the application of one of the following rules:

1. The characteristic of the logarithm of a number greater than 1 is positive, and is one less than the number of places to the left of the decimal point.

2. The characteristic of a number less than 1 (i.e., a decimal fraction) is negative and is one more than the number of zeros immediately to the right of the decimal point. For example, $\log_{10} .05 = -2.6990$.

The values of mantissas can be found through rather complex mathematical operations, but since they are constants this work has already been done and presented in tabular form as shown in Fig. 203. Using this table, let us determine the logarithm of the number 483. In the left-hand column of the tables, designated "No.", will be found the first two digits of the number. The third digit is located in the appropriate column to the right. To find $\log_{10} 483$, then, we first find 48 in the left column of the table and then move across to intersect the column headed 3. At this point we find that the mantissa of the logarithm is 6839. However, since 483 has three places to the left of the decimal point, we determine from rule 1 that the characteristic is 2, with the result that $\log_{10} 483 = 2.6839$.

To find the logarithm of a number less than 1, such as .000286, we quickly find from the table that the mantissa is 4564, and rule 2 informs us that the characteristic is –4. Hence the logarithm of .000286 is –4.4564.

Since the tables of Fig. 203 have been carried only to three significant figures, the greater accuracy needed in the case of larger numbers can be obtained through the use of more complete tables. If no such tables are readily available, however, it is possible to calculate the logarithms of larger numbers through a process known as *interpolation.*

Suppose that we wish to find the log of 4943. Obviously, since this number lies between 4940 and 4950, its mantissa likewise must lie between the mantissas of those two numbers. From the table we find that these mantissas are .6937 and .6946, respectively, with the difference between them (.6946 – .6937) being .0009. This figure is known as the *tabular difference.* We can also see that 4943 is 3/10 of the distance between 4940 and 4950, and we infer from this that the desired mantissa is 3/10 of the tabular distance. The mantissa we want is therefore .6937 plus 3/10 of .0009, or .69397. The characteristic is 3, with the resultant $\log_{10}$ 4943 = 3.69397. This same process of interpolation can be employed with any number of integral places greater than three.

It is equally important to be able to find the antilogarithm, employing a method which is the inverse of that just described. Suppose it is necessary to find the number denoted by the logarithm 4.5729. Consulting the log table we find the mantissa 5729 corresponds to the digits 374, and from the characteristic we know that the number has five integral places. Hence the number whose antilog (to the base 10) is 4.5729 = 37,400.

Interpolation must be used in determining antilogarithms whenever the mantissa of the log does not appear in the tables. For example, find the antilog_{10} of .0905. The table shows no such mantissa, the closest being 0899 and 0934, with the corresponding numbers being 123 and 124. The difference between 0934 and 0899 is 0035, and the difference between 0905 and 0899 is 0006. Hence the number corresponding to 0905 is found by adding 6/35 or .17 to the number, giving 123.17. But since the characteristic is 0, the required number is 1.2317.

The decibel

Now we come to the practical audio application of logarithms to the measurement of sound. We have already seen the necessity for this as being dictated by the approximate logarithmic characteristic of the human ear. In recognition of this fact a logarithmic unit known as the *bel* (named in honor of Alexander Graham Bell) was adopted as the standard unit of measurement of sound levels, superseding the earlier "transmission unit." This unit, universally adopted by international agreement in 1928, is still the fundamental means of sound measurement. It was soon found, however, to be rather unwieldy, since the ear readily detects differences in volume level of 0.1 to 0.3 bel. Thus, to avoid the use of cumbersome fractions whenever making sound level calculations, the audio engineer customarily employs the *decibel* or one-tenth of a bel.

The decibel is not an expression of a fixed amount of volume level (except under certain special conditions which will be described later) but is normally an indication of the *ratios* between the powers, voltages or currents at two given points in an audio transmission system. It is extremely useful in expressing the gain of an amplifier or the losses in attenuators, filters and equalizers.

The bel is defined simply as the common logarithm of the ratios of the powers at two given points:

$$\text{bel} = \log_{10} \frac{P_1}{P_2} \tag{2}$$

However, since there are 10 decibels in each bel, the basic decibel equation becomes

$$\text{db} = 10 \log_{10} \frac{P_1}{P_2} \tag{3}$$

In each case it is good practice to regard P_1 as the larger power, not necessarily the input power. In the case of a loss device such as an attenuator, P_1, the larger power, is actually at the input, the output power P_2 being less by the amount of circuit loss. But in an amplifier if P_1 is the input power, then P_2 will be greater by the amount of amplifier gain. Then the quantity (P_1/P_2) woud be a number less than one, and its logarithm would have a negative characteristic. All of this difficulty can be avoided by making certain that the numerator of the fraction is always the larger number, for the facts of the problem itself will always indicate whether the calculated decibels represent a gain or a loss.

As a practical example of decibel calculation, suppose that a power amplifier has a rated output of 40 watts when 0.5 watt is fed into it. What is the gain of the amplifier in db? Substituting values in formula (3) we get:

$$\begin{aligned}\text{db (gain)} &= 10 \log_{10} \frac{40}{0.5} = 10 \log_{10} 80 \\ &= 10 \times 1.9031 = 19.03\end{aligned}$$

When we wish to determine decibels in terms of voltage or current ratios, we substitute the standard power equations for P_1 and P_2. Then we have:

$$db = 10 \log_{10} \frac{\dfrac{E_1^2}{R_1}}{\dfrac{E_2^2}{R_2}} \qquad (4)$$

and

$$db = 10 \log_{10} \frac{I_1^2 R_1}{I_2^2 R_2} \qquad (5)$$

In the case of *equal impedances* at input and output, that is, when R1 = R2, these equations reduce simply to

$$db = 20 \log_{10} \frac{E_1}{E_2} \qquad\qquad db = 20 \log_{10} \frac{I_1}{I_2}$$

Here R1 and R2 have cancelled and the exponents affecting the voltage and current values have been removed by the simple operation of doubling the factor of the logarithm.

Since the decibel is logarithmic in character, a series of gains and losses in transmission are totaled simply by algebraic addition. For example, consider Fig. 204, showing a cascade of an amplifier of 50-db gain, followed by an equalizer with an insertion loss of 20 db, followed by an amplifier having 20-db gain and finally another amplifier with a gain of 24 db. The total gain of the entire system is then $+50-20+20+24 = 74$ db. In this case the second amplifier is probably a booster, specifically designed to overcome the insertion loss of the equalizer, and included as an integral part of the equalizer unit.

Another unit of sound intensity, now rather rare, is the *neper*. This is based upon the earlier system of logarithms introduced in the seventeenth century by John Napier which does not use the common log base of 10. The neper formula, based upon Napierian logarithms, is

$$N = \tfrac{1}{2}\log_e \frac{P_1}{P_2} \qquad (6)$$

Since nepers may be encountered in old texts or in a few foreign circuit diagrams, it is convenient to know that they may be converted to decibels simply by multiplying them by the factor 8.686. Db may be converted to nepers by multiplying by 0.1151.

Reference levels

In many cases it is desirable to employ the decibel system to indicate an absolute level of sound power rather than a ratio. This can easily be done if we adopt a standard *reference level* as a basis of comparison. Thus if it is agreed that 0 db shall represent a power of .006 watt in a circuit resistance of 500 ohms, then this figure can be substituted for P_2 in our standard decibel equation and P_1 can then be expressed in terms of absolute decibels rather than watts.

This reference level of 6 milliwatts across 500 ohms was actually employed for many years, especially in the telephone industry, for it represented the amount of power required for the satisfactory operation of the equipment then in use. It is still being used widely in the motion picture industry, but elsewhere it has been largely superseded.

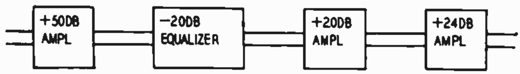


Fig. 204. *The use of decibels (db) makes it easier to determine overall loss and gain.*

RCA introduced a 12.5-milliwatt reference, which was used for a time by broadcasters and recording companies, but the reference most commonly employed today is 1 milliwatt of power in a 600-ohm circuit. This value is designated as *zero level.* Thus the statement that the signal level out of an amplifier is +12 dbm means that the output is 12 db above the power of 1 milliwatt in 600 ohms. The use of the 1-milliwatt reference has the advantage that mathematical calculations involving it are exceedingly simple since it is a unit quantity. Furthermore, it is so small that the majority of devices will have a positive sign with respect to it. Thus most pieces of equipment have outputs which are said to be a number of db *above* zero level. But if a device has an output which is less than the 1-milliwatt reference, then this is said to be a number of db *below* zero level. It should be understood that this does *not* imply a negative quantity nor does it necessarily indicate a loss in a given device. It simply means that the level of audio power at a given point in a circuit is less than that of the zero-level reference point of 1 milliwatt in 600 ohms.

To round out the absolute values of 0 dbm, it is helpful to know the voltage across a circuit at this level as well as the current flow. Since in an a.c. circuit $E = \sqrt{PZ}$, then in this case $E = \sqrt{.001 \times 600} = \sqrt{0.6} = 0.775$ volt. Similarly, solving for $I = \sqrt{P/Z}$, the rate of current flow is .00129 ampere or 1.29 milliamperes.

Level diagrams

A complete picture of an audio system and the way in which a signal passes through it may be seen through the combination of a block schematic with a level diagram, as shown in Fig. 205. Directly below each unit in the channel is indicated the peak operating level at that point under normal conditions. The levels representing fixed gains or losses are connected by vertical lines, while those which are variable due to a fader or volume control are shown as sloping lines. The extent of the slope does not necessarily, however, indicate the extent of the possible loss in the device. Thus in Fig. 205 the gain control is shown as having a loss of 20 db, which is the amount of loss under peak signal conditions. But if this control were to be closed so that no signal were permitted to pass through, then the loss would become infinite.

Measuring instruments

The measurement of sound levels in an amplifier can be accomplished with a simple rectifier type a.c. voltmeter, provided that only a steady tone of constant level is under consideration. However, the nature of a sound program is very much more complex, having frequency components from 20 to 20,000 c.p.s. or more under highly transient conditions. In applications where it is required to know at all times the level of these constantly changing sounds, it is necessary to employ a special type of meter commonly called a volume indicator or VI meter.

In a broadcast or recording control room, the engineer may adjust the balance of his various sound sources by ear, but he must also maintain an adequate program level to keep well above the innate noise of his audio system while at the same time keeping just below the overmodulation point. He accomplishes this by properly adjusting the gain controls while making constant reference to the VI meter.

This meter must have a very quick response to enable it to respond to all of the rapid variations in program level, but at the same time it must not operate so rapidly that the eye is unable to follow it. A standard meter having these desirable characteristics has been devised and is now in general use throughout the broadcast, recording and telephone industries. In the motion picture field, however, conformity has been somewhat slower, many of the studios still adhering to the old db reference of 6 milliwatts measured by a meter of much more rapid ballistics.

The new standard meter, introduced in 1940 and known as the VU – (volume unit) meter, is a copper-oxide rectifier type instrument with a dynamic characteristic in which, upon the instantaneous application of a signal, the pointer will swing to 99% of the exact value in 0.3 second ±.03 second, and then will overswing by 1 to 1.5%. It must also have a frequency response flat within 0.5 db from 25 to 16,000 c.p.s.

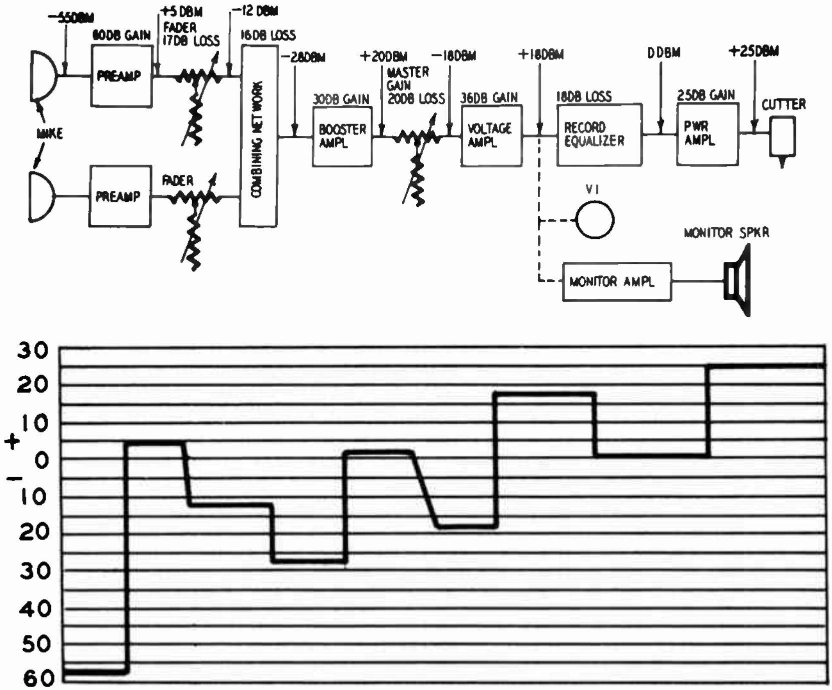


Fig. 205. *Block diagram of an audio-amplifier system. The graph indicates the loss or gain of the various units in the system.*

A typical VU meter scale is shown in Fig. 206, which indicates sound level both in terms of volume units and percentage of modulation. While a difference in volume units is equivalent to the same difference in decibels, the new designation should properly be used only in connection with the VU meter. Since the power in complex program material is constantly changing in an unpredictable and nonrecurring fashion, it is only when a steady signal is applied to the meter that the indication in VU can be taken as numerically equal to dbm.

Under such a steady state condition the meter indicates 0 VU when reading a signal which dissipates 1 milliwatt in a 600-ohm circuit. Usually the meter also contains a multiplier circuit which will insert losses in successive steps of 2 db. Thus, if it is desired to maintain a peak program level of +6 VU, the multiplier will be set to +6 and thereby insert a loss in the meter circuit of 6 db. Then the basic instrument will not indicate 0 level until peaks of +6 VU are reached.

Miscellaneous volume indicators

The cathode-ray oscilloscope makes an excellent volume indicator, being capable of showing instantaneous peaks at all audio frequencies. It is practically inertialess, and its usefulness is limited only by the ability of the eyes of the operator to perceive and retain the rapidly changing information presented on the screen.

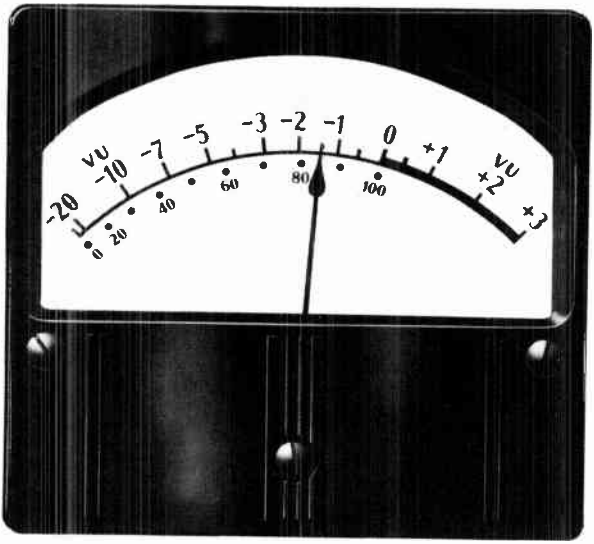


Fig. 206. *Volume-unit meter. The upper scale indicates VU and the lower scale shows percentage of modulation.*

Another device (commonly found in amateur recording equipment) employs the neon tube, which fires whenever a predetermined voltage is impressed across it. Since this tube strikes at peak potential, it has the important advantage that it can be used as the basis of a peak-reading instrument. But, as used in home equipment, with only one or two tubes, it has the disadvantage that no one ever knows for certain whether the lamp is igniting at about the desired peak or whether it is actually being caused to fire by a considerable overvoltage.

chapter

3

audio-frequency amplifiers

An audio amplifier is one which is designed to amplify sound frequencies in the form of feeble voltages up to the point where they can be reconverted into usable amounts of acoustical power. The important requirements for this class of service are:

1. The system must be capable of passing the full range of audible frequencies without discrimination or alteration. A frequency response of from 30 to 15,000 c.p.s. or better is highly desirable.

2. The equipment must operate with equal ease throughout a wide range of amplitudes. Since a full symphony orchestra may generate peak powers on the order of 70 watts while a solo instrument playing pianissimo may develop only a few milliwatts, it follows that the system must operate within a dynamic range of 50 db or better.

3. The output waveform must be as nearly as possible an undistorted reproduction of the original. At the present state of the art, total distortion in an excellent amplifier is 1% or less.

4. The amplifier gain must conform to a given set of conditions in terms of the desired amount of output power to be delivered from the available input voltage.

5. Noise and hum must be absolutely minimum, preferably 65 to 70 db below peak signal level.

6. Input and output impedances must be designed to match those of the devices immediately preceding and following the amplifier equipment.

Classes of audio amplifiers

Any vacuum-tube amplifier may be thought of as a power inverter in that it receives a steady supply of d.c. power from its own power supply system and then converts this to a.c. power in accordance with an a.c. signal at its input. This a.c. output may be used, depending upon the purpose of the tube, to deliver a larger voltage across an impedance or to supply power to a load (such as a speaker). If the tube is used simply to develop a voltage which will be impressed across the input of a succeeding amplification stage, it is then known as a *voltage amplifier*. If it is required to deliver rather large amounts of power to an electromechanical device such as a speaker or disc-recording cutter, it is called a *power amplifier*. In either case, each stage of amplification must have an input coupling device, an output coupling device and a suitable source of power for operating the tube.

The distinction between voltage and power amplifiers is not always clear cut. Generally, the tube immediately preceding the speaker and feeding power to it is called the power amplifier, all other tubes in the system being known as voltage amplifiers. However, all power amplifiers also give some voltage amplification, and in some instances a voltage amplifier may also be called upon to deliver a modest amount of power. Often erroneously referred to as a power tube, the rectifier has nothing to do with either voltage or power amplification. Its job is to rectify, to change a.c. to d.c. and that is all.

The Institute of Radio Engineers (IRE) has adopted designations for four basic classes of amplifier service. These are known as classes A, AB, B and C. Class-A operation is often defined as that in which the grid bias and a.c. signal voltages are such that plate current flows *continuously,* with or without signal input. Much more important is the fact that the tube is operating over the *linear* (straight line) portion of its dynamic characteristic curve. Under these conditions the output waveform should be a precise facsimile of that at the input. Thus, class A will deliver the most faithful reproduction of all classes of amplifiers.

In class-AB operation the grid is made considerably more negative (stronger negative bias) than in class A and a fairly large signal is fed to the input, sufficient to drive the grid positive at the peak of the positive half of the wave. Plate current flows for *less than the full 360° of the input cycle,* resulting in waveform distortion. This class of operation is unsatisfactory for audio use, except when two tubes are connected in what is known as a push-pull circuit.

A class-B amplifier has its grid biased to *cutoff;* that is, the grid voltage is so negative that no plate current will flow under no-signal conditions. Since additional negative voltage will have no effect on a plate current which is already zero, only the positive alternation of the input cycle will cause a current flow in the output. Thus the class-B amplifier might be thought of as a half-wave rectifier. This system is also totally unsatisfactory for audio except in a push-pull circuit.

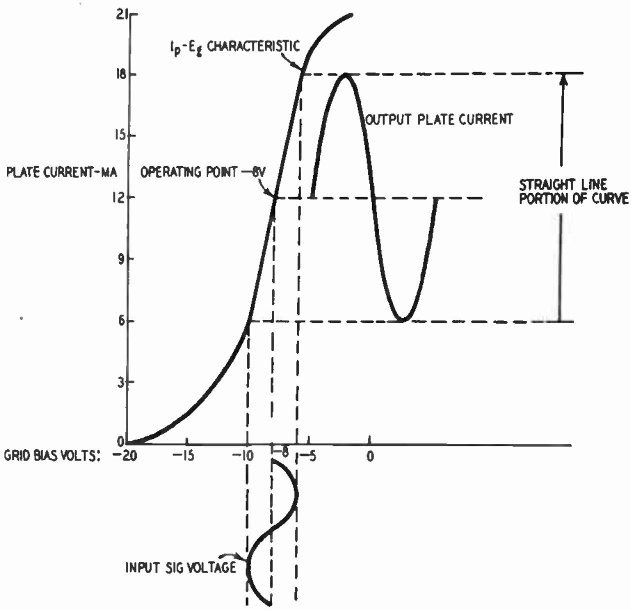


Fig. 301. *The class-A amplifier produces an undistorted output waveform but is characterized by very low efficiency.*

Class-C operation uses tubes biased considerably beyond cutoff, with the result that plate current flows for *less than half of the input cycle.* Even when used in push-pull, class-C operation requires such special circuitry at the output for a barely acceptable waveform that it is hardly ever used for audio work. Class-C amplifiers are most often used in the r.f. amplifier stages of transmitters.

The class-A amplifier

The operation of the class-A amplifier can be better understood by reference to its characteristic curves, an example of which is shown in Fig. 301. This shows how the plate current varies with grid voltage, the plate voltage remaining constant at a given value. This particular curve would be one of an entire family for a given tube wherein each curve would be predicated upon a different value of plate voltage.

With the fixed negative potential on the grid—known as the grid bias—amounting to −8 volts, the plate current will remain steady at 12 ma. The point at which these two values intersect on the curve is known as the *operating point*. Class-A operation, then, occurs when an operating point is selected which is approximately at the middle of the straight-line portion of the characteristic curve. When an a.c. signal voltage is impressed upon the

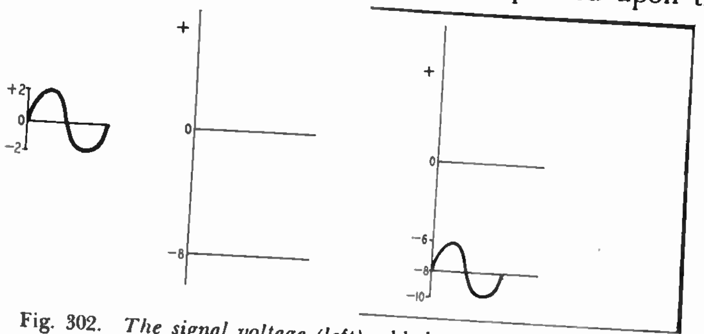


Fig. 302. *The signal voltage (left) added to the d.c. bias (center) gives us an a.c. wave with a d.c. component, (right).*

grid, it alternately adds to and subtracts from the 8-volt bias point of Fig. 301. A signal of 2 volts peak would cause the grid to swing between −6 and −10 volts (see Fig. 302) and this in turn causes the plate current to swing between 6 and 18 ma. The class-A amplifier tube always has a negative bias stronger than the maximum positive peak value of the signal. At no time is the signal able to make the control grid positive.

With a properly selected operating point and an input signal which is not sufficient to drive the operation into the bent portions of the curve, the output will be a perfectly enlarged replica of the input. This means complete freedom from distortion, which of course is the desired ideal.

The class-B amplifier

Class-B operation is consistently different from that of class A, as seen from Fig. 303. In this case the same tube is more heavily biased (very nearly to cutoff) so that the negative pulse of the plate current flow is practically eliminated. Obviously the output wave is far from an exact replica of the input, which means that severe distortion is present. Much of this distortion can be eliminated, however, by the use of the push-pull circuit. This uses two tubes so connected that each of them handles only half of the signal, one tube amplifying only the positive halves of the wave while the other one amplifies the negative halves. The outputs of the two tubes are combined in a common output circuit, where a voltage is developed which has reasonably low

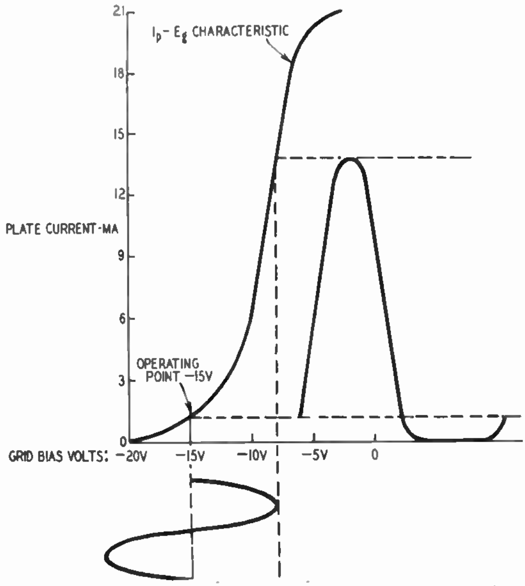


Fig. 303. *Class-B operation uses more of the tube's characteristic curve, is more efficient, but produces more distortion than class A.*

distortion. This effect is illustrated in Fig. 304. The advantage of the class-B system is its greater efficiency compared to class A. Thus for a given amount of power output, less stages of amplification are necessary, resulting in a simpler and cheaper system. Although its fidelity is not perfect, the class-B audio amplifier is quite acceptable for less critical requirements such as public-address amplifiers or as modulators of amateur radio transmitters.

The class-AB amplifier

The fundamental difference between a class-A and a class-B amplifier is in the amount of negative bias with which we start.

As the bias is increased, the operating point on the characteristic curve is moved downward toward the cutoff point. If we assume that class-A operation places the operating point at the exact center of the characteristic curve and that class-B puts that point down at the bottom of the curve, we then have what might be

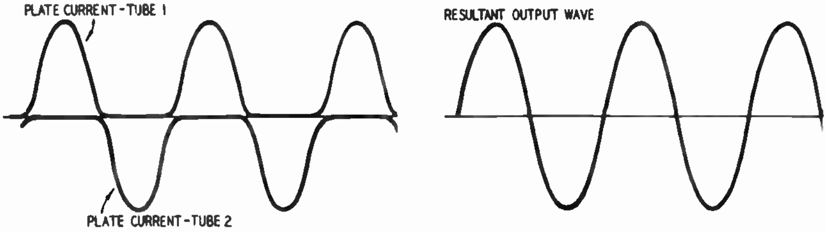


Fig. 304. *The distortion of class-B operation can be largely overcome through the use of a push-pull circuit.*

considered as two limits of operation. As you might readily expect, we can compel an amplifier tube to work somewhere in between these extremes. Such an amplifier tube would have some of the characteristics of both class A and class B, and quite logically is referred to as class AB.

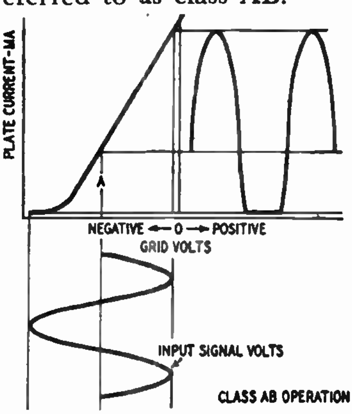


Fig. 305. *In the illustration at the left we have a graph indicating class AB operation. The class of operation of a tube is determined by the amount of d.c. bias existing between control grid and cathode of the tube. As the bias is increased (is made more negative) the operating point is moved down on the characteristic curve (toward the left). In this particular case the bias is halfway between that required for class A and class B — hence the name class AB. The operating point on the characteristic curve is immediately above letter A shown in the drawing.*

The relationship between the input signal and the output voltage of a class AB amplifier is shown in Fig. 305. Note the limitation we have imposed on ourselves. We have deliberately restricted our use of the linear portion of the characteristic curve. We still have the advantage of that part of the curve that is *above* the operating point, but the effect of moving the operating point down on the curve means that we are now using the non-linear section of the characteristic. The result is readily seen in the illustration. The top half of the output signal is linear (not distorted) but the lower half does not resemble the input signal waveform. As in the case of class-B operation, this defect is readily overcome through the use of push-pull tubes.

Since vacuum tubes working in class A have the decided advantage of distortion-free operation, you may very well wonder why we bother with other classes (such as AB, or B) when such classes are known to produce distortion, plus the requirement for an additional tube for push-pull operation. The answer lies quite simply in power output. Fig. 306 gives comparison graphs between class A and class AB amplifiers. In Fig. 306-a we see that the amount of signal swing (signal amplitude) is strictly limited

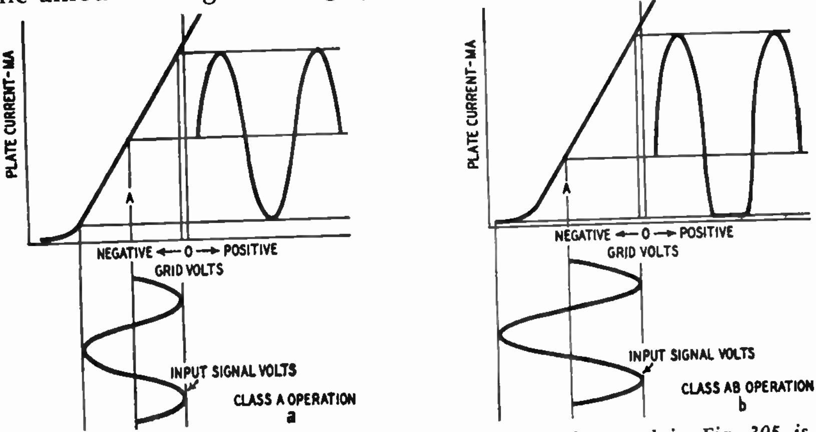


Fig. 306-a,-b. *The drawing for class AB operation illustrated in Fig. 305 is repeated here so that you may compare it with class A (shown at the left). The letter A is directly below the operating point.*

by the length of the linear part of the characteristic curve. If the signal strength is deliberately increased, as in Fig. 306-b, the bottom (and sometimes the top) of the output waveform is clipped. With a class-AB amplifier you can have a much greater signal swing before you reach the bent top portion of the characteristic curve. We don't worry about the lower part of the waveform since two tubes in push-pull will take care of that problem.

Class-AB amplifiers can be subdivided into class-AB_1 and class-AB_2. These amplifiers (AB_1 and AB_2) operate somewhere between class A and class B. The distinction between class AB_1 and class AB_2 is a matter of grid bias and signal swing. If the amount of negative bias is always greater than the maximum positive voltage of the signal, the control grid will always remain negative. This is class AB_1. If, however, the positive peak of the signal is larger than the negative bias, the control grid will be made positive. This means that the control grid of the tube will draw current for a small portion of the time. Such operation is called class AB_2.

The direct-coupled amplifier

The simplest type of audio amplifier is one which has only a single tube, the source signal being applied across its input and the load such as a loudspeaker being connected to its output. However, there has not as yet been invented any single tube with

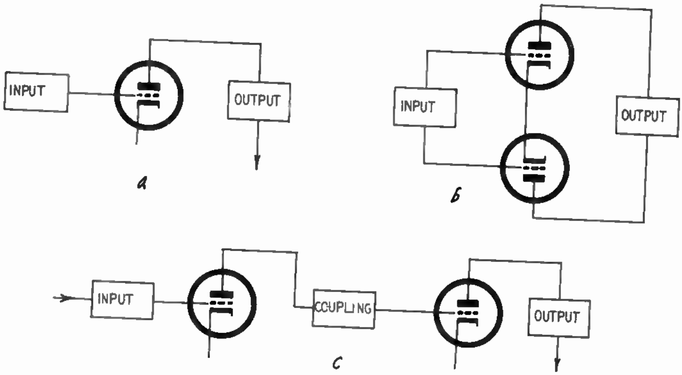


Fig. 307-a,-b,-c. *The circuits shown in a and b provide only a single stage of amplification. In c we have a cascade arrangement of two stages of amplification.*

sufficient amplification combined with the necessary large power-handling capabilities to permit suitable operation of one-tube amplifiers. Therefore, it is customary to connect one or more tubes following the initial tube to provide additional amplification and power. Each single tube or combination of tubes which provide a single stepup in the signal constitutes a *stage* of amplification. Thus, if two or more tubes are so connected that they work together to provide a given amount of amplification from a common input to a common output, they provide only a single stage of amplification. See Fig. 307-a,-b. When a number of stages of amplification are connected together in sequence, they are said to be connected in *cascade* (Fig. 307-c) and the particular method of connection between stages is known as the *coupling*. In audio work, usually all stages but the final one are employed as voltage amplifiers, as contrasted against high-power radio transmitters which may have several power stages preceding the final one. The basic differences in circuitry and operation between voltage and power amplifiers are illustrated in Fig. 308-a,-b.

The simplest method of interstage coupling occurs when the plate output circuit of a stage is connected directly to the grid input of the succeeding stage as shown in Fig. 309. The particular advantage of this circuit is its very excellent frequency response and low distortion. This is especially true at the lower frequencies, where it is very difficult to obtain by other coupling methods a high degree of amplification without encountering considerable phase distortion. This amplifier operates well all the way down to 0 c.p.s. (direct current).

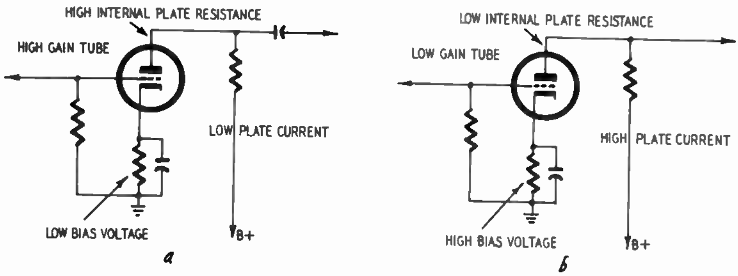


Fig. 308-a,-b. *The circuit at the left is a voltage amplifier; that at the right is a power amplifier. The circuits are drawn the same way, but component values and type of operation differ.*

One of the greatest disadvantages of this system is its requirement of exceedingly high voltages. Since the grid of each succeeding stage is at the same positive potential as the plate of the pre-

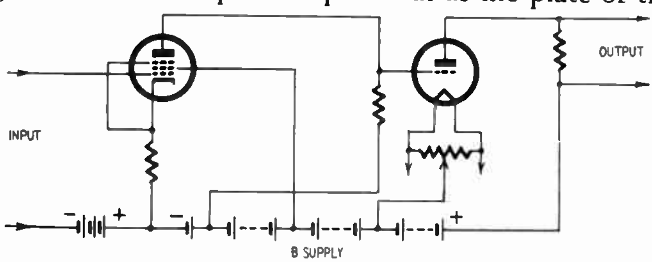


Fig. 309. *Circuit of a direct-coupled amplifier. The capacitor between the plate of the first tube and the control grid of the following tube is eliminated.*

ceding stage, its cathode must be even more positive with respect to the grid, and of course the plate also must become much more positive. In the case of several stages of this continual stepping up of the voltage requirements, a very large supply is necessary, providing dangerously high potentials. Another serious disadvantage of this system for audio use is its lack of stability. Performance varies rather widely with fairly small variations in supply voltages, aging of tubes or tube replacements. For these reasons it is seldom used in audio, although it is quite useful in such d.c. applications as photocell devices and control circuits. For audio, a similar system known as the resistance-coupled amplifier is the most commonly used.

Resistance coupling

This method is more properly known as resistance–capacitance coupling for the capacitor, inserted between the plate load circuit of one stage and the grid input of the next stage is a very essential part of the system. Its purpose is to allow all audio frequencies to pass between stages without impairment, while at the same time providing complete d.c. isolation between output and input circuits of the coupled stages.

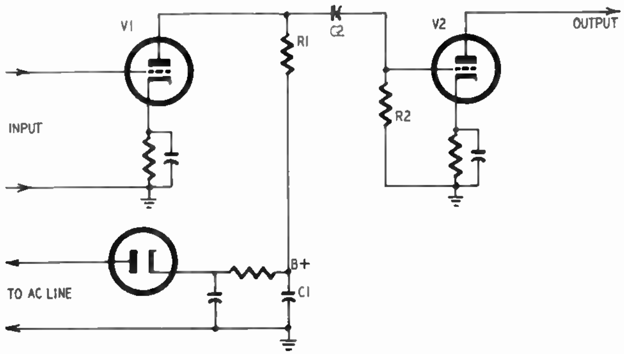


Fig. 310. *Elementary audio-amplifier system. Triodes are used here for the sake of simplicity.*

The theory of operation of this circuit may be understood by reference to Fig. 310. The signal to be amplified is applied across the grid input circuit of the first tube, V1, resulting in a variation in the amount of current normally flowing through the plate load resistor, R1. The audio currents flowing through this resistor cause an audio voltage to be set up across it, this voltage being an amplified replica of the input voltage.

The voltage applied to the plate of the tube, therefore, varies in exact accordance with the audio voltage, for the plate voltage at any instant will differ from the supply voltage by the amount of the drop across the plate load. This audio-frequency voltage will then tend to charge the plates of the coupling capacitor, C2 with the result that the voltage on the grid of the next tube will vary similarly. A grid leak resistor (R2) is included in the circuit to prevent grid blocking by providing the customary return from grid to cathode.

The resistance value of the plate load resistor should theoretically be as high as possible for the gain per stage is determined in part by the ratio of the plate load to the internal plate resistance of the tube. But as the load resistance is increased the supply voltage must be similarly increased. In practice the plate load is usually around three times the internal plate resistance, a value which allows the tube to deliver about 75% of its theoretical voltage amplification.

The plate current of V1 (in Fig. 310) varied by the input signal also flows through R1. It is this variation in voltage across R1 that is our audio signal and it is this signal that is impressed on the control grid of the following tube. Actually, the coupling capacitor C2, and the grid leak resistor, R2, form an a.c. voltage divider network as shown in Fig. 311. Note that in this illustration we have the load resistor, R1, in series with the filter capacitor, C1, of the power supply. As far as the signal is concerned, C1 looks like a very small impedance compared to R1. Usually C1 is a filter-type unit of some 40 μf or more. Its capacitive reactance (in ohms) at the frequency of the audio signal, is quite small. Let us assume that at this particular moment the audio signal has

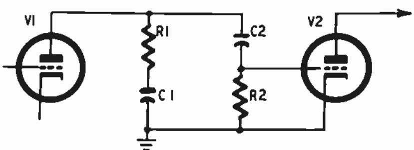


Fig. 311. *This is the same circuit as that shown in Fig. 310. The power supply has been omitted.*

a frequency of 1,000 c.p.s. and that C1 has a value of 40 μf. Under these operating conditions C1 has a reactance of about 4 ohms. If R1 is 470,000 ohms (a common value) then almost the entire audio signal will appear across R1. The value of C1 (in ohms) is so very small that the bottom end of R1 thinks it is tied right to ground.

Now let us take a look at C2 and R2. If you will compare Fig. 311 with Fig. 310 you will see quite readily that all we have done is turned C2 around a bit but that the circuit has not been altered otherwise. C2 and R2 are in series. This series combination is shunted right across R1 just as shown in Fig. 312. We have omitted C1 in this illustration since its reactance value in ohms was small enough to be disregarded. Any audio voltage appearing across R1 will now also be impressed across C2 and R2. The question now arises as to how much of this audio voltage will be dropped by C2 and how much will appear across R2.

There is no question but that if the capacitive reactance of C2 (in ohms) should happen to be exactly equal to the value of R2 (also in ohms) the signal voltage will divide equally between these two components. Thus, if the audio voltage across R1 were 10 volts, then under these conditions we would have 5 volts of signal across C2 and 5 volts across R2. By looking at Fig. 312 we can easily see that the only signal voltage that can do us any good is that appearing across R2. This is the component that is connected between control grid and cathode of the following tube; hence it is R2 that is responsible for delivery of the signal. The ideal situation is to make the capacitance of C2 as large as possible. The larger the capacitance, the smaller the reactance, and the less signal we lose across C2. Similarly, we want to try to make R2 just as large as possible (high value of resistance) so that we get the maximum signal across it.

The correct values of coupling capacitor and grid leak resistor are very important to the frequency response and stability of the amplifier. The capacitor must be quite large to permit the passage of the lowest audio frequencies, and the grid leak must be large enough to provide a sizable voltage across the input of its

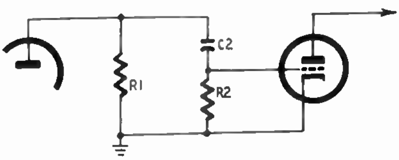


Fig. 312. *Here we have a further simplification of the circuit of Fig. 310.*

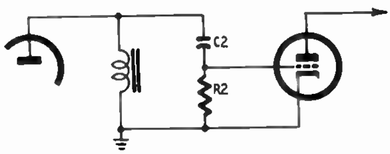


Fig. 313. *An iron-core coil can be substituted for resistor R1.*

stage. But as this combination becomes substantial its time constant is a serious factor in causing blocking of the grid because of the inability of grid-leak return circuit to discharge rapidly enough to follow the signal at high frequencies with high levels. In practice, this time constant is limited to a maximum of .05 second. The coupling capacitor is kept as small as possible consistent with the requirement for good low-frequency response. A fairly common set of values employs a coupling capacitor of .01 μf (microfarad) with a grid leak of 470,000 ohms.

The impedance-coupled amplifier

Impedance coupling is very similar to resistance coupling, the essential difference being that an inductor (or coil) is employed as the plate load in place of the resistor. See Fig. 313. This impedance device consists of a core of laminated silicon steel around which is wound a large number of turns of fine wire. The coupling capacitor and grid leak resistor are the same as in a resistance-coupled stage.

The basic principle of operation is also similar, except that the audio voltage developed across the load is equal to I $\times$ Z (current times impedance), instead of I $\times$ R (current times resistance). Now if the impedance is so designed that it has a high a.c. impedance but a low d.c. resistance, then it should be possible to develop a sizable a.c. voltage drop across the load with a minimum d.c. loss. This will enable the use of a lower plate supply voltage for the same amount of potential actually appearing on the plate and the same amount of signal developed at the output.

The very fact that this impedance is an a.c. device means that it is reactive, with its opposition to current flow being directly proportional to the frequency of that current. Hence the voltage

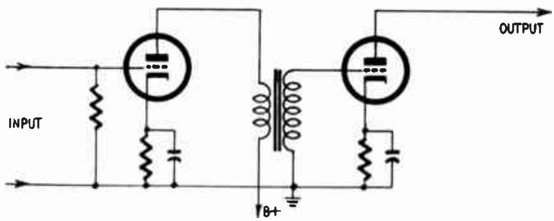


Fig. 314. *Transformer-coupled audio amplifier.*

developed across the load will depend largely upon its frequency, a condition quite the opposite from the ideal. For this reason, along with the relatively high cost of the inductor as compared with a resistor, the impedance-coupled amplifier is hardly ever used for quality audio work.

The transformer-coupled amplifier

A method of interstage coupling which has some advantages involves the use of audio transformers between plate and grid circuits. The schematic of such a method is shown in Fig. 314. The transformer usually has a step-up ratio, due to a larger number of turns on the secondary than on the primary, with a consequent voltage gain in addition to that provided by the tube. The transformer thus acts as an impedance-matching device between the plate and grid circuits, and it may also provide the d.c. path for the plate current.

Any tube will operate efficiently as an amplifier when the load connected across its output is high compared to the internal plate resistance of that tube. In fact, the maximum theoretical gain from the tube is possible only when the plate load is infinite. However, under these conditions the tube would become inoperative, and it is therefore necessary for us to find the best compromise point short of this. An idea of how this works in practice can be gathered by reference to Fig. 315, which shows how the gain of a typical triode, having an internal load impedance of 10,000 ohms and an amplification factor of 10, will vary with a changing load or impedance. We can see that for low values of load impedance only a small fraction of this amplification factor can be realized, but at values of three to five times the plate impedance over 75% of the theoretical gain is possible. Beyond this the law of diminishing returns sets in, and there is little practical value in attempting to raise the gain further.

Getting back to our transformer-coupled amplifier, we now en-

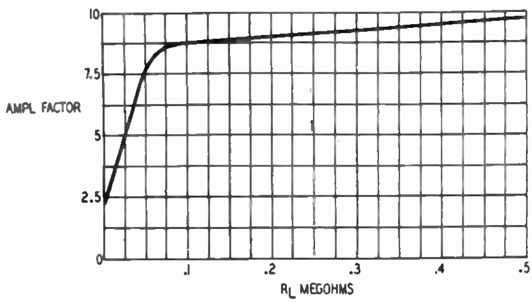


Fig. 315. ***The gain of an audio circuit increases as the resistance or impedance of the load is raised.***

counter the practical problem of building a transformer whose primary has sufficient self-impedance to load the tube properly. A very high inductance involving a tremendous number of turns of wire is necessary, and then several times that number will be necessary on the secondary if a voltage stepup is to be provided. For this reason transformer coupling is most practical when used with triode tubes, the pentode simply having too much internal impedance to be properly loaded by a practical transformer.

Another difficulty with transformer coupling is identical with that encountered in the impedance-coupled method. The transformer is a reactive device whose impedance increases with frequency, which makes for considerable difficulty in loading the tube correctly at the bass frequencies. At the other end of the spectrum the distributed capacitance between the turns of the windings may begin to act as a shunt path for the high frequencies. The result of all this is that it is extremely difficult to achieve a wide-range frequency response with the transformer-coupled method.

The current flowing through the primary winding of the transformer actually consists of two components, a.c. and d.c., with the alternating current component superimposed upon the d.c. This d.c. component is of no use whatever in inducing a voltage over into the secondary, but it does have the undesirable effect of magnetically saturating the core. This results in an attenuation of the positive peaks of the signal with consequent waveform distortion. It also reduces the inductance and reactance of the primary winding, resulting in even poorer low-frequency response.

The problem of saturation is attacked in two ways. First the transformer is built with a large amount of steel in the core, which raises the saturation point but also increases the size, weight and cost. The amount of saturation can be further reduced by

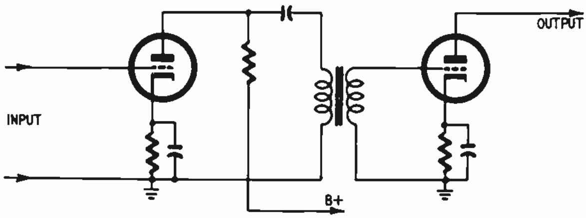


Fig. 316. *The capacitor connected between the plate of the first tube and the primary of the transformer permits passage of the signal, but blocks d.c. from the audio transformer.*

keeping the d.c. component completely out of the winding. This is accomplished by providing a different path for the plate supply in a method known as *shunt feed,* illustrated in Fig. 316. A capacitor is inserted in series with the primary to block the d.c. from flowing through the winding, but the capacitor is large enough to pass all audio frequencies. The d.c. meanwhile flows through the dropping resistor to the power supply. Since a part of the a.c. component will also appear across the dropping resistor, this system provides somewhat lower gain but it does at least eliminate d.c. saturation of the transformer core.

Motorboating

Whenever two circuits operating at the same frequency have an impedance which is common to both, there is coupling between them. In audio amplifiers having a common power supply for all stages, coupling will exist through the high internal resistance of the power supply itself. Then, depending upon the phase relationships of the voltages present, there may be either regeneration or degeneration. In the former case instability may result. The amplifier becomes a form of relaxation oscillator at a low frequency, emitting a sort of *putt-putt* sound often called *motorboating.*

Since there is ordinarily a phase reversal of 180° in each stage, the coupling between adjacent stages through the common power supply is degenerative and of little consequence. But when more than two stages are employed, alternate stages will have in-phase signals and regenerative coupling may then result.

The commonest method of avoiding motorboating involves the use of a decoupling filter consisting of a resistance and a capacitance. Such a circuit is shown in Fig. 317. This is shown as part of a resistance-coupled amplifier, but the method is applicable to any type of coupling. Circuit constants are so chosen that the reactance of the bypass capacitor to all audio frequencies is con-

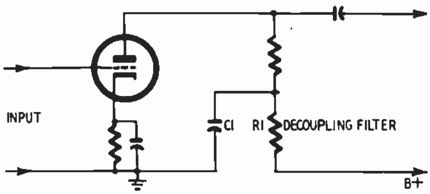


Fig. 317. *A decoupling filter consists of a resistor and a capacitor.*

siderably less than the resistance of the decoupling resistor plus the internal resistance of the plate supply. In practice a decoupling resistor around one-fifth the size of the load resistor is used in conjunction with a fairly large capacitor, perhaps on the order of 8 μf. This circuit also has a second desirable effect in that it aids considerably in hum reduction.

Like almost everything else in electronics, if you want a desirable effect you must pay for it. The decoupling network, C1 and R1 in Fig. 317, will minimize or eliminate motorboating, reduce hum, improve filtering—but at a price. A part of the B plus voltage will drop across R1. This means that the available B plus supply voltage is reduced by just that amount. As an example, if the supply voltage is 300 and you get a 50-volt drop across R1, then your supply voltage is actually only 250 volts.

This disadvantage, however, is minor compared to the benefits, and most audio systems therefore use some decoupling.

Push-pull amplification

The simplest type of push-pull amplifier employs transformer coupling, as shown in Fig. 318, in which two tubes act to provide a single stage of amplification. When an input signal appears across the primary of the transformer, there is first a voltage stepup by induction from primary to secondary and then the signal appears at the grids of the two push-pull tubes. Since there is a difference of potential between the ends of the transformer secondary, the voltages on the grids must be 180° out of phase with each other. Thus when a signal is applied, one grid will become more negative at the same time that the other one becomes less negative, and *vice versa.*

Similarly, in the plate circuits of each tube, as the plate current in one tube becomes greater due to a *less negative* voltage on its control grid, the plate current in the other tube at the same instant will be decreasing by a like amount due to a *more negative* voltage

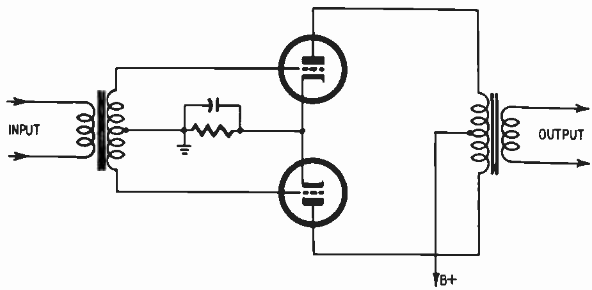


Fig. 318. *Fundamental diagram of a push-pull circuit. Center-tapped transformers are used.*

on its grid. These actions result in voltages which are out of phase with each other at opposite ends of the output transformer. This is precisely the desired condition. The voltages will then combine additively in the output transformer and the resultant will be exactly twice what it would have been with a single tube operating under identical voltage and signal conditions. Since it doesn't appear that we have something for nothing in the combination of two tubes in push-pull compared with two single-ended stages in cascade, we must look elsewhere for the advantages of this circuit.

Referring again to Fig. 303, we observed the condition in which the negative pulse of the signal was nearly eliminated due to an excessive amount of grid bias on the tube. And even if the bias had not been so very great, the negative pulse would have been deformed as long as the operating point was allowed to remain in the vicinity of the lower bend of the curve. Similarly, if the bias were very small (less negative) then the operating point would be along the upper bend of the characteristic curve, resulting in waveform distortion of the positive half of the cycle. In the third instance, even if the operating point were correctly set at the middle of the straight-line portion of the characteristic curve, a very heavy signal would swing the grid so far that the tube would operate into *both* bent portions of the curve and cause both negative and positive peaks to be flattened. Finally if the load impedance connected to the tube is rather low, about equal to the plate resistance instead of the customary two to five times that amount, the dynamic characteristic curve of the tube much more resembles the letter *S*, with almost no linear portion at all. In each of these four cases, then, waveform distortion occurs within the tube itself: (1) excessive grid bias; (2) insufficient bias; (3) excessive signal and (4) low load impedance.

In all these instances the waveform can be analyzed to show that the output signal has had added to it a number of harmonics which were not present in the original, resulting in *harmonic distortion*. Furthermore, the sort of distortion condition shown in Fig. 303 is largely the result of the addition of the second harmonic to the signal fundamental, as we see in Fig. 319-a, -b.

A further study of Fig. 319 shows that the fundamentals and resultants are 180° out of phase, as they should be for additive combination at the output of a push-pull circuit. But when this is true we also note that the spurious second harmonic components are still exactly in phase. It would then seem to be a safe assumption that if out-of-phase voltages can combine additively in a push-pull circuit, equal in-phase voltages will cancel each other. This is exactly the case, and we can now extend it further to state that *all even-order harmonic distortion generated within a push-pull circuit will be cancelled and eliminated at the output*. This is unfortunately not true of odd harmonics nor is it true of distortion present at the input signal and fed into the system in correct phase.

From Fig. 318 we can see that the plate voltage is applied to the two tubes by means of a center tap at the primary of the output transformer. Therefore the d.c. component as well as any extraneous voltages which accompany it will arrive at both plates simultaneously, which is to say they will be in phase. This fact provides us with several additional advantages in the push-pull system. Since the two d.c. plate currents in each half of the winding flow in opposite directions, the magnetization of the core is very slight, which means almost no danger of magnetic saturation. Thus, the turns of wire on the transformer are able to provide more effective inductance, with the result that our transformer needs proportionately less iron and less copper than for the equivalent performance in a single-ended stage.

Also as a result of this center-tap feed, any hum voltage applied to the plates from the power supply will be cancelled. Similarly any other fluctuations in the plate voltage, such as those due to interstage coupling within the power supply, will also be eliminated. Thus the possibility of instability or motorboating due to coupling is practically nonexistent in the push-pull amplifier.

Phase inversion or splitting

From our discussion of the push-pull system it is obvious that some method of phase splitting is necessary to provide two voltages which are 180° out of phase before they are applied to the

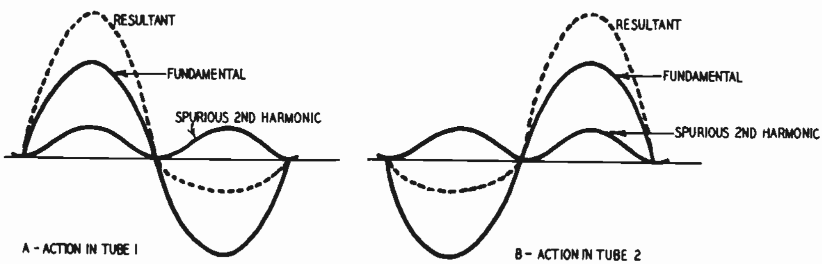


Fig. 319-a-b. *Waveforms in the output of a push-pull circuit.*

two control grids. The most obvious method of accomplishing this is by the use of the center-tapped input transformer already discussed. Other methods involve a modification of impedance coupling in which a center-tapped choke coil is connected to the input or a pair of equal resistors in series may be connected across an ordinary transformer with a connection to the center of them providing the ground return for the cathodes.

The commonest method of phase inversion, however, employs a modification of resistance coupling. This eliminates the expensive transformer and provides better fidelity at the same time. A typical resistance-coupled phase inverter is shown in Fig. 320.

To understand its principle of operation consider an instant when the alternating input signal is becoming positive, thereby decreasing the negative voltage on the grid of the phase inverter. At this point the plate current will increase, causing a larger voltage drop across resistors R1, R2 and R3. Then the voltage appearing at point A will have decreased (will have become more negative) while the voltage at point B will have increased (will have become more positive). Thus these two voltages are out of phase and will be suitable for exciting a push-pull amplifier, provided they are of equal value. Since the same current flows through R2 and R3, this condition is met by making these resistors equal.

This particular phase inverter circuit has a few minor disadvantages, one of them being that it cannot be used with a device having an unbalanced output. If a pickup, for example, having one side grounded, were to be connected across this phase inverter, it would short R2 to ground and thus kill one side of the split circuit. Also it does very little amplifying, as degeneration is characteristic of its operation. In practice the gain is about 0.9 for each side of the split, or a total of 1.8 times from grid to grid of the push-pull output tubes. Therefore the phase inverter is

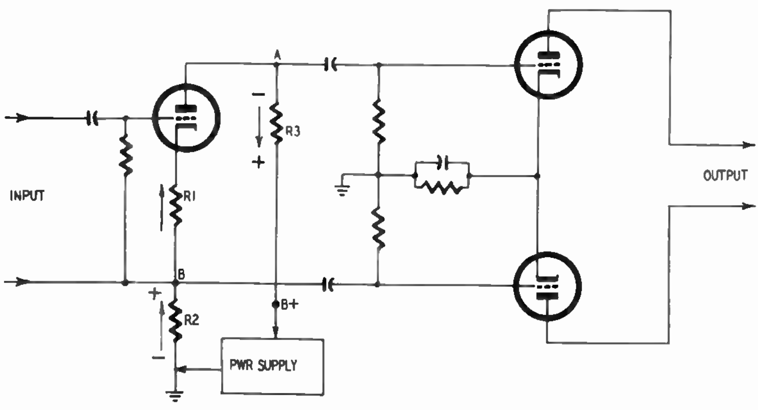


Fig. 320. *A single tube can be used to supply out-of-phase signals for driving tubes in push-pull.*

best regarded simply as a substitute for the center-tapped input transformer. This is not nearly as extravagant as it may seem: the transformer will usually cost considerably more than the tube and its associated circuits and still the results would not be as satisfactory. Furthermore, the degeneration is actually rather desirable, for it possesses all of the usual features of negative feedback (inverse feedback). These advantages include the reduction of harmonic distortion and an improved frequency response.

Inverse feedback

Whenever a part of the signal appearing at the output of a circuit is returned to its input, we have a condition of *feedback.* If the feedback signal is *in phase* with the incoming voltage and thus adds to it, the feedback is said to be *regenerative* or *positive.* If the feedback is out of phase with the incoming signal, it is called *degenerative, inverse* or *negative.* Either of these effects may occur accidentally within an amplifier due to faulty design, but they may also be deliberately introduced to serve some useful purpose.

Regeneration is common in radio transmitters, communication receivers and audio oscillators while *both* positive and negative feedback are sometimes employed in quality amplifiers. Most amplifiers use some form of negative feedback. The system of negative feedback essentially involves an arrangement whereby energy from the plate output circuit is fed back to the grid input with such a phase relationship as to have the feedback voltage opposing the input signal. This circuit is shown in Fig. 321.

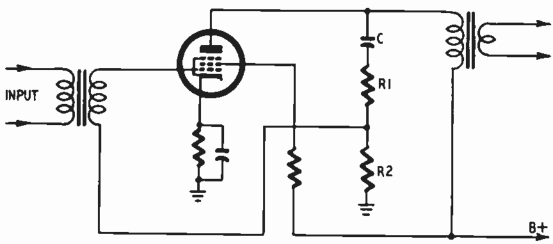


Fig. 321. *Typical negative-feedback arrangement. Frequency response improves at expense of gain.*

While this circuit is relatively simple, the amount of feedback and its phase with respect to the signal must be carefully engineered. Otherwise, serious instability and distortion may occur due to uncontrolled regeneration within the audio spectrum or even somewhere outside it. The arrangement of Fig. 321, in which the feedback occurs all in one stage, is the least complicated, but often in high-quality applications the feedback signal may be returned to a circuit several stages preceding the point of feedback pickup.

Referring to Fig. 321, the feedback circuit consists of the voltage divider R1 and R2, plus the blocking capacitor C, which prevents the d.c. plate supply from being applied to the control grid of the tube. But the a.c. component in the plate circuit readily passes through the capacitor and appears across the combination of R1 and R2. By varying the values of these resistors we can feed back any desired percentage of the output signal. The actual amount of voltage fed back will be proportional to the ratio R1/(R1 + R2). For example, if the total audio voltage across the output without feedback is 100, R1 = 10,000 ohms and R2 = 190,000 ohms, then the feedback factor will equal 10,000/(10,000 + 190,000) = 0.05. This is a voltage ratio of 20 to 1, or about 26 db, a fairly common value. In practice the amount of feedback is usually somewhere between 16 and 40 db.

We have already indicated that inverse feedback causes a reduction in harmonic distortion. The manner in which this is achieved is illustrated in Fig. 322. The waveform of the output without feedback is considerably distorted due to the addition of spurious harmonics. When feedback voltage is applied to the input, it will be similarly distorted and this distortion will oppose that generated in the amplifier, resulting in some cancellation of the distortion at the output. Of course the output is reduced in the process, but the important thing is that the distortion is reduced by exactly the same proportion. That is, if the signal is reduced by just 6 db (voltage ratio of 2 to 1), the distortion will also be cut in half. It is then a simple matter to provide sufficient gain to overcome the loss in the feedback circuit, and the final signal will still have considerably less distortion than it would have had without feedback. In similar fashion all other spurious components generated within the amplifier, such as noise or hum, will be reduced in like amounts.

We have also said that inverse feedback improves the frequency response of an amplifier. The reason for this will become clear when we refine our previous statements concerning the phase relationships between the input and output voltages. The tube itself is said to be phase-reversing because its plate current varies in a manner exactly the opposite to the voltage on the grid. But there are also other factors at work here in the shape of capacitors and inductances in the circuit as well as stray values of both inductance and capacitance, even including the internal structure of the tube itself. These reactive components will take an active part in affecting the phase relationships of the voltages across them and the currents flowing through them. As a result of these several forces working simultaneously, the actual phase relationship of the feedback voltage with respect to the input may be something quite different from an exact 180°.

Now when the phase angle is anywhere from 90° around through 180° and up to 270°, the input and the feedback will combine more or less negatively and the net result will be degeneration. But if the phase angle is upward of 270° around to 90°, the two signals will add and produce regeneration. The only exceptions are at precisely 90° and 270°, where the feedback will have no effect. We also know that the reactances of the circuit capacitance and inductance will vary with frequency, the inductive reactance increasing at the higher frequencies and the capacitive reactance increasing at the lower end. Therefore the phase angle of the feedback signal will swing closer to 90° or 270° from the 180° condition, and less degeneration will result. Thus the losses at the extreme ends of the range will be less than at the middle, resulting in a stronger bass and high-frequency response. If the phase angle swings too far, however, then real trouble will be encountered in the form of regeneration. This may be an audible squeal or it may be an oscillation outside the audio range which will seriously overload the amplifier. In the design of an inverse feedback system, you must be careful in your choice of values for reactive components.

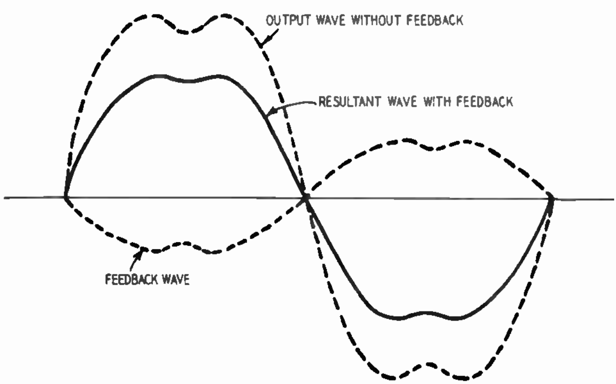


Fig. 322. *The effect of feedback is to reduce the gain and improve the waveform*

Volume compression, limiting and expansion

Volume compression and limiting are necessary because of inherent shortcomings in even the finest available audio transmission equipment. The equipment, in short, is simply not capable of handling satisfactorily the entire useful dynamic range of the human ear.

Since the full volume range of a symphony orchestra, for example, may be as much as 80 db, while the peak signal-to-noise ratio of even the best equipment may be no better than 60 or 70 db, it is obvious that sounds of least intensity will tend to disappear below the innate noise level of the system. If they are to be heard at all, therefore, they must be amplified to a greater degree than the louder passages.

At the other end of the scale, every system has a maximum point beyond which the signal level through it cannot be increased without encountering excessive distortion or other undesirable effects. A lateral disc recorder, for example, will undergo such wide excursions that the adjacent grooves it cuts will run into one another, so that the playback will simply be unable to track and will *jump* and *skate* grooves. Overmodulation of radio transmitters, both AM and FM, will result in severe distortion, interference to signals on adjacent channels and possible damage to equipment. Magnetic and optical sound recorders will similarly overload and distort, optical systems being particularly sensitive.

All of this means that it is necessary to compress the range of incoming sound volume to an area where the softest passages are kept up out of the noise region while the loudest peaks are prevented from spilling over into the overload zone. Much of this can be accomplished by a competent control engineer who attempts to anticipate rapid level changes which will be beyond the useful range of his system and who "rides gain" on the signal by adjusting levels as he deems necessary. But unless he has had a number of rehearsals it is highly unlikely that he will be able to do a perfect job.

One of the handicaps to manual volume compression lies in the fact that the American standard VU (volume unit) meter does not indicate excessive peaks of very short duration, the instrument being an average-reading device. Furthermore, the human reflexes are such that it is extremely difficult to detect and prevent trouble before the damage has been done. For these reasons, equipment has been developed which will automatically control the gain of an amplifier by electronic means. Basically, this is simply a refinement of the automatic volume control commonly employed in radio and TV receivers. Such equipment is known as a compression or limiting amplifier, the two terms having the same general meaning with the difference between them being largely one of degree.

The basic operation of a typical volume compressor can be understood through reference to Fig. 323. This is simply a conventional audio amplifier to which has been added a control tube and circuit which works between the output and input. The push-pull tetrodes are of the variable-μ type, which means that their mutual conductance and amplifier gain as determined by the bias appearing on their grids are variable over a much wider range than usual. The bias voltage is obtained, in this case, across a resistor–capacitor combination (R-C) between the control grids and ground, but the grids and this R-C circuit also connect to the plates of a dual-diode control rectifier. The cathodes of this rectifier are biased positively by an amount determined by the bias control. They also receive a sample of the audio passing through the system as delivered by the control amplifier bridged across the output. The magnitude of the audio signal applied to the diode is determined by the setting of the gain control.

When a signal passes through the amplifier, a portion of it

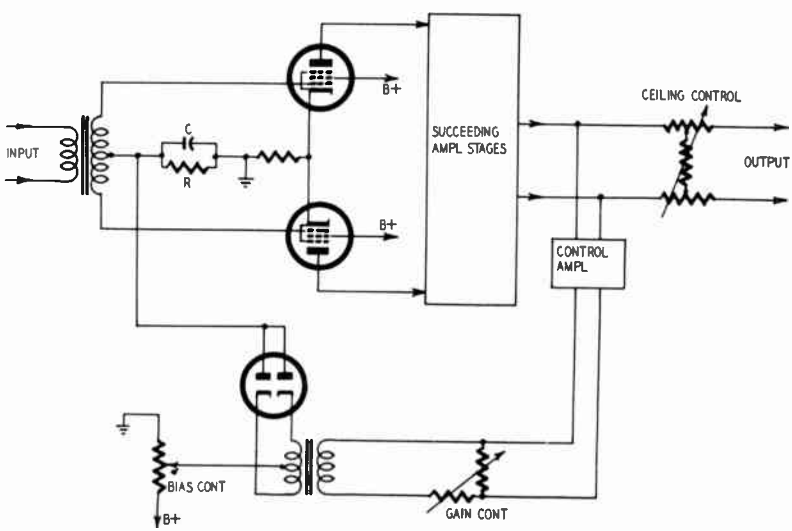


Fig. 323. *In this circuit we have a volume compressor added to a typical push-pull unit.*

will appear across the transformer at the cathodes of the rectifier. At medium signal levels there will be no effect on the diode, but when heavy peaks occur, the signal delivered to the cathodes will exceed the control tube bias, one of the cathodes will become negative with respect to ground, current will flow through the diode and a negative voltage will appear at the grid end of the R-C filter circuit. This will act to bias the tubes in the variable-μ stage more negatively, with a consequent reduction in overall gain. We now have an amplifier whose input–output characteristic appears as in Fig. 324. The system behaves quite conventionally up to the point (known as the *breakaway point)* where the diode bias is exceeded and compression begins. Beyond this point the characteristic breaks away from the 45° slope (1/1 ratio), with a given increase in input level being altered to a smaller increase in output level. The reciprocals of the slope lines of these various curves are known as *compression ratio,* and two of the commonest of these ratios are shown in the diagram. For the gradual compression characteristic, the expression "2/1 ratio" means that beyond the breakaway point there must be 2-db increase in input for every 1-db increase in output. The limiting slope is much more severe, requiring 4-db increase at the input for each 1 db more at the output.

The important characteristics of a compression amplifier are: attack time; release time; breakaway point; compression ratio.

Consider what happens when an excessive peak is applied suddenly to the input of the amplifier. The system will be operating under its normal gain condition along the line AB of Fig. 324. When the peak hits, a measurable period of time will pass before the diode operates and the grid capacitor (C in Fig. 323) of the controlled stage becomes charged. This period, known as the *operate* or *attack time,* is a function of the value of capacitance in the R-C filter. It is desirable to have as short an attack time as possible; in commercial equipment attack times vary downward from 10 milliseconds, with values well under 1 millisecond most common.

While a fast reaction time is most desirable in a compressor, it must not act so quickly that it is able to follow every single alternation of the lower audio frequencies. One obvious method of avoiding this is simply to eliminate the most susceptible frequencies and this is actually sometimes done by means of a high-pass filter with a cutoff of 45 c.p.s. or so. Also the *release time* is increased to something between 25 milliseconds and several seconds, thus avoiding audio-frequency fluctuation of the grid control bias.

A compressor may be responsible for a modification in the balance between frequencies, a condition known as *spectral energy distortion*. This is most noticeable to the ear at the upper middle frequencies, where the hearing is most sensitive, because the relatively low powers at these frequencies are incapable of actuating the compressor as readily as are the more powerful low frequencies. This condition is corrected by the insertion in the control circuit of an equalizer which boosts these upper-middle frequencies and make them more effective in operating the compressor. This usually takes the form of a 3–8-db accentuation at 6,000 c.p.s. Since the compressor without equalization causes dialogue to be excessively sibilant, this circuit is often referred to as a *de-esser*. The design of a compressor is predicated on a reasonably well-balanced signal at its input and for this reason it is not advisable to alter this characteristic by equalization or filtering until after the signal has undergone compression.

A variable attenuator or *ceiling control* is connected at the compressor output to control the compression ratio. Increasing the loss in this ceiling control will require an increased signal at the input to the compressor for the same peak output level. This therefore effectively increases the overall amount of compression applied to the signal as it passes through the system. For maximum flexibility and ease of operation, the ceiling control should be within easy reach of the control engineer at the mixing console.

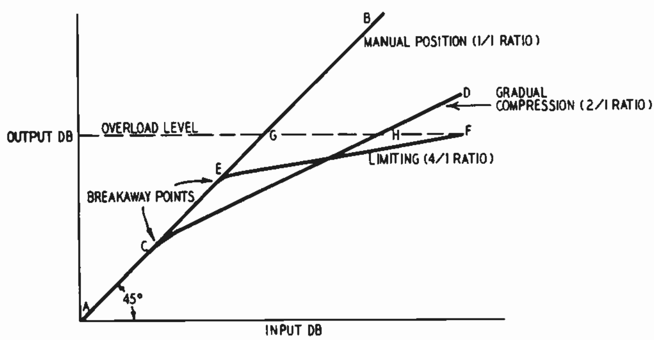


Fig. 324. *Characteristics of an amplifier using volume compression.*

The other two controls of Fig. 323 affect the compression ratio and the breakaway point. The gain control, which determines the amount of signal applied to the cathodes of the control tube, will affect the breakaway point only. If the setting of this attenuator is reduced by 6 db, for example, then a control signal 6 db higher will be applied to the rectifier and the breakaway point will occur 6 db sooner. That is, compression will begin at a peak signal level 6 db lower than the previous breakaway point.

Adjusting the bias control for a higher fixed voltage on the rectifier cathodes, on the other hand, will result in both a higher compression ratio and a higher breakaway point. Thus when adjusting a compressor to a desired characteristic, it is necessary to make several successive adjustments, alternating between the gain control and the bias control.

It is obvious now that the actual operating characteristic of a compressor or limiter is a function of both the compression ratio and the breakaway point. At least a couple of methods are employed for describing the shape of this characteristic. Referring again to Fig. 324, a compression ratio of 2 to 1 tells us that beyond the breakaway point, a 2-db increase in input level is necessary to cause a 1-db increase at the output. Then suppose that the *overload level* line intersects the *manual position* line at a point G; 8 db above breakaway point C. Then the *gradual compression* line will intersect the overload line at an input level 16 db higher, due to the 2 to 1 ratio. Then it is said that for this given overload level and this compressor adjustment, the upper 16 db have been compressed into 8 db. If the overload level in this case were +12 dbm, we would then describe our characteristic by stating that "16 db are compressed into 8 db at an output level of +12 dbm." The compression characteristic is then simply stated in terms of the difference in db between the overload level without compression and the output level with compression for the same value of input signal.

Volume limiting is somewhat different from compression and it employs a much more severe compression ratio, as seen in Fig. 324. It is intended primarily to set a maximum on the output level, above which it is almost impossible to go. Thus it simply clips off the signal peaks which exceed a predetermined value, thereby preventing overmodulation of recorders and transmitters.

The ratio of 4 to 1 is quite common although ratios as high as 10 to 1 have been used. Thus the breakaway point can be set at about the overload level, resulting in a limiter working range which is much narrower than that of the compressor. While the compressor might be said to be working almost constantly, the limiter is actuated only by overload peaks. Distortion and loss of speech intelligibility result when limiters have as fast release times as are commonly used with compressors.

Most units of this type can be used as either limiters or compressors. Throwing a single switch will ordinarily change the compression ratio and the release and attack times as well. Compression amplifiers are usually designed as no-gain devices, containing only sufficient booster amplification to overcome the circuit losses. This arrangement makes it possible to switch the device in and out of an audio channel at will without seriously upsetting level adjustments.

Volume expansion performs precisely the opposite function of compression and limiting. The incoming signal, instead of tending to cause a reduction in its own level through the control circuits, now produces a change in output proportionately greater than the change in input which caused it. Under ideal conditions the expander characteristic should be the exact complement of the compression, in which case the original dynamic balance would be fully restored. But as a purely practical matter, it is hardly ever possible to know what characteristics were employed at the transmission end and thereby to establish the correct complementary characteristic required for reproduction. Under these circumstances the results are often poorer than without expansion. For these reasons volume expansion is seldom useful in reproduction and has little popularity.

Generator concept of a vacuum-tube amplifier

A vacuum-tube amplifier in operation is in many ways analogous to an a.c. generator and its associated load circuit, and a study of

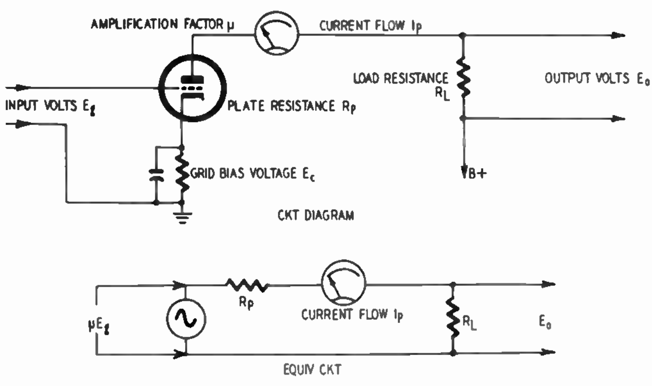


Fig. 325. *The vacuum-tube amplifier can be simplified to an equivalent circuit.*

it from this viewpoint provides much useful background information when you attempt the design of your own amplifier circuits. This concept, sometimes referred to as the μE_g approach to amplifier operation, is based on the notion that the tube itself is an a.c. generator developing a voltage equal to the product of the input signal E_g times the voltage amplification factor μ. The generator has an internal resistance equal to the plate resistance R_p of the tube and works into a load represented by the load impedance R_L. Fig. 325 is a schematic illustration of this basic idea.

Now as a practical example of the usefulness of this concept, let us consider a commercially available tube, such as the 6C5, and examine its operation. From a tube manual we learn that a typical recommended condition would yield the following constants: $R_p = 10{,}000$ ohms, $E_c = 8$ volts bias, $\mu = 20$.

With a negative 8-volt bias, we can assume that no serious overloading will occur with a signal (E_g) of 5 volts, so we'll arbitrarily establish that value as one of our operating conditions. Finally we'll start with an assumed value for R_L equal to R_p, or 10,000 ohms.

As a first step in our analysis we must establish the value of the voltage which our hypothetical generator is developing:

We'd like to know just how much of this internally developed voltage is actually useful to us at the load. To learn this we begin by calculating the amount of current flowing in the circuit:

$$I_p = \frac{\mu E_g}{R_p + R_L} = \frac{100}{10{,}000 + 10{,}000} = .005 \text{ ampere}$$

Now we can determine the value of the output voltage E_o across the load:

$$E_o = I_p \times R_L = .005 \times 10{,}000 = 50 \text{ volts}$$

Here it is evident that when the load impedance equals the plate resistance, exactly half of the generator voltage is delivered to the output, the other half being lost within the generator itself. It is evident that increasing the load will increase the voltage across it, but it will at the same time reduce the current which produces that voltage drop. Calculating E_o for a number of values of R_L and plotting the results will produce a curve having the identical shape of that of Fig. 315, thus confirming our earlier statement that the load must be large with respect to the plate resistance to realize much of the theoretical amplification factor but that in practice it is not worth the trouble and expense to attempt to get more than about 75% of this figure.

Now let's see what happens to the power developed in the output circuit.

$$P_o = \frac{E_o^2}{R_L} = \frac{2{,}500}{10{,}000} = 0.25 \text{ watt}$$

It is obvious that if E_o were to remain constant and R_L decreased, then the power output P_o would increase. But there is an interdependence between I_p, E_o and R_L, as I_p decreases and E_o increases with an increase in R_L: If a number of other values were assumed for R_L and the preceding series of calculations worked out successively for a set of values for P_o, the result would be a decrease in power output for P_o for all conditions other than $R_L = R_p$. From this we can infer the rule that *for any given input signal voltage the power output will be maximum when the load impedance R_L equals the internal plate resistance R_p.*

This would then seem to be the ideal condition, except for the fact that we have already learned that such a low load impedance bends the tube characteristic curves into an S-shape, with high distortion resulting from nonlinearity. The operating condition for the maximum power output consistent with minimum distortion has been determined experimentally and provides us with the rule: *for any given input signal voltage the undistorted power output will be maximum when the load impedance R_L equals twice the internal plate resistance R_p.* Under these specific conditions the power output can be calculated from the formula:

$$P_o = \frac{\mu^2 E_{gp}^2}{9R_p} \tag{7}$$

wherein E_{gp} is the *peak* value of the signal input, equal to $E_g \times \sqrt{2}$.

As an example of the application of this equation, let us see how much power we must sacrifice in this case when operating under the maximum undistorted conditions:

$$P_o = \frac{20^2 \times (5 \times 1.414)^2}{9 \times 10{,}000} = \frac{400 \times 49.98}{90{,}000} = 0.2221 \text{ watt}$$

As compared with the maximum value of 0.25, it is clear that the maximum undistorted value of 0.2221 represents a decrease of only about 11% (a power ratio of 1.124), a drop in amplitude of 0.5 db. The power loss is insignificant, being unnoticeable to the ear, but the improvement in fidelity is considerable. Even at this point, however, the distortion is around 5%, although further improvement through increasing the load will not be very satisfactory. For this reason it is best to attack whatever remaining distortion it is desired to eliminate by the use of push-pull stages and inverse feedback.

Principles of audio-amplifier design

Before you can begin to design an audio amplifier for any specific purpose, you must consider the requirements of the given application in terms of the job which must be done and the standards which must be maintained. At the beginning of this chapter we enumerated the fundamental design considerations faced by the amplifier engineer. Let us now see how these factors apply to a practical design problem.

Suppose we require a portable remote broadcast amplifier with a single channel working from a crystal microphone into a 600-ohm telephone line requiring a program level of +6 VU. We'll begin by deciding that, since this equipment must be able to operate wherever there is a telephone line available but where there may not be any handy a.c. outlets around, the power shall be furnished by batteries.

We learn from the specifications of the particular microphone we intend using that its output is about 5 millivolts across 500,000 ohms. This is a level of —73 dbm. Since the level at the output must be +6 VU, our overall gain requirement is 79 db. The impedance of this device is such that it operates very well when connected directly into the input circuit of a tube having a grid-leak resistance of about 500,000 ohms and a glance at a tube manual shows us that this is not an unusual value. We therefore anticipate no trouble with impedance matching at this end. The load impedance of 600 ohms, however, is too low for efficient operation of the output tube, and we realize that a matching device, probably a transformer, will be needed between the plate circuit and the line.

We also find that it would be rather difficult to obtain all of the gain we require within a single stage. Bearing in mind that high gain along with even a modicum of power is also difficult in the same stage, we decide tentatively to design a two-stage amplifier, with the first stage providing most of the necessary gain, while the second shall complete the gain requirement with sufficient power to maintain the correct line level.

Since a plate-to-line transformer must be used at the output, this dictates the use of a triode in the second stage. The plate resistance of a pentode is too high to permit a transformer to load it properly. But then in order to insure adequate gain, we decide that a pentode should be used in the input stage. For purposes of simplicity and economy, we would like both tubes to have the same filament voltage, so we will confine our search for appropriate tubes to the 6.3-volt group. Finally, since we know resistance coupling to be simple, inexpensive and capable of excellent quality, we decide to employ this method between stages. This amplifier will be employed primarily for speech amplification, so we will forego such refinements as push-pull and inverse feedback.

We begin our selection of the tube type best suited to our needs for the input stage by consulting manufacturer's literature, such as the *RCA Receiving Tube Manual, RC17*. We start by checking the "Receiving Tube Classification Chart," and by a process of elimination find the box on the chart which applies to our particular problem. The process would go something like this: "Pentodes—sharp cutoff—single unit—6.3 volts." In this box we find about a dozen types listed, and to eliminate further we next consult the "Resistance-Coupled Amplifiers" section with its "Key to Charts." Then we compare our types in the classification chart against this second list and find that many from the first box are not listed here and can therefore be disregarded. Now we look at the main body of the manual and read the descriptions of each of the several types we are still considering. Only one of these, the 6J7, is described as being used in *high-gain audio amplifiers,* and we therefore settle on that type.

Since there is a certain amount of interdependence between the

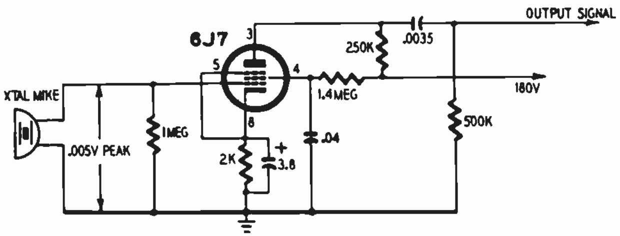


Fig. 326. *Resistance-coupled audio amplifier. The design of this circuit is described in the text.*

values of the various components and operating voltages, our design calculations would be quite tedious but for the fact that much of the work has been done for us and presented in tabular form in the resistance-coupled amplifier charts previously mentioned. Referring to the chart which shows the circuit constants for our 6J7 tube, we find that we have one more fact which must first be established—the plate voltage. We would like to keep our battery pack as light and portable as possible, of course, so we decide to try a supply voltage of 180 volts, a value which will provide a signal voltage gain of 140.

With these factors established, we now find that our first stage is already designed for us, up to and including the grid leak of the second stage. Our circuit up to this point, then, using the values supplied in the tube manual, will look like that in Fig. 326.

Before proceeding with the design, let us determine what is happening to the signal and if the voltages developed will be within the operating tolerance of the tube. With a microphone peak signal of .005 volt and a gain of 140, the voltage developed across the output will be approximately 0.7 volt peak $(.005 \times 140)$.

Since the tube manual tells us that the maximum distortion-free voltage allowable under these conditions is 60, we are working very well within the rated tolerance. Being satisfied as to these conditions, we can now proceed to the design of the final stage.

The first step must be the selection of a tube, and we once again begin by consulting the tube classification chart under the subject headings "Triodes—medium-mu—single unit—6.3 volts." Then the resistance-coupled chart key and the technical data pages are used again to narrow the choice down to a single type, in this case obviously the 6C5.

Now, however, we run into a little difficulty, for our resistance-coupled data cannot be employed with a stage using transformer output. So we consult the technical data page for the 6C5 and find the following typical operation: plate volts, 250; grid volts, —8 (grid-circuit resistance should not exceed 1.0 megohm); amplification factor, 20; plate resistance, 10,000 ohms; transconductance, 2,000 micromhos; plate current, 8 ma.

We still aren't quite ready to proceed, however, for we note that this typical condition is predicated upon a plate voltage of 250, while we had previously decided that our power supply would be only 180 volts. In order to design a stage under these conditions, we must go elsewhere in the manual to the set of *conversion curves,* which are used for calculating approximate operating conditions for a plate voltage which is not included in the published data. These curves are reproduced in Fig. 327.

The first step in converting the given figures is computing the ratio of the new plate voltage to the published data, in this case $180/250 = 0.72$. We assume negligible d.c. drop across the transformer winding, with the full supply voltage consequently being applied to the plate. This figure of 0.72 is known as the *voltage conversion factor* (F_e), which is multiplied by the given value of grid bias to ascertain the new voltage, in this way: $8 \times 0.72 = 5.76$ volts. We can then obtain the balance of the new conditions simply by multiplying the published values by the factors given on the graph for the voltage conversion factor of 0.72. The key to the various curves is as follows:

1. F_i applies to plate and screen currents
2. F_p applies to power output
3. F_r applies to plate and load resistances
4. F_{gm} applies to transconductance

From these we can calculate that the new plate current will equal $0.6 \times 8.0 = 4.8$ ma, and the new plate resistance will be $1.3 \times 10{,}000$ ohms. We can now determine the value of cathode bias resistance by the formula:

$$R_k = E_c / I_p \tag{8}$$

$$R_k = 5.76/.0048 = 1{,}200 \text{ ohms}$$

It is standard practice to limit the peak signal voltage to within 95% of the grid bias, and since in this case the signal delivered from the first stage is only 0.7 volt, the tube is operating well within tolerance.

For the choice of the cathode bypass capacitor the general rule is that the reactance of this unit at the lowest frequency it is desired to amplify should be held to within 10% of the value of the resistance which it is shunting. Since in this case the resistance is rather small, the capacitance will have to be quite large to provide adequate bass response. When the reactance exceeds

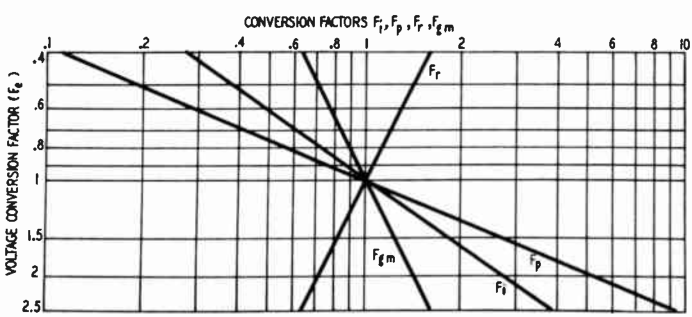


Fig. 327. *Conversion curves are used when designing for particular operating conditions.*

the 10% figure, degeneration begins and bass response falls off rapidly. In this case we select a capacitor of 15 μf, affording good bass response down to around 60 c.p.s., quite adequate for speech.

About all that remains now is the design of the output transformer, which good audio practice dictates should have a primary impedance equal to twice the plate resistance. In this case, then, the transformer should have a primary of 26,000 ohms with a secondary of 600 ohms. To determine just how much signal our amplifier will deliver to the line through this transformer, we must first determine the effective voltage in this stage:

$$VG = \frac{\mu R_L}{R_L + R_p} \tag{9}$$

$$= \frac{20 \times 26{,}000}{39{,}000} = 13.33$$

Then the a.c. voltage appearing across the primary of the output transformer would be:

$$E_L = E_s \times VG \tag{10}$$

$$= 0.7 \times 13.33 = 9.33 \text{ volts}$$

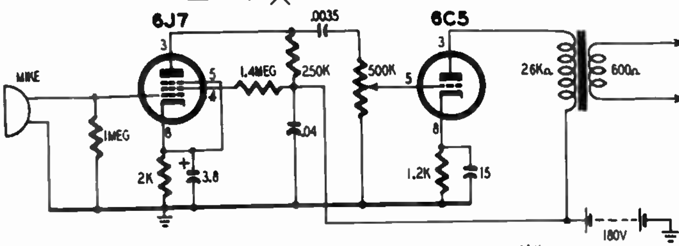


Fig. 328. *The final design of the amplifier.*

Now to determine the voltage across the secondary and fed to the line, it is first necessary to calculate the turns ratio:

$$\frac{N_p}{N_s} = \sqrt{\frac{Z_p}{Z_s}} \tag{11}$$

$$= \sqrt{\frac{26{,}000}{600}} = 6.58{:}1$$

Then the secondary voltage becomes:

$$E_o = 9.33 \times \left(\frac{1}{1/6.58}\right) = 1.42 \text{ volts}$$

This value will provide an output just barely short of our +6 VU requirements, but it will be adequate for most telephone loops. If it is determined that it is not, then it will be necessary to use higher plate voltages and rework the entire design, choose higher gain tubes or even an additional stage.

A volume control is desirable, especially if fade-ins and fade-outs are necessary, and this can be simply accomplished by making the grid leak resistor in the 6C5 stage variable. Since only two stages are being used, there are no decoupling filters. Our design is therefore now complete and is shown schematically in Fig. 328.

chapter

4

electronic power supplies

All radio receivers (with the exception of the simple crystal and transistor types), all transmitters and all audio amplifiers employ vacuum tubes. Each of these tubes requires some sort of voltage at most of its elements.

Of itself, no tube is capable of generating power, but simply acts to control and modify those voltages which are supplied to it. The process begins with the *thermionic emission* of electrons, whereby a specially constructed cathode acts as the source of current flow when its temperature is raised. Any source of heat could theoretically excite this cathode to the point of emission, but the method actually employed uses a filament of high-resistance wire which becomes very hot when electric current passes through it. Hence a source of voltage is required for heating the filament and causing the cathode to begin emitting electrons.

These electrons will flow through the vacuum in the tube and into an anode or plate, provided this plate is positively charged. Another source of voltage, positive at all times with respect to the cathode, is therefore necessary to permit the tube to conduct current.

The flow of current through the tube is controlled and modified mainly by a third electrode within the tube, known as the control grid. The action of the grid in controlling the flow from cathode to plate is determined by the incoming signal voltage. The control grid will do its job much better if a bias voltage is applied to it which will cause it to be at a negative potential at all times with respect to the cathode. Thus the control grid also must have a source of steady voltage.

If the tubes in an audio amplifier have more than the minimum triode elements, as is most often the case, the additional electrodes must be maintained at given potentials also. The screen grid is normally maintained positive, although usually less so than the plate. The suppressor grid (or beam electrodes in a beam-power tube) has no requirement of an external voltage. This electrode is maintained at cathode potential, usually through an internal connection between the elements in the tube.

All of these voltages may be supplied by cells and batteries. A basic audio amplifier operating in this fashion is shown in Fig. 401. Such a system has a number of disadvantages, however, especially when several tubes are required in the circuit. Methods

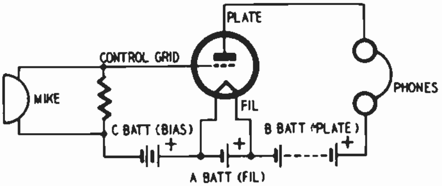


Fig. 401. *There are three voltage supplies in this circuit. The A battery furnishes filament voltage, the B battery is for plate voltage while the C battery biases the control grid.*

have therefore been developed which permit the use of ordinary house current, which is usually supplied at 117 volts at 60 c.p.s.

Design features of power supplies

It would seem that the filament of a vacuum tube should heat just as readily with a.c. as with d.c., and that we could eliminate the A-battery simply by reducing the 110 volts to whatever is required by the filament, using either a dropping resistor or a transformer. But in tubes designed for battery operation, the filament itself acts as the emitter. The filaments of such tubes cool and heat much too quickly for a.c. operation, their temperature and emission varying in step with the a.c. cycle causing hum.

The solution to this problem was twofold and resulted in some redesign of the cathodes of tubes intended for a.c. operation. Since greater voltages could now be used than the 1.5 to 2 volts supplied by most A-batteries, it was possible to increase the size and capacity of the filament, so that it would have considerable thermal lag and respond much less readily to the variations in current flowing through it. Secondly the heater and cathode functions of the old filament were separated, the emitter being indirectly heated and therefore not subject to the rapid filament current variations. The use of an a.c. heater voltage is therefore now perfectly feasible, and one of the functions of our power supply system is to provide a means of reducing the incoming voltage to values which permit the correct operation of the various filaments. This is performed by the power transformer, which may have several low-voltage secondary windings of various values as required.

Negative grid voltage in audio amplifiers is generally provided by an R-C circuit connected in the cathode lead and for this reason the system is named *cathode bias.* The plate current, in returning to the cathode, causes a voltage drop across the resistor and the required potential difference between grid and cathode. A capaci-

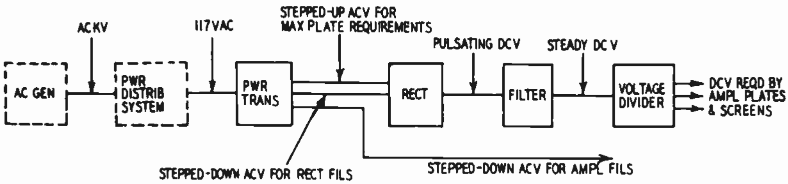


Fig. 402. *Block diagram of a power supply. Almost all supplies follow this same general pattern.*

tor across the resistor acts to smooth out the audio fluctuations in the plate current at this point so that the grid will not be modulated. Design considerations of the cathode-bias system were discussed in Chapter 3.

With the A- and C-batteries gone, we are now faced with the problem of eliminating the battery which supplies the plate and screen voltages and which in the early days was known as the "B-eliminator." Referring to Fig. 402, we see that such a system includes these four basic units: (1) power transformer, (2) rectifier, (3) filter and (4) voltage divider. Now let us see the functions performed by each of these elements and learn something about their design.

Since we require operating voltages greater than 110, we take advantage of the fact that our supply is a.c. at this point and very simply step it up by means of a transformer. At the same time of course, as we have already said, we also step down the incoming voltage to the low values required by the various tube filaments. Since the plate voltages required for proper operation of the tubes in an audio amplifier run upward to 300, one of the functions of the power transformer is to change the line voltage to this value, allowing sufficient leeway for succeeding circuit losses.

The power transformer will therefore often have several windings. The primary, of course, will be designed to operate at the value of the incoming voltage, usually 110. The main secondary will be a high-voltage plate winding to provide what will ultimately become several hundred volts of pure d.c. for the tube plates and screen grids. One or more additional secondary windings will provide the required filament voltages. These voltages are not converted to d.c. but are applied to the filaments just as they are. More than one winding for this purpose may be necessary either because differing values of filament voltage are required or because so many tubes are involved that a single winding

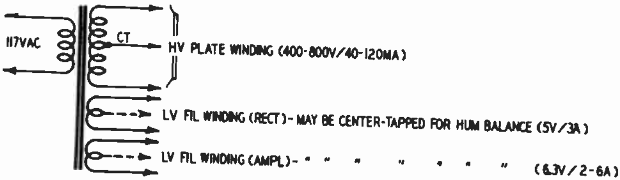


Fig. 403. *Typical power transformer circuit. The size of the transformer will be determined by the amount of power to be delivered by it.*

will not have sufficient current-carrying capacity. A typical power transformer as used in audio amplifiers is shown in Fig. 403.

Rectification

The high voltage from the secondary plate winding is converted to d.c. by means of a *rectifier*, which in most cases is a thermionic vacuum tube. Since the tube as a conductor is a strictly one-way device, it can be used quite readily for this application. As a quick review of this important concept, let us refer to Fig. 404, which shows the simplest type of battery-operated diode, but with a switch in the plate circuit which can reverse the polarity of the B-battery.

When the double-pole double-throw switch is in the No. 1 position, as shown, the plate will be positive, will collect the electrons emitted by the filament and a flow of current will travel through the load resistor in the direction shown. But when the switch is thrown to the No. 2 position, the battery polarity will be reversed and the plate will become negative with respect to the cathode. At this point no current will flow, for the plate is not an emitter and there will be no way for the electrons to get from plate to cathode. Under these conditions, therefore, the tube has become a nonconductor. Then if the switch were to be thrown back and forth repeatedly and very rapidly, the voltage appearing across its center pair of terminals might be said to be *alternating*, but still current would flow only during the positive alternations. Similarly, if a true a.c. voltage were to be applied to the plate by a generator replacing the battery and switch arrangement, the flow of current would still be confined to the positive halves of the cycles. Reducing this to practical terms results in a transformer applying the

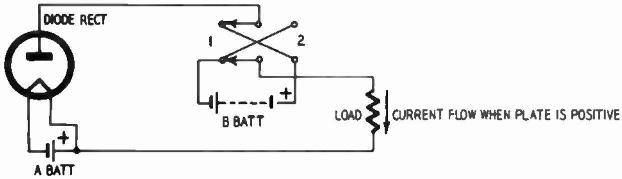


Fig. 404. *Current will flow through the diode only when the plate is positive with respect to the filament (or cathode). The diode behaves like a valve.*

a.c. plate and filament voltages as shown in Fig. 405.

The voltage across the plate winding is still a.c., of course, which means that opposite ends of the winding will be alternately negative and positive. Now during the half-cycle when the lower end of the winding is positive, the plate is also positive and current flows through the load. But when the upper end of the winding is positive, the lower end is negative, the plate is negative and no plate current flows. Thus we have effectively cut off the nega-

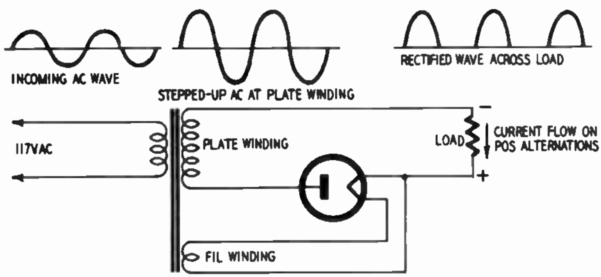


Fig. 405. *The line voltage is stepped up by the transformer and is rectified by the diode. The output consists of pulses.*

tive halves of each cycle, as shown in the illustration of the rectified wave across the load. Since this current is all in the same direction, even though it is of varying value, it is properly regarded as d.c. Since only a half of each cycle appears at the output, this system is known as *half-wave rectification.*

The full-wave rectifier

Obviously we could exactly double the usefulness of this voltage if we could somehow cause the negative half-cycles to "flop over" and become positive, thereby filling in the gaps between the positive output pulses of Fig. 405. A little thought on the problem might suggest a possible solution involving a second rectifier system, so connected that its positive pulses occurred exactly when those of the first system were negative. Then the two rectifiers could alternately impress positive voltages across the common load. Such a system might look something like that of Fig. 406.

Here we can see that as the bottom of plate winding No. 1 becomes positive, the same condition will be obtained as in the system of Fig. 405. V1 will conduct and current will flow through

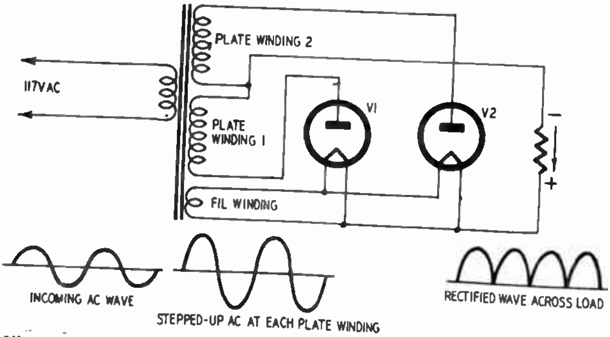


Fig. 406. *The action here is the same as that shown in Fig. 405 except that we have more pulses in the output. This type of supply is more desirable for audio work.*

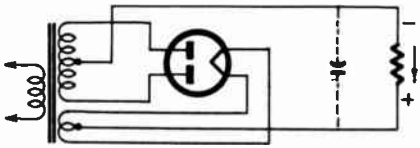
the load. At the same time, the bottom of plate winding No. 2 is also positive, thus leaving the plate negative and V2 therefore nonconductive. But on the other half of the cycle, when the plate of V1 goes negative, the top ends of each winding are positive, which is the correct condition for conduction through V2. Thus current continues to flow through the load in the same direction, first through V1 while V2 remains idle and then the reverse as the polarity of the cycle changes.

We have now achieved the *full-wave rectification* which we sought, but at the expense of an additional tube and an additional transformer winding. The system is quite workable, but in actual practice it is only used in high-power applications such as radio and television transmitters. In smaller installations we can eliminate the two parallel power leads going to the same point on the load through the use of a center-tapped transformer and a single conductor. Furthermore, since the tube filaments are in parallel, we can replace them with a single larger filament supplying electrons to two separate plates within the same tube. The use of such a tube, sometimes known as a *dual-diode,* as a full-wave rectifier is shown in Fig. 407.

Two terms which help to define the merit of a given rectifier tube are *inverse peak voltage* and *peak current*. The inverse peak voltage is the maximum e.m.f. which the tube can withstand between plate and cathode on the negative half of the wave, when the cathode is positive and the plate negative. Voltages in excess of this value may cause an arc or *flashover* between the electrodes, with consequent damage to the tube.

There is a maximum amount of current which the rectifier tube can carry between cathode and plate, and exceeding this value even momentarily may also result in severe damage to the tube structure. This quantity is known as the *maximum peak plate*

Fig. 407. *The two diodes can be combined into a single tube. This is a standard power supply circuit.*



current. The actual amount of current which will flow in any given rectifier circuit will depend in part upon the type of filter used with it and the values of the filter components.

Filtering

Although we have achieved full-wave rectification, the *pulsating d.c.* output wave of Fig. 406 is still not nearly good enough to apply to the tubes of our amplifier circuit. Under these conditions, the plates of the tubes would be very heavily modulated at the *ripple* frequency, and the resultant audible output would consist very largely of hum. Thus, some means must be found for smoothing this pulsating d.c. and changing it to pure steady d.c.

Ideally we might pass this voltage through an audio filter, which would be of the low-pass type with a cutoff frequency of exactly 0 c.p.s. While such a filter might not be altogether practical, we can approach it, and this is the basic notion of the method used in power-supply filtering.

Suppose we begin by connecting a capacitor of suitable size across the load resistor of Fig. 407. Already we have a sort of parallel R-C filter, which we know is useful in smoothing audio variations when used as a cathode-bias arrangement. Now let's take a little closer look and try to learn just what occurs across the combination when a pulsating d.c. wave passes through it.

When the voltage is first applied, the capacitor charges to its peak value. Then as the voltage decreases, the capacitor begins to discharge, but not as rapidly as the pulsating wave. Instead, the capacitor follows its characteristic *exponential discharge curve,* as shown in Fig. 408. The capacitor continues to discharge just to the point where the rising voltage of the next incoming peak equals the value to which it has discharged. Then the charge on the capacitor goes back up to the peak and the whole process is repeated.

This system smoothes the ripple content of the rectifier output quite noticeably, but it still isn't good enough in actual practice. Since an inductance has some rather interesting effects on a varying current passing through it, perhaps it too would be useful in this application. Lenz's law states that an inductance tends to oppose any change in the amount of current flowing through it, and this is the very characteristic we require. Because this action

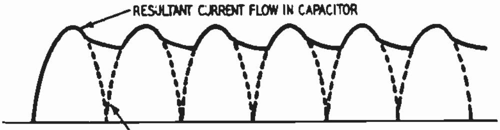


Fig. 408. *The filter changes the output of the rectifier from pulsating d.c. to a somewhat smoother voltage.*

is directly proportional to the size of the inductance, we should use a rather large one, with an iron core, commonly known as a *filter choke coil.* This would give us the basic filter circuit shown in Fig. 409.

Ripple voltage

It may not be clear why the load is connected across the capacitor only, rather than across the L-C combination or in some other fashion. To understand this, we must think of full-wave pulsating d.c. as having two components. One of these is the pure d.c. and the other a ripple component of 120-c.p.s. frequency. Now consider the L-C circuit as a voltage divider for the ripple and see what happens to this component. Assume a 30-henry choke, which has a reactance at 120 c.p.s. of about 22,000 ohms. A capacitor of 10 μf would have a reactance of around 130 ohms at this same frequency. Obviously, then, most of the ripple voltage is being dropped across the reactance of the choke. Numerically, the amount of the total ripple voltage appearing across the capacitor will be proportional to the ratio of its own reactance to that of the entire circuit, or $(130/22{,}130) = .006 = 0.6\%$.

This last figure is known as the *ripple percentage*. Where the inductive reactance is much greater than the capacitive reactance, it may be expressed as the simple ratio X_C/X_L, from which we derive the following equation:

$$\frac{X_C}{X_L} = \frac{\frac{1}{2\pi fC}}{2\pi fL} = \frac{1}{(2\pi f)^2 LC} \qquad (12)$$

The ripple percentage then expresses the ratio of the a.c. ripple at the load to the ripple at the output of the rectifier—in a good

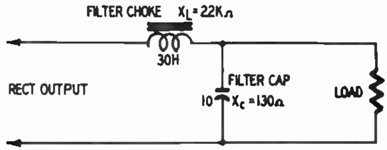


Fig. 409. *Simple power-supply filter circuit. This is known as an L-section or choke-input type.*

system well under a few hundredths of 1%. This means that our simple L-section filter of Fig. 409 would probably not be satisfactory in a high-quality audio setup. We can improve on it, though, simply by inserting additional sections of L and C, with three chokes and three capacitors not being at all uncommon. Other filter configurations are shown in Fig. 410.

You will see that some of these filters have a series choke first following the rectifier. These are known as *choke-input* filters, while those which have a parallel capacitor directly across the

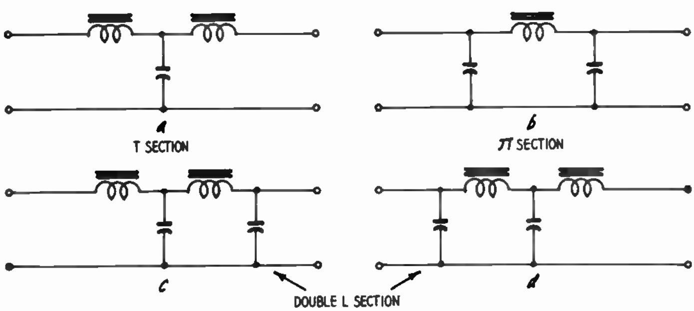


Fig. 410-a,-b,-c,-d. *Various types of power supply filters.*

rectifier output are known as *capacitor-input filters*. Under otherwise equal conditions, the capacitor-input filter will provide the higher output voltage of the two, but the choke-input system will provide better *voltage regulation*.

Voltage regulation

As soon as the load connected across a power supply begins to draw current, the terminal voltage at the output of the filter will decrease somewhat. This is due to voltage drops across the various resistances in the power supply circuit through which the current must flow, such as the ohmic resistance of the transformer and choke windings, as well as the drop across the rectifier tube itself due to its own plate resistance.

Voltage regulation is expressed as the ratio of the voltage drop at full load to the voltage under no-load conditions. For example, if a power supply delivers 375 volts at no load and 350 at full load, then the regulation is

$$\frac{375 - 350}{375} = .066 = 6\tfrac{2}{3}\%$$

A regulation figure greater than 15% is intolerable in an audio

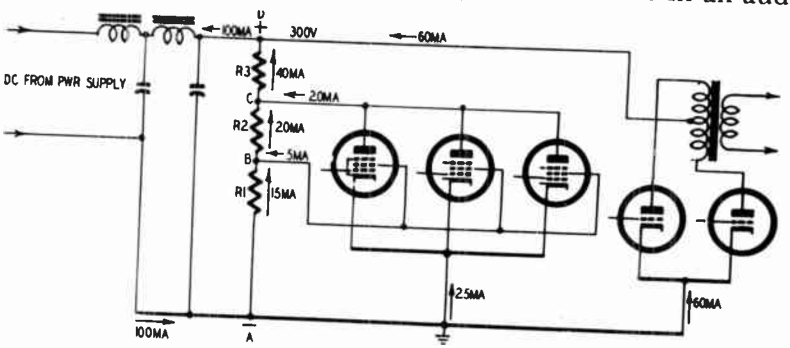


Fig. 411. *The power supply furnishes plate and screen voltages for the tubes. The tubes may be considered as a load in parallel with the output of the power supply.*

amplifier, and preferably it should be much less than that. In some applications, such as the oscillator circuit of a transmitter or the local oscillator in a superheterodyne receiver, greater voltage stability is necessary in order for the system to remain on frequency. This is accomplished by means of a gaseous voltage-regulator tube. In class-A audio operation, however, this refinement is usually unnecessary.

In most power supplies, the rectifier is of the filament-emitter type which begins to pass current almost the instant voltage is applied. But the tubes used in the amplifiers are normally of the indirectly heated type, which take considerably longer to warm up and begin conducting. For this reason a *bleeder resistor* is placed across the power supply output to minimize high-voltage surges through the rectifier when it is first turned on. It also aids in voltage regulation and discharges the filter capacitors after the equipment has been turned off, thus eliminating shock hazard during servicing operations. The best value of bleeder resistance has a bleeder current which is 10 to 20% of the total load current.

The voltage divider

This same bleeder resistor can also be used as a voltage divider if properly designed and provided with convenient taps, and this

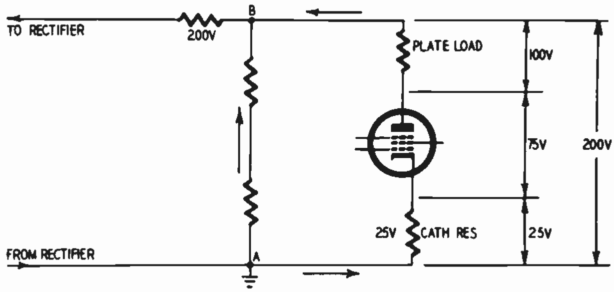


Fig. 412. *The cathode resistor, tube, and plate load form a voltage divider across the output of the power supply.*

is the method most often employed in actual practice. A typical circuit is shown in Fig. 411 wherein the supply provides plate and screen voltages for three resistance-coupled pentodes and plate voltage for a pair of push-pull triodes. The actual amplifier circuit has been deleted for simplicity.

In this case the triodes require 60 ma of plate current at 300 volts, the pentode plates a total of 20 ma at 200 volts, and the pentode screens 5 ma at 90 volts. Thus the maximum voltage required is 300 and the total tube current is 85 ma. To this we will add 15 ma of bleeder current, making a total current drain of 100 ma. Now let's design the necessary voltage divider.

The 100 ma of current in the lower leg of the filter circuit splits at point A, 85 ma going to the tube cathodes and the 15-ma bleeder current going through R1. The tube current further divides, with 25 ma flowing to the cathodes of the pentodes and 60 ma to the cathodes of the triodes. The pentode current divides in the tubes, with 5 ma going to the screens and 20 ma to the plates. The screen current returns to the voltage divider at point B, which must be 90 volts above ground. Then we can calculate R1 by Ohm's law: $R = E/I = 90/.015 = 6{,}000$ ohms.

The pentode plate current rejoins the divider at point C, which is at +200 volts. Since the drop across R1 is 90 volts, the drop across R2 must be 200 – 90, or 110 volts. The current through

R2 comprises 15 ma of bleeder current and 5 ma from the screens, a total of 20 ma. Then R2 = 110/.020 = 5,500 ohms.

R3 carries 15 ma of bleeder current, plus 5 ma of screen current, plus 20 ma of plate current—a total of 40 ma. The drop across this resistor must be 300 – 200 = 100 volts. Then R3 = 100/.010 = 2,500 ohms.

The remaining question is the power rating of the resistors, which should allow a safety margin of at least 100%. The power dissipated in R1 is E×I = 90 × .015 = 1.35 watts. Thus the resistor employed for R1 should be rated at around 3 watts. Similarly, R2 dissipates 110 × 0.020, or 2.2 watts, and should therefore be rated at about 5 watts. Finally, R3 uses 100 × 0.040, or 4 watts, requiring a resistor rated at 8 watts or better.

The above calculations have presupposed that the supply voltage requirements took into account any losses in decoupling filters. Since in this case we have more than two stages operating from

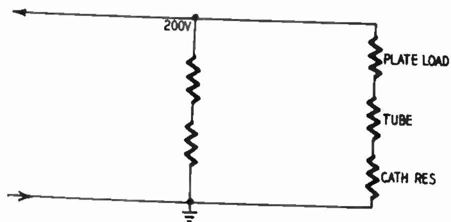


Fig. 413. *The circuit of Fig. 412 becomes simplified if we consider the tube as a resistor.*

a common power supply, such filters will be necessary to prevent motorboating. For the sake of simplicity, plate-load resistors, screen-dropping resistors and cathode resistors have been omitted.

If you will examine Fig. 412 you will see the relationship of the plate-load and cathode resistors to the power-supply voltage divider. These resistors are in series with the tube, but the entire combination of tube and resistors is shunted by (in parallel with) the power supply output. The current, coming from the rectifier tube, divides at point A. Some of the current flows up through the power-supply voltage divider, while the balance of the current goes through the cathode resistor, through the tube, through the plate-load, rejoining the divider current at point B.

If we consider the audio tube as a resistor, then the circuit of Fig. 412 can be simplified to appear as in Fig. 413. The amount of current flowing through each branch network will depend upon the total resistance in each branch. Fig. 413 further emphasizes the fact that the tube, its load and its cathode resistor are shunted across the power-supply output.

chapter 5

distortion and noise

True fidelity of sound reproduction will occur only at the time when the ear and brain receive precisely the same aural impression during reproduction as they would if actually present at the sound source. While fidelity of reproduction has been improving constantly, the ultimate in perfection has not yet been achieved. This means that we still have remaining some *distortion*.

A reproduced sound is distorted if it contains additional components which were not present in the original or if it lacks certain features which were present in the original. As fidelity has improved, some well known forms of distortion have become much less significant but at the same time other more obscure types have become more apparent, and so the battle goes on. Since distortion is the arch-enemy in the quest for perfect reproduction, it behooves us to know as much as possible about its character, its causes and cures; so let's first survey the distortion picture and then examine the more troublesome forms in greater detail.

Forms of distortion

One of the first criteria of audio reproduction is the frequency response of the entire system. Fig. 501 shows the frequency response curve of a typical amplifier system with satisfactory performance over the audio range. It is known that frequency response in humans varies between trained and untrained ears and between those of young people and of the aged. It is also known that the frequency range required of a system for speech reproduction only is much narrower than for music. A system incapable of carrying all of the audible sounds present in the original is guilty of *frequency distortion.*

We also know that an improperly designed amplifier may produce spurious harmonics within its own circuitry, these harmonics serving to alter the waveshape of the incoming signal. It has been determined by listening tests that when the full audible range is reproduced, distortion on the order of 3% to 5% is noticeable and 10% distortion is highly objectionable. We already know a number of ways to eliminate even *harmonic distortion* in amplifier design, but unfortunately the odd-order components are more

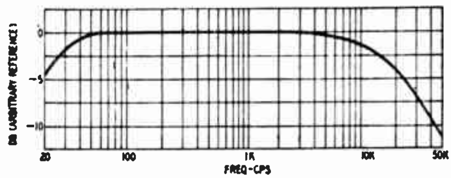


Fig. 501. *Frequency-response curve of a typical audio amplifier system.*

annoying to the ear. We have learned by experiment that more of this type of distortion, sometimes known as *amplitude distortion* (Fig. 502), can be tolerated if the frequency range of the system is simultaneously reduced.

When sound is reproduced at a different volume level from the original, the ear will be operating with a differing sensitivity characteristic and there will appear to be an imbalance among the various frequency components of the original sound. This is especially apparent during low-level reproduction, when the high and low ends seem to disappear, due to the insensitivity of the ear itself. This phenomenon is sometimes referred to as *scale distortion* and may be corrected by a tone-compensating volume control known as a *loudness control.*

The acoustics of the studio or auditorium where the original sound pickup is made, as well as the environment of the listening room, will have a considerable effect upon the fidelity of the various sounds which are to be reproduced. The ear is quite able to concentrate upon and bring into focus the sound emanating from a given source and at the same time can discriminate against noise and reverberation arriving along other paths. The microphone, however, is unable to make this discrimination, with the result that the reproduced noise and reverberation sound quite different from their original form. Much of the ability of the human hearing apparatus to make such distinctions, and to identify the direction of arrival of a sound source results from the fact that we possess two ears. But an ordinary audio system is only *monaural* in that all of the reproduced sound goes through a single

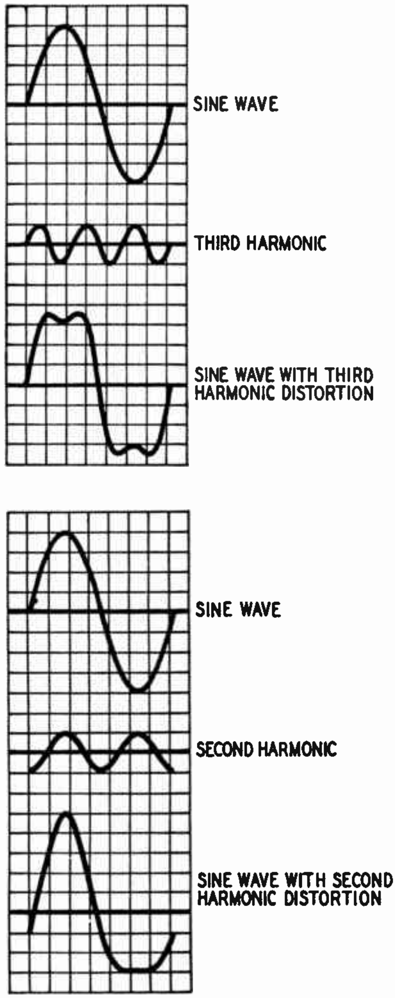


Fig. 502. *Effect of second and third harmonic distortion on a sine wave.*

channel and is emitted from a single source, regardless of the location of the original sounds with respect to the microphone. An illusion of space and direction is possible through *binaural* or *stereophonic* systems which employ two or more channels up to and including the reproducing loudspeakers. These systems, however, are expensive, difficult to adjust and still not altogether satisfactory.

If limiting or compression has been employed anywhere in the system, the original dynamic nuances have been deliberately distorted. The only cure for this is the use of volume expansion with a dynamic characteristic which is the precise complement of the original compression. As a practical matter, this is a near impossibility.

With so many intricate means by which distortion can creep into the system, it is obvious that the audio man has his hands full. Until fairly recently, frequency response and harmonic distortion were thought to be the full criteria of audio system performance. Now we know that this is only the beginning. These two fundamental forms of distortion are not to be ignored. On the contrary, they remain of fundamental importance although they have become less troublesome as design techniques have improved and as the full significance of other forms of distortion has been recognized.

Frequency distortion

Frequency discrimination is most commonly observed as the cutting or rolling off of the extreme high- and low-ends of the audio spectrum. It may, however, also be represented in the forms of peaks and valleys somewhere in the middle of the range. It is most pronounced when attempts are made to obtain too much gain per stage in an amplifier. Due to the usual discriminating tendency of the reactive components of a circuit (when the gain of a given circuit is raised to a very high value) the relative amplification of the middle tones with respect to the extreme highs and lows is increased out of proportion. Of course electromechanical devices, such as microphones, pickups and loudspeakers are particularly susceptible, due to such physical problems as mass, inertia and resonance. However, as far as audio amplifiers are concerned, as long as extremely high gain per stage is avoided and properly designed resistance coupling is employed, frequency distortion is no longer a very serious problem.

Harmonic distortion

In addition to the nonlinearity concept of harmonic distortion, we can also approach the problem from the idea of *resonance*. Every mechanical device has some sympathetic period at which it will vibrate much more strongly than at other frequencies, and every electrical circuit has a characteristic resonant frequency at which it will develop much greater voltages. In mechanical systems, such as pickups and speakers, every effort is made to damp out these resonance effects or to have them occur at some frequency outside the audible range. The same principles hold for electrical circuits, the resonance effects being defeated wherever possible by holding down the gain (Q) of the circuit.

Rating an amplifier in terms of overall harmonic distortion is quite fallacious, for it ignores the important fact that some harmonics are much more obnoxious than others. As an example, let us refer again to Fig. 103 where we find that a low A has a fundamental frequency of 55 c.p.s. The second harmonic of this note would of course be an octave above that, or 110 c.p.s. Since most instruments have a rather strong second harmonic, and some may very likely be doubling the note in that octave anyway, small additional amounts of the second harmonic will not be particularly objectionable. The third harmonic will be 3×55, or 165 c.p.s., an octave and a fifth above the fundamental. Now as long as the chord being played contains this interval, a little boosting of the third harmonic will be tolerable. But if any other harmonies are being played, then some very unpleasant dissonance effects will be observed.

The fourth harmonic will be 220 c.p.s., exactly two octaves above the fundamental. This, too, can stand a little more emphasis without offending the listener, but as a practical matter it is usually of rather insignificant magnitude anyway. The fifth harmonic, 275 c.p.s., doesn't even occur on the musical scale and so is particularly undesirable. The sixth harmonic, 330 c.p.s., is an interval of two octaves and a fifth, and the same comment applies as for the third harmonic. The seventh harmonic is 385 c.p.s. and is another of those which falls midway between two notes on the scale. This, too, must be particularly avoided.

It is now quite obvious why we can say that the even harmonics are, on the whole, much less objectionable than the odd-order products. Therefore, how much total harmonic distortion is present is not so important as *which* particular harmonics predominate.

Phase shift

When complex musical tones are passed through an audio amplifying system, a certain amount of phase shift will occur as a result of the reactive components (coils and capacitors) of the apparatus. But phase displacement is a function of frequency, and the shift will therefore vary for the several frequencies. That is, the amount of time required for a signal to pass through an amplifier will vary with frequency, with the result that not all of the frequencies fed into the system simultaneously will emerge with an identical time relationship. This condition is known as *phase distortion.* As a practical matter, it is relatively unimportant in most audio applications, since the ear will tolerate rather large amounts of this distortion without even recognizing it. It does become troublesome, however, when the total transmission time approaches the periods of the signal frequencies themselves. This sort of condition is encountered in long-distance network program lines. It is also of considerable importance in the video amplifiers of TV receivers.

Transient response

A very important criterion of an audio system is its ability to handle transients. These are the steep wavefronts generated by attacks and releases of musical tones. Whenever a musician tongues a mouthpiece, depresses a key or strikes a percussion instrument, there will be a very brief period between this initial attack and the point of peak amplitude. The sound generated during this period is really a noise, and as such it comprises many more and higher frequencies than are present in the tone itself. A system unable to cope with these instantaneous effects without "hash" or "ringing" is said to exhibit *transient distortion.*

Due to its mechanical inertia the loudspeaker is probably the worst offender in this area, but fortunately proper amplifier design can help to alleviate this difficulty. The usual 1:1 impedance match between circuits is not the best condition for minimum distortion when coupling the speaker to the amplifier. Instead, the output impedance is actually made several times larger than the speaker impedance, this ratio being known as the *damping factor.* Under these conditions the amplifier output will tend to behave as a short circuit to the transient distortions set up in the speaker system. In practice, damping factors between 3:1 and 5:1 are about the maximum useful figures. Beyond this, amplifier instability and decreased bass response assume serious importance.

Probably the two most serious types of distortion problems facing the audio designer and experimenter today are those of transients and intermodulation. Perhaps this is true because their full significance has only recently been recognized. *Intermodula-*

tion distortion (IM), for example, has long been known, but for many years it was simply thrown within the broad category of harmonic distortion and left there in relative obscurity. Of course it would be periodically "discovered" and reported upon by one experimenter or another, but these discussions seemed to leave only few lasting impressions upon the art until recent times when high fidelity itself was "discovered."

Cross-modulation

Before any confusion develops in the matter of terminology, we should note the important idea that intermodulation is not to be confused with *cross-modulation* distortion, which is still another phenomenon. Cross-modulation occurs when variations in the amplitude of one audio signal impress themselves upon another audio signal of a different frequency, with the result that the

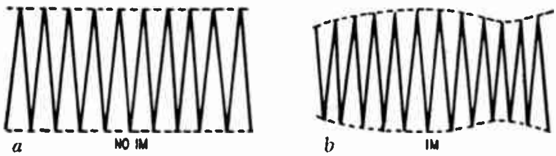


Fig. 503-a,-b. *Undistorted waveform (a) shows intermodulation distortion (b).*

amplitude of the second signal varies in accordance with the first. It is particularly noticeable in the case of two sustained tones which differ rather widely in frequency. If a violin and organ are played together, for example, the high-frequency tones of the violin may appear to vary in accordance with the more powerful low-frequency organ accompaniment.

Cross-modulation seems to be proportional to the rate of change of curvature in tube operating characteristics. It is therefore avoided by restricting the operating range to the area of the straight-line portion of the characteristic curve of the tubes in the amplifier.

Intermodulation

Intermodulation distortion is also the result of nonlinearity, but its manifestations are quite different from either cross-modulation or harmonic distortion, although it may be regarded as a secondary effect of harmonic distortion. When two or more frequencies pass through an amplifier which is producing considerable harmonic distortion, the output will contain, not only the original tones and the spurious harmonics, but also various *combination tones* as shown in Fig. 503-a,-b. This is analogous to heterodyning at radio frequencies and may result in various sum-and-difference combinations between the fundamentals and any of their harmonics. The number and amplitudes of these combinations increase not only with the percentage of harmonic distortion, but also with the order of the harmonics. Thus, a given percentage of fifth-harmonic distortion will produce greater amounts of intermodulation than the same percentage of third-harmonic distortion, while sixth, seventh, eighth and higher harmonics will produce increasingly greater amounts of intermodulation distortion for the same amount of harmonic distortion. This is one more reason why the percentages of the higher-order spurious harmonics must be dealt with so severely.

This secondary effect of harmonic distortion is really far more distasteful to the listener than are the harmonics themselves, and the reasons should be fairly obvious. Suppose that a simple sine-wave tone of a few hundred cycles is fed into an audio amplifier and that under these conditions the harmonic distortion content of the output is measured at 2.5%. Now suppose that this signal is removed and another signal of something less than 100 c.p.s. is applied, producing 4% distortion. It would seem then that if both tones were to be applied to the input simultaneously, the maximum possible distortion overall might be 6.5%. But in actual practice it might well be that the total distortion will have actually jumped from two to four times this figure. Obviously another and more serious factor has suddenly entered the picture. Spurious combination tones are being generated which have absolutely no harmonic relationship to the signal frequencies, and which produce much more ear-jangling results upon the listener.

The measurement of intermodulation distortion in amplifiers can be readily and accurately determined, although it requires equipment not normally within the reach of the amateur experimenter. But it can be done, and this is probably the explanation for the fact that the amplifier is by far the most perfectly refined of all the audio-frequency components. In the mechanical and acoustical elements, such as microphones, pickups and loudspeakers, distortion is determined more or less empirically. That is, the method is "cut and try," with the final proof being the effect upon the ear itself. In this area particularly, the art has still much room for improvement.

Flutter

Sound recording, whether the medium be disc, tape or film, involves the use of moving parts, and with this comes a group of mechanical troubles which may produce some of the most annoying forms of distortion. The recording medium must move during reproduction in *precisely* the same manner as it moved during recording, and any deviation from this will produce some form of distortion.

The simplest form of distortion of this type occurs when the rate of movement during recording and reproduction are both unvarying, but are at different speeds. Suppose, for example, that a disc recording turntable rotates precisely at the standard of 78.26 r.p.m., but that the record which it cuts is then played on a turntable rotating at 75 r.p.m. Then middle *A* will no longer vibrate at 440 c.p.s., but will have a pitch which is determined as follows:

$$440:A_1 = 78.26:75$$

$$78.26 \times A_1 = 75 \times 440$$

$$78.26 \times A_1 = 33{,}000$$

and if we now divide both sides of this equation by 78.26, we will get:

$$A_1 = 421.6 \text{ c.p.s.}$$

Thus, we have lowered the pitch by nearly a half-tone from that of the original performance. A constant change in pitch like this may be detected only by persons having highly trained musical ears, but to such people this form of distortion is absolutely intolerable. There are, however, related forms of distortion which do not require such acute perception but are equally objectionable to most listeners.

If the speed of the reproducer driving mechanism does not remain constant but instead varies periodically, it is obvious that the pitch of a sustained tone will also vary at the same rate as that of the speed change. Various terms are employed to describe the several forms this distortion takes, including *wow, flutter, hash, gargle, whiskers, waver* and *wobble*. There is considerable overlapping in the definitions of these terms, but the differences are based primarily upon the rate of change of the speed of the drive mechanism.

The total number of excursions in speed that the drive mechanism goes through in a second, as from maximum speed to minimum speed and then back to maximum, is known as the *flutter rate*. Perhaps the most common occurrence of distortion of this type is found in cheap phonograph turntables, where the rate is usually once per revolution. If the turntable has an operating speed of 78 r.p.m., it is 1.3 r.p.s., and the flutter rate of the once-around type would then be 1.3 c.p.s. We have learned experimentally that for a flutter rate of around 20 c.p.s. or less, the distortion which the ear hears takes the form of tones which sweep up and down across a spectrum whose limits are determined by the flutter amplitude. The further the speed varies above and below the norm, the greater will be the total pitch variation. As the flutter rate increases to about 40 c.p.s. or higher, the effect as perceived by the ear is somewhat different. At these frequencies, the pitch variations are too rapid to be heard as such and the ear hears instead a whole group of frequencies which have no harmonic relationship, and these effects are very similar to those of intermodulation.

This is only a coincident similarity, however, for the causes are quite dissimilar. The particular varieties of distortion presently under discussion are really forms of *frequency-modulation* distortion. In this case, the frequencies of the reproduced sound are varied or modulated in accordance with the rate of variation of the speed of the driving force. It is important once again to differentiate between this type of distortion and cross-modulation. The distinction is simple enough when it is remembered that cross-modulation is AM, wherein the *amplitude* is modulated, while we are here concerned with FM distortion, wherein the *frequency* is spuriously varied.

Perception of flutter

While the ear hears the higher-rate flutters as groups of spurious tones, in practice flutter at such a rate is of fairly low amplitude. It is certainly not negligible, but experience has shown that the average listener is most sensitive to flutter (or wow) in the region of 1 to 8 c.p.s. Since the once-around type on a 78-r.p.m. turntable falls within this area, it is apparent why this form of distortion is so well known.

Flutter is generally the result of mechanical aberrations in some part of the drive mechanism, due either to faulty design or excessive wear. (See Fig. 504) Any roughness or poor fit in the motor (or anywhere in the linkage) may cause it, whether the driven member be a phonograph turntable, a tape recorder capstan or the sprockets and claws of a motion picture projector. In the case of motion pictures, the source of trouble may even be in the medium itself, for a poor fit between the sprockets and the sprocket holes on the film is a quite common source of flutter. In the case of 35-mm film, the flutter rate is 96 c.p.s., which is not too readily perceptible to the ear and which may be simply removed with a 100-c.p.s. high-pass filter without a great deal of detriment to the

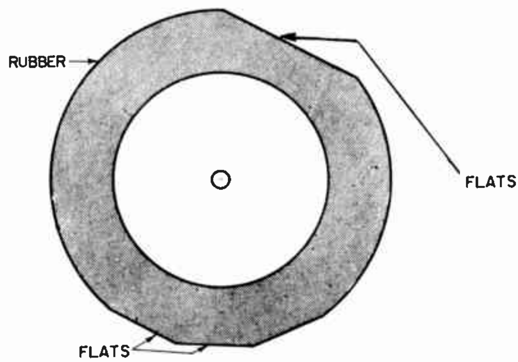


Fig. 504. *Flat spots in phono idler wheel produce sound distortion.*

fidelity of reproduction. But in 16 mm work, the flutter rate is 24 c.p.s., and here it does pose a serious problem.

FM distortion

We have described the several forms of flutter as being all members of the family of frequency-modulation distortion. Perhaps this may become a little clearer when we consider how this form of distortion occurs in a loudspeaker. In the case of a single-element speaker system whose voice coil receives signals encompassing the entire audio range, the speaker cone cannot vibrate as a unit for all frequencies. At the higher frequencies its mass is simply too great, and it will respond to these higher tones only in the region of its apex and around the voice coil. But at the same time the cone will vibrate as a whole on the lower frequencies, for its own mechanical period of resonance is in that region. As a result, the high-frequency vibrating segment will be frequency-modulated by the entire cone vibrating as a whole at the lower frequencies. This is the reason for multiple-speaker systems in which each radiator covers a range of only a few octaves. Unfortunately, no such easy solution exists for application to microphones and pickups, where the same trouble is encountered.

Here again, as in the case of intermodulation distortion, measurements of the actual amount and character of the distortion may be carried on much more successfully with the amplifier than with electromechanical devices. Since flutter is a form of frequency modulation, FM detection techniques are employed to measure that which exists at the output of the amplifier. The audio signal is regarded as the "carrier" and the flutter rate is the "modulation." The output of the amplifier is then passed through a conventional FM discriminator circuit whose output in turn provides a voltage which is a function of the *flutter percentage*. This figure is simply the ratio of the amount of frequency deviation to the mean frequency, expressed as a percentage.

Amplifier noise

All audio amplifiers deliver a certain amount of output voltage to their loudspeakers even when there is no signal being impressed across the input. This residual signal, whatever its origin, is generally classified as *noise*, and every effort is made to keep it as small as possible. The figure of merit for the noise level in amplifiers is the *signal-to-noise ratio,* expressed in decibels. A figure of 65 to 70 db is not uncommon in a good modern amplifier.

Probably one of the commonest sources of noise trouble in amplifiers is *hum*. This may be due to induction from neighboring circuits carrying alternating currents or by direct transmission from an improperly filtered power supply. It is particularly troublesome in high-gain audio amplifiers, for any hum which gets into the low-level stages at the front end will be tremendously magnified as it passes through the succeeding stages. The slight hum developed in these early stages by a.c. heater voltages may become troublesome, and for this reason the low-level preamplifier stages are often supplied with d.c. heater voltages.

Magnetic induction may occur in audio transformers, amplifier wiring or even inside the tubes. Principal sources of induction voltage are the a.c. power cable, power transformers, filter chokes and filament leads. Trouble from the filament leads may be minimized by twisting and properly routing them with respect to the components. Induction from transformers and chokes is controlled by very careful placement and by designing them to have as low a leakage flux as possible. Audio transformers are used sparingly, and any that are used are heavily shielded and kept at some distance from the power transformer and chokes.

Electrostatic induction can also cause trouble, particularly with those parts of the amplifier which have a high impedance to ground, for an induced current will flow and develop a hum which will be directly proportional to the impedance through which it flows. Trouble from this source is also most serious at the front end of the amplifier, where it is usually necessary to shield the tubes and grid leads or even to enclose in a metal box an entire stage or more along with the associated wiring. And finally, the chassis itself must be carefully grounded.

Noise produced by the mechanical vibration of the tube elements is known as *microphonics.* It may be transmitted to the tube either mechanically through its base or envelope or acoustically through sound waves in air. The former difficulty is encountered when the equipment is mounted in the vicinity of rather intense vibration, as in an automobile or airplane. while acoustic microphonics will always occur when an amplifier and loudspeaker are mounted in a common cabinet. The only cure for the latter type of trouble is separate mounting locations for speaker and amplifier. The direct transmission of vibration will be minimized by shock-mounting the amplifier and by the choice of tubes which are least susceptible to microphonic disturbances. Individual tubes, even of the same type, will vary tremendously in their microphonic sensitivity. If troubles of this sort are encountered, a number of tubes of the same type should be tried, or special nonmicrophonic tubes should be employed.

Other sources of noise in amplifiers are bad contacts, thermal agitation, shot effect, faulty resistors, leaky capacitors and weak batteries. Obviously all points of contact, such as switches, sockets and joints, must be clean and electrically sound. And it is equally obvious that components whose performance has deteriorated must be replaced.

In the case of faulty carbon resistors, the contact resistance varies constantly between adjacent granules of carbon, and any voltage appearing across such resistors will cause a current flow which varies slightly with the varying resistance. This produces a hissing sound in the output which can only be overcome by replacing the resistor with another type.

Another source of noise within the tube is *shot effect.* This results from the fact that the electron flow consists of a series of particles battering the plate, somewhat like hail on a tin roof. Most of this irregularity is smoothed out by the presence of a good space charge within the tube, however, and it is therefore imperative that the electron emission from the cathode be adequate to maintain this space charge. This explains why some tubes get noisy with age and why an emission test is a useful function on any tube checker.

With all other sources of noise eliminated, there is still the residual output produced by *thermal agitation,* for which no known cure exists. This is due to the random flow of electrons in the amplifier wiring, particularly at the input circuit. The noise generated in this fashion is uniform throughout the spectrum, from 0 c.p.s. right up through radio frequencies and beyond. While we know no way to avoid it, we do know that the amount of agitation, and therefore the amount of the noise, is directly proportional to the temperature. This is another good argument for adequate ventilation and for keeping operating temperatures as low as possible.

chapter
6

attenuators and equalizers

A fixed attenuator or *pad* is simply a network of resistance elements which is intended to lower the audio signal level between two points without introducing any frequency or phase distortion. Sometimes the loss is held to just a few db when it is desired to provide a degree of isolation between two circuits as, for example, between the output of a line amplifier and the program line itself. A pad may also serve as a very effective impedance-matching device, in which case the loss in the circuit is kept as low as possible.

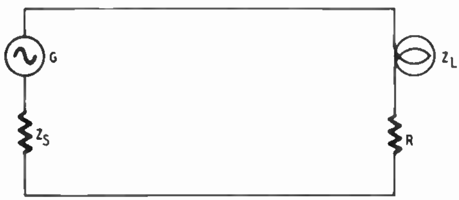


Fig. 601. *In its simplest form an attenuator is a single fixed resistor as shown in the illustration at the right. Attenuators can be used to provide a desired reduction in gain, or for matching the impedances of two circuits, or both.*

The simplest sort of fixed attenuator is the single resistor connected in series with the load, as shown in Fig. 601. In this circuit a generator G, with an internal resistance Z_s, supplies a voltage which is too high for the safe operation of the lamp Z_L. A resistor R is therefore inserted in series to drop the voltage across the lamp to a safe value. Now it is obvious that the internal resistance of Z_L will be a fixed value determined by its design and construction, and the value of R will be determined by the amount of current flow and the value of the desired voltage drop. Hence if the values of the source and load impedances should happen to match, it would be purely accidental. This is a commonly accepted fact in power circuits, but for audio work it is not good enough.

There is a definite optimum load impedance which an amplifier must see at its output if harmonic distortion is to be held to the minimum. Furthermore, filters and equalizers are designed for certain fixed values of source and load impedance, and a change in either or both of these values will alter the intended frequency characteristics. For this reason, the more complex pads and variable attenuators which we will discuss are quite necessary and justified, for they can be designed for both the proper attenuation and impedance characteristics.

Insertion loss

The loss introduced by a pad is known as the *insertion loss* and is defined as the db ratio of the power input to the power output at the attenuator, when the source and load impedances are equal. If the impedances are not equal, the concept is altered slightly to define the loss as that which we would obtain if the attenuator were replacing a perfect matching transformer inserted between source and load. The usual db formula then applies, except that for convenience in further calculations the power ratio is designated as k^2. Then

$$k^2 = \frac{\text{attenuator power output}}{\text{attenuator power input}} \tag{13}$$

and

$$\text{db (loss)} = 10 \log k^2 = 20 \log k \tag{14}$$

One of the simplest pads which will provide an impedance match along with any desired value of attenuation is the T-network shown in Fig. 602. The design problem here is to obtain values for R1, R2 and and R3, when the pad is actually connected between source and load and when the following requirements are met: (1) the impedance presented by the pad to the source across terminals A and B is equal to Z_s; (2) the impedance presented by the pad to the load across terminals C and D is equal to Z_L and (3) the db attenuation must be the figure called for in the design specifications, or the power delivered to Z_L must equal attenuator power input $\times$ k^2. These conditions will be satisfied by the following design equations:

$$R1 = 2\sqrt{Z_s Z_L}\left(\frac{k}{k^2 - 1}\right) \tag{15}$$

$$R2 = Z_s \left(\frac{k^2 + 1}{k^2 - 1}\right) - R1 \qquad (16)$$

$$R3 = Z_L \left(\frac{k^2 + 1}{k^2 - 1}\right) - R1 \qquad (17)$$

Fixed attenuator design

As a practical example of the use of equations (14), (15), (16) and (17), let us design a T-pad having a 25-db loss and matching

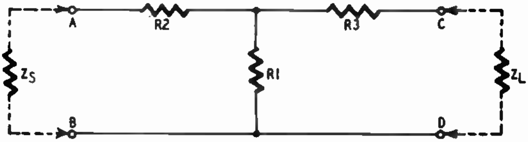


Fig. 602. The basic attenuator circuit at the right resembles the letter T, hence its name. The design of a typical T-pad is described in the text.

a source impedance Z_s of 600 ohms to a load impedance Z_L of 250 ohms. Since we already have decided upon a value for k^2 of 25 db, we do not need to work out equation (13). Using equation (14) to solve for k we get:

$$10 \log k^2 = 25 \qquad \log k^2 = 2.5 \qquad k^2 = \text{antilog } 2.5 = 316.2$$

$$k = \sqrt{316.2} = 17.8$$

Equation (15) is used to solve for R1 as follows:

$$R1 = 2\sqrt{600 \times 250}\left(\frac{17.8}{315.2}\right) = 43.3 \text{ ohms}$$

The value of R2 is found using equation (16). This gives:

$$R2 = 600\left(\frac{317.2}{315.2}\right) - 43.3 = 556.7 \text{ ohms}$$

Using equation (17) to solve for R3 we get:

$$R3 = 250\left(\frac{317.2}{315.2}\right) - 43.3 = 206.7 \text{ ohms}$$

A further glance at our original equations will indicate that, when Z_s and Z_L are equal, R_2 and R_3 will also be equal. This con-

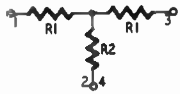
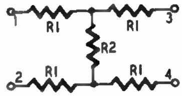
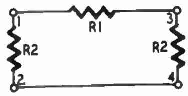
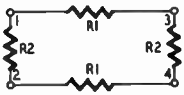
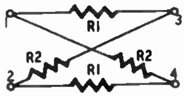
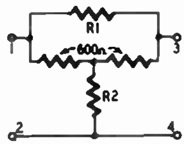
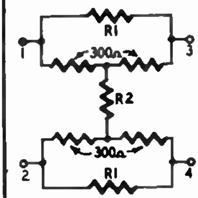
						
Impedance	600 Ohms		600 Ohms		600 Ohms	
Loss, db	R_1 Ohms	R_2 Ohms	R_1 Ohms	R_2 Ohms	R_1 Ohms	R_2 Ohms
0	0	∞	0	∞	0	∞
0.1	3.58	50204	1.79	50204	7.20	100500
0.2	6.82	26280	3.41	26280	13.70	57380
0.3	10.32	17460	5.16	17460	20.55	34900
0.4	13.79	13068	6.90	13068	27.50	26100
0.5	17.20	10464	8.60	10464	34.40	20920
0.6	20.9	8640	10.45	8640	41.7	17230
0.7	24.2	7428	12.1	7428	48.5	14880
0.8	27.5	6540	13.75	6540	55.05	13100
0.9	31.02	5787	15.51	5787	62.3	11600
1.0	34.5	5208	17.25	5208	68.6	10440
1.5	51.8	3452	25.9	3452	104.3	6950
2.0	68.8	2582	34.4	2582	139.4	5232
2.5	85.9	2053	42.9	2053	175.4	4195
3.0	102.7	1703	51.3	1703	212.5	3505
3.5	119.2	1448	59.6	1448	258.0	3021
4.0	135.8	1249	67.9	1249	287.5	2651
4.5	152.2	1109	76.1	1109	324.6	2365
5.0	168.1	987.6	84.1	987.6	364.5	2141
5.5	184.0	886.8	92.0	886.8	405.9	1956
6.0	199.3	803.4	99.7	803.4	447.5	1807
6.5	214.6	730.8	107.3	730.8	492.6	1679
7.0	229.7	685.2	114.8	685.2	537.0	1569
7.5	244.2	615.6	122.1	615.6	584.7	1475
8.0	258.4	567.6	129.2	567.6	634.2	1393
8.5	272.3	525.0	136.1	525.0	685.5	1322
9.0	285.8	487.2	142.9	487.2	738.9	1260
9.5	298.9	453.0	149.5	453.0	794.4	1204
10.0	312.0	421.6	156.0	421.6	854.1	1154
11.0	336.1	367.4	168.1	367.4	979.8	1071
12.0	359.1	321.7	179.5	321.7	1119	1002
13.0	380.5	282.8	190.3	282.8	1273	946.1
14.0	400.4	249.4	200.2	249.4	1443	899.1
15.0	418.8	220.4	209.4	220.4	1632	859.6
16.0	435.8	195.1	217.9	195.1	1847	826.0
17.0	451.5	172.9	225.7	172.9	2083	797.3
18.0	465.8	152.5	232.9	152.5	2344	772.8
19.0	479.0	136.4	239.5	136.4	2670	751.7
20.0	490.4	121.2	245.2	121.2	2970	733.3
22.0	511.7	95.9	255.9	95.9	3753	703.6
24.0	528.8	76.0	264.4	76.0	4737	680.8
26.0	542.7	60.3	271.4	60.3	5985	663.4
28.0	554.1	47.8	277.0	47.8	7550	649.7
30.0	563.0	37.99	281.6	37.99	9500	639.2
32.0	570.6	30.16	285.3	30.16	11930	630.9
34.0	576.5	23.95	288.3	23.95	15000	624.4
36.0	581.1	18.98	290.6	18.98	18960	619.3
38.0	585.1	15.11	292.5	15.11	23820	615.3
40.0	588.1	12.00	294.1	12.00	30000	612.1

Fig. 603. *This table of design values is convenient to use for the various attenuator networks illustrated above.*









600 Ohms		600 Ohms		600 Ohms		600 Ohms	
R_1 Ohms	R_2 Ohms	R_1 Ohms	R_2 Ohms	R_1 Ohms	R_2 Ohms	R_1 Ohms	R_2 Ohms
0	∞	0	∞	0	∞	0	∞
3.60	100500	3.58	100500	7.2	50000	3.6	50000
6.85	57380	6.82	57380	13.8	26086	6.9	26086
10.28	34900	10.32	34900	21.0	17143	10.5	17143
13.80	26100	13.79	26100	28.2	12766	14.1	12766
17.20	20920	17.20	20920	35.4	10169	17.7	10169
20.85	17230	20.9	17230	43.2	8333	21.6	8333
24.25	14880	24.2	14880	50.4	7143	25.2	7143
27.53	13100	27.5	13100	57.6	6250	28.8	6250
31.2	11600	31.02	11600	65.4	5504	32.7	5504
34.3	10440	34.5	10440	73.2	4918	36.6	4918
52.1	6950	51.8	6950	113.4	3174	56.7	3174
69.7	5232	68.8	5232	155.4	2316	77.7	2316
87.7	4195	85.9	4195	200.4	1796	100.2	1796
106.2	3505	102.7	3505	247.8	1452	123.9	1452
129.0	3021	119.2	3021	297.6	1209	148.8	1209
143.8	2651	135.8	2651	351.0	1025	175.5	1025
162.3	2365	152.2	2365	407.4	883.7	203.7	883.7
182.3	2141	168.1	2141	466.8	771.2	233.4	771.2
203.0	1956	184.0	1956	530.4	678.7	265.2	678.7
223.8	1807	199.3	1807	597.0	603.0	298.5	603.0
246.3	1679	214.6	1679	667.8	539.8	333.9	539.8
268.5	1569	229.7	1569	743.4	484.3	371.7	484.3
292.4	1475	244.2	1475	822.6	437.6	411.3	437.6
317.1	1393	258.4	1393	907.2	396.8	453.6	396.8
342.8	1322	272.3	1322	996.6	361.2	498.3	361.2
369.4	1260	285.8	1260	1091	329.9	545.5	329.9
397.2	1204	298.9	1204	1191	302.2	595.5	302.2
427.0	1154	312.0	1154	1297	277.5	648.5	277.5
489.9	1071	336.1	1071	1529	235.5	764.5	235.5
559.5	1002	359.1	1002	1788	201.3	894	201.3
636.3	946.1	380.5	946.1	2080	173.1	1040	173.1
721.5	899.1	400.4	899.1	2407	149.6	1204	149.6
816.0	859.6	418.8	859.6	2773	129.8	1387	129.8
923.2	826.0	435.8	826.0	3186	113.0	1598	113.0
1042	797.3	451.5	797.3	3648	98.68	1824	98.68
1172	772.8	465.8	772.8	4166	86.4	2083	86.4
1335	751.7	479.0	751.7	4748	75.8	2374	75.8
1485	733.3	490.4	733.3	5400	66.66	2700	66.66
1877	703.6	511.7	703.6	6954	51.72	3477	51.72
2369	680.8	528.8	680.8	8910	40.4	4455	40.4
2992	663.4	542.7	663.4	11370	31.66	5685	31.66
3775	649.7	554.1	649.7	14472	24.87	7236	24.87
4750	639.2	563.2	639.2	18372	19.58	9186	19.58
5967	630.9	570.6	630.9	23286	15.46	11643	15.46
7500	624.4	576.5	624.4	29472	12.21	14736	12.21
9480	619.3	581.1	619.3	37260	9.66	18630	9.66
11910	615.3	585.1	615.3	47058	7.65	23529	7.65
15000	612.1	588.1	612.1	59400	6.06	29700	6.06

The design of attenuator pads is considerably simplified when several of the elements have the same value of resistance.

siderably simplifies the design problem and since equal impedances are the general rule, the convenient table of Fig. 603 may be employed as a ready source of the most commonly required values.

Note that there are numerous additional types of networks illustrated other than the simple T, but each of them is intended to do the same basic job. The actual configuration chosen will depend upon whether the circuits with which it is working are balanced, and which of the several types provides the most convenient values for the available resistors.

Minimum-loss pads

When a resistance network is to be used solely as an impedance-matching device, it is desirable that the insertion loss of the pad

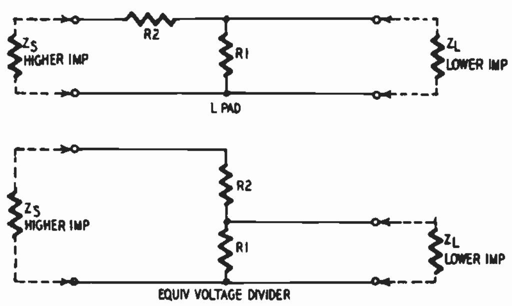


Fig. 604. *The L-pad is a voltage divider.*

be as slight as possible. Obviously, however, there must be some loss involved when any resistance at all is inserted into the circuit. This loss will be minimum when the simple T-network has one of its series arms reduced in value to zero, the configuration then becoming an L, as shown in Fig. 604. When this circuit is slightly redrawn, we see that it is actually a simple voltage divider. Note that in this special case Z_s is taken as the larger of the two impedances, whether it be at the input or output. That is, if the source impedance is the lower of the two, it will be represented by Z_L, and the series arm of the L will always connect to the higher-impedance side. Then under these conditions,

$$k = \sqrt{\frac{Z_s}{Z_L}} + \sqrt{\frac{Z_s}{Z_L} - 1} \qquad (18)$$

To solve for the values of R_1 and R_2 which will provide minimum attenuation, substitute the given values in the T-pad equations previously given. As an example, let us design a minimum-loss pad to match a source impedance Z_S of 250 ohms to a load impedance Z_L of 500 ohms. Using equation (18) to solve for k gives the following result:

$$k = \sqrt{\frac{500}{250}} + \sqrt{\frac{500}{250} - 1} = \sqrt{2} + \sqrt{1} = 1.414 + 1 = 2.414$$

Using equations (15), (16) and (17) again we can now find the

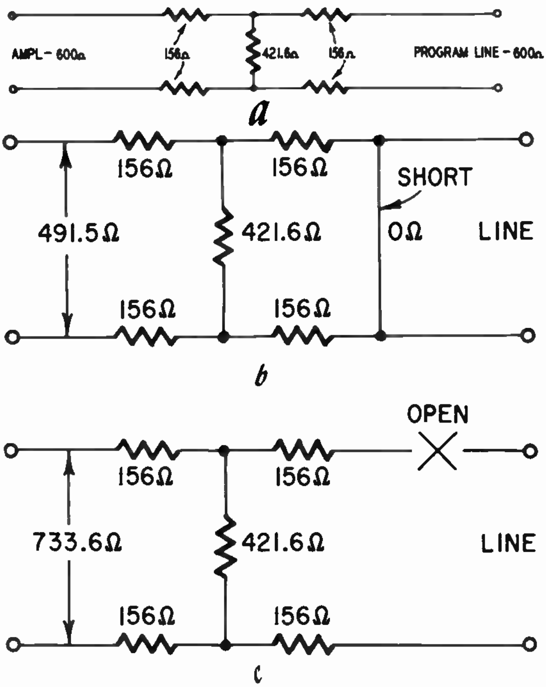


Fig. 605-a, -b, -c. *Attenuators can be used as isolation networks.*

necessary values of R1, R2 and R3. The calculations are given as follows:

$$R1 = 2\sqrt{500 \times 250}\left(\frac{2.414}{4.76}\right) = 355 \text{ ohms}$$

$$R2 = 500\left(\frac{6.76}{4.76}\right) - 355 = 355 \text{ ohms}$$

$$R3 = 250\left(\frac{6.76}{4.76}\right) - 355 = 0 \text{ ohms}$$

These figures are worked out within the limits of slide-rule accuracy. Thus, the results obtained in solving these problems are very close approximations. It all depends upon the number of decimal places you feel like using in your arithmetic. For all practical purposes, the results shown here are considered satisfactory.

The proof that R3 = 0 is simply a double check on the accuracy of the solution. And since this is a voltage divider circuit with a 2-to-1 mismatch, it is logical that the total pad resistance should be center-tapped for the feed to the lower-impedance element, that is, that R1 and R2 should be equal. The final problem is to determine the value of the minimum insertion loss. This calculation is obtained by utilizing equation (14) as follows:

$$\text{db (loss)} = 20 \log k = 20 \times 0.38 = 7.6$$

Isolation pads

Pads are often employed simply as isolation devices to prevent impedance variations in one circuit from reacting adversely upon another. Since equalizers, filters and program lines, for example, are reactive devices, their impedances will vary somewhat with frequency. It is therefore standard practice to insert (following an amplifier which is to work into these devices) a pad of a size adequate to minimize these reactive effects.

To understand the function of an isolation network, let us consider the case of a program amplifier feeding a 600-ohm line, with a 10-db H-pad inserted in the circuit. From the table in Fig. 603 we can construct such a pad (shown in Fig. 605-a). Now let us consider the extreme case of variation in line impedance, from short to open circuit. In this case the impedance would vary from zero to infinity; and if the amplifier were connected directly to the line, it would see this impedance. But see what happens with our pad in the circuit, as illustrated in Fig. 605-b,-c. When the line impedance goes to zero (short circuit), the load on the amplifier goes to 491.5 ohms. When the impedance increases to infinity, the amplifier load goes only to 733.6 ohms. Thus, the amplifier stability will be considerably better under these conditions. In actual practice, of course, extremely wide fluctuations such as these are not encountered and for this reason isolating pads of from 6 to 10 db are ordinarily regarded as entirely adequate.

Splitting and combining networks

In some cases it is desired to feed the same signal into several different circuits simultaneously or to combine several signals into a common load, as in the mixers of control consoles. The simplest way of accomplishing this is by means of series resistors, as shown in Fig. 606-a. Such a network will provide perfect impedance matching for all circuits concerned without frequency discrimination, provided noninductive resistors are employed in its

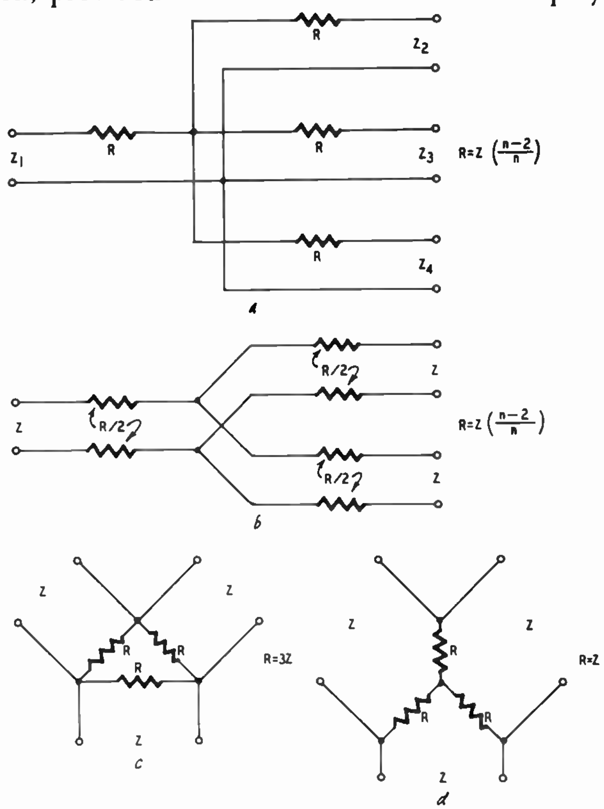


Fig. 606-a, -b, -c, -d. *Attenuators can be used to divide the signal.*

construction. It is not convenient, nor is it ordinarily necessary, to employ unlike impedances with this circuit, which means that $Z_1 = Z_2 = Z_3 = Z_4$. Under these conditions and when n represents the number of circuits, the formula for R is

$$R = Z \left(\frac{n-2}{n}\right) \tag{19}$$

In the illustration (Fig. 606) four circuits were used, so $R = Z \times (2/4) = Z/2$. Thus if the circuits were 600 ohms each, then each resistance in the network would be 300 ohms.

This circuit is useful only with unbalanced lines, but it may be easily modified for balanced operation simply by putting half of R in each leg, as shown in Fig. 606-b. This network will combine any number of circuits, up to the point where the losses become prohibitive, but in this case three circuits are used. Then if Z is once again 600 ohms, then $R = 600 \times (1/3) = 200$ ohms. Since each resistor is R/2, its actual value then will be 100 ohms.

Other forms of combining networks are the delta and star (or Y) configurations shown in Fig. 606-c, -d. These circuits exhibit somewhat greater losses than those previously discussed, but they do provide balanced operation with fewer resistors. In the case of the delta connection, $R = 3Z$, and with the Y connection $R = Z$.

Variable attenuators

It is often necessary to control the volume level in audio circuits, and this is accomplished by means of some sort of variable attenuator. Going back to our original lamp circuit of Fig. 601, the most direct means of varying the voltage across the load is by means of a rheostat, as demonstrated in Fig. 607-a. This arrangement, however, is even worse than when a fixed resistor was used. Neither the source nor the load will see their own impedances as R is varied and, if Z_L should be very high, R must be unduly large to produce adequate attenuation. For these reasons use of the rheostat is confined to electric power circuits.

A somewhat more satisfactory approach to the problem is seen in Fig. 607-b in which a potentiometer has replaced the rheostat. With this arrangement it is possible to obtain full attenuation control without unreasonably large values of R, but Z_L will still not look into a constant source impedance nor will Z_s, unless the ratio Z_L/R is very high. This latter condition occurs when Z_L is the input circuit of a vacuum tube, and this arrangement is most commonly used as an amplifier volume control.

Mixers

The variable attenuators most commonly encountered in professional audio work are called *faders*. These are really groups of fixed resistance pads, any of which may be inserted in the circuit by means of a slider or tapped switch. They are found in mixers used in broadcasting and recording to provide control over a number of program sources, such as microphones and turntables, either simultaneously or sequentially. For example there are occasions, such as the live pickup of a large orchestra, for example—where several microphones may be in use simultaneously. To provide the correct acoustic balance, not only must the microphones be placed correctly, but the amount of signal from each which is used to make up the single composite signal must be minutely controlled. At other times it may be desired to use only a single announce microphone or phono playback channel. Determination as to which channels are to be in operation at any given time rests with the control engineer and his use of faders.

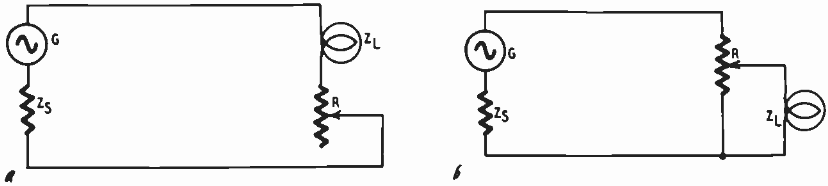


Fig. 607-a,-b. *The variable attenuator at the left is not as suitable as that shown at the right.*

The mixer circuit must be such that each channel operates independently of all the others, without any crossing effects whatever. Still, all of the channels must ultimately combine at a common point for further amplification and transmission of the composite signal. The fader must also have uniform frequency characteristics throughout the audible range and an exceedingly low noise level. A well-designed and well-maintained fader will exhibit a noise level of −120 dbm or even less. Shield cans are usually placed around the attenuators to protect against dirt as well as electrostatic and electromagnetic fields. The circuits of typical attenuating structures are illustrated in Fig. 608. Each of these types has a direct predecessor in the fixed pads of Fig. 603.

Generally speaking, mixer circuits employed in public-address systems are of the high-impedance type, while those used in broadcasting and recording are low-impedance. The former can be used with high-impedance microphones and pickups working directly into the amplifier input without a transformer. Such a circuit is shown in Fig. 609-a. This employs a dual triode, such as the 6C8–G or 6SN7–GT. It is excellent as it stands for a two-position mixer, and more channels can readily be added by the simple expedient of using more tubes. Since each of the inputs is totally isolated from all of the others, there is no crosstalk or other interaction and because the plates are in parallel, each section sees a load impedance consisting of the load resistance shunted by the other section's plate resistance (R_p). The load resistance that one section sees is always less than R_p, and the gain of each section must in every case be less than $\mu/2$. Similarly, for a three-channel mixer the gain will be less than $\mu/3$, for four channels less than $\mu/4$, and so forth.

Fig. 609-b illustrates a system which works fairly well for almost any number of positions when the series resistors have the same

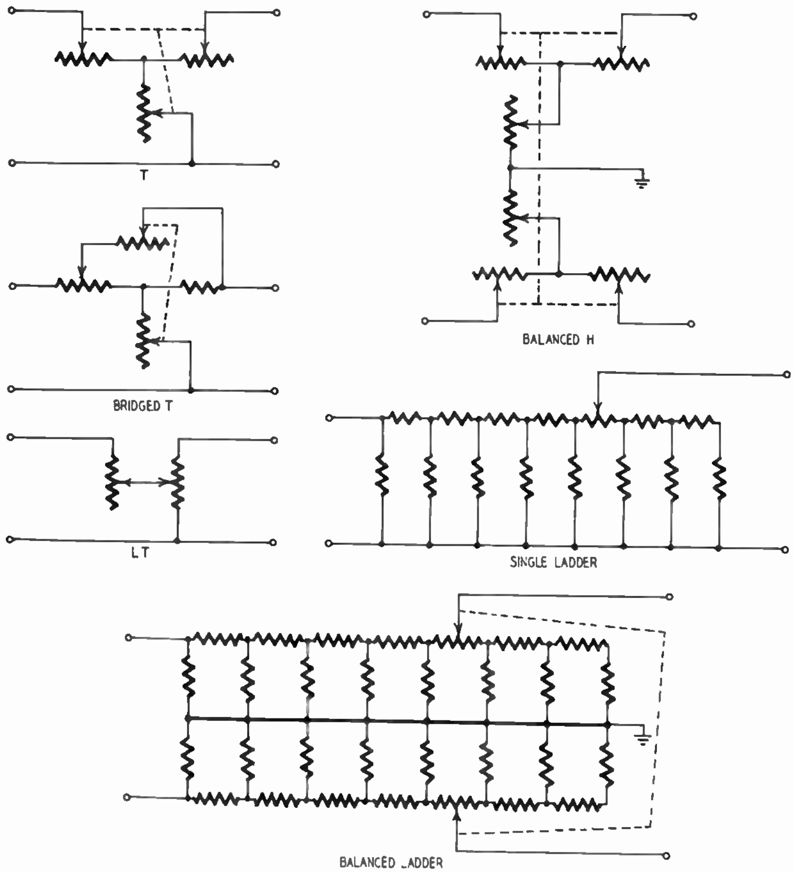


Fig. 608. *Typical attenuating networks are shown above.*

value as the potentiometers. There will be some interaction between channels with this arrangement, however, with the insertion loss in each channel varying by several db. This system is used only where continuous and precise control is not required.

Since the mixers just described make no effort to maintain a constant impedance, nor to match impedance, they are unsatisfactory for critical work. Low-impedance mixers which meet these requirements may be further subdivided into *low-level* and *high-level* systems. The mixers studied up to now are low-level systems in that the output of the source device was fed directly into the attenuator. In high-level mixing, the source voltage is always fed into a preamplifier before entering the mixer channel. This provides a much improved signal-to-noise ratio and is used exclusively in broadcasting and recording.

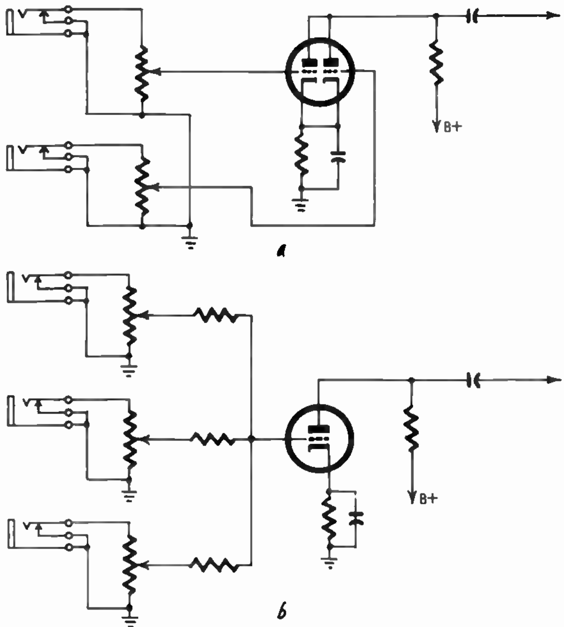


Fig. 609-a, -b. *Attenuators can be used for proper mixing of signals.*

Mixer circuits

Fig. 610 shows the three basic connections for attenuators in mixer circuits: series, parallel and series-parallel. Each has its characteristic advantages and disadvantages. All other factors being equal, the parallel mixer (Fig. 610-a) is unable to maintain a very constant impedance throughout the operating range of its several controls. The series mixer (Fig. 610-b), while it is more stable in this regard, cannot have its faders grounded and when in this "floating" condition they are more subject to induced noise as well as leakage at the high frequencies. An attempt at a compromise, with some of the best features of both types, is the series-parallel mixer illustrated in Fig. 610-c.

Note that each type of circuit employs resistors R_B in series with each attenuator and also in the common output circuit. Known as *building-out* resistors, they serve a dual purpose. Their values are chosen so that they maintain the correct impedance relationships between the several sources and the common load. Also they serve to provide a degree of isolation between faders so as to minimize crosstalk and other interaction.

Filters

In audio work it is sometimes desirable to alter the frequency transmission characteristics of a system for some special applica-

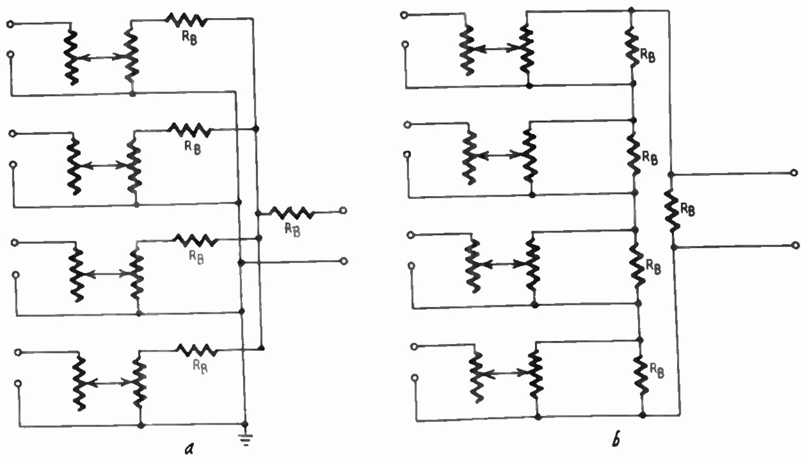


Fig. 610-a, -b. *Circuit arrangements for mixer type attenuators.*

tion. One device employed for this purpose is the *filter*, which may be any one of four basic types:

(1) The *low-pass* filter, which readily transmits all frequencies below a certain value, but which attenuates all frequencies above this value.

(2) The *high-pass* filter, which readily transmits all frequencies above a certain value, but which attenuates all frequencies below this value.

(3) The *band-pass* filter, which readily transmits all frequencies between two values, but which attenuates all frequencies above and below this pass band.

(4) The *band-elimination* filter, which attenuates all frequencies between two values, but which readily transmits all frequencies above and below this elimination band.

In every case, the frequency at the junction between the transmission and attenuation bands is known as the *cutoff* frequency.

Filters consist of networks of reactive elements (capacitance and inductance) connected in configurations similar to those for the attenuators already discussed. We begin with simple combinations, known as *sections,* which may be connected in tandem when more complicated structures are required. In less stringent applications the entire filter may comprise only a single section, while more severe requirements may dictate a multisection type. Design and analysis under this method are greatly simplified, since the loss of each individual section may be considered alone, and the total insertion loss of the structure then simply becomes the sum of the individual losses. We may then consider the sections as the building blocks of filter design.

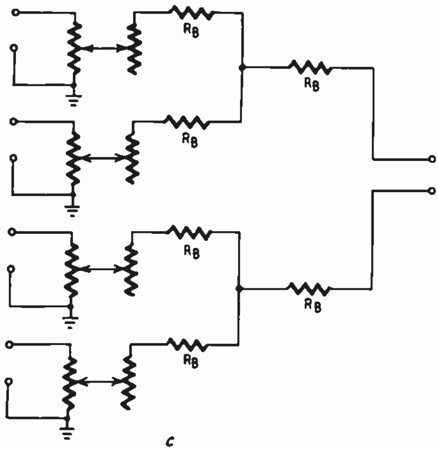


Fig. 610-c. *Series-parallel mixer circuit.*

The basic element of all filter design is the simple L section shown in Fig. 611-a, with the conventional symbol notation identifying the several components. The impedance elements, of course, are actually reactive but, since their form will depend upon the filter application, the simple resistance symbol is employed to represent each impedance. The way in which the basic L section (sometimes called the half-section) can be combined to form T and π sections can be understood by reference to Fig. 611-b -c. By the correct use of these three basic sections, any filter ordinarily required in audio work can be constructed.

Filter characteristics

The self-impedances of these filter sections are designated by the symbols Z_I and $Z_{I'}$. The Z_I impedance of the L section is referred to as its *mid-series* impedance, while the $Z_{I'}$ value is called the *mid-shunt* impedance. When L sections are combined to form a T or π as in Fig. 611-b-c, each half-section provides the correct matching impedance for the other. In the case of the T section, the mid-shunt impedance has disappeared in the combination, and the network is now symmetrical with its input and output impedances both equal to Z_I. In the π section, Z_I disappears and the unit becomes symmetrical with both image impedances equal to Z_I'.

To determine the values of Z_A and Z_B, it is necessary, of course, to have given the value of f_c, the cutoff frequency. But, to understand fully the theory of filter design, let us first assume values for

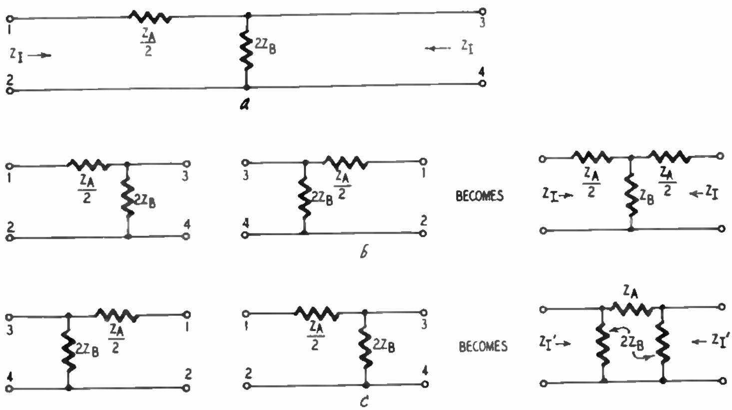


Fig. 611-a, -b, -c. *The L-pad can be used to build a T- or pi attenuator.*

Z_A and Z_B, and then determine the behavior of various frequencies as they are fed into the network.

We recall from an earlier chapter that the attenuation loss in any circuit expressed in decibels is equal to 10 log (P_1/P_2). This assumes that the system is operating under matched-impedance conditions which are mandatory for correct filter operation. Then in the region where the filter is to be freely transmitting the attenuation is ideally zero, which means that the power output equals the power input.

Now let us consider what happens when we apply a signal of variable frequency to the input terminals of the filter. As the frequency is varied, the values of Z_A and Z_B will likewise vary, for they are reactive components whose impedance is a function of the frequency. Furthermore, they will change inversely, the inductive reactance characteristically increasing with a higher frequency while the capacitive reactance simultaneously decreases. Under these conditions the ratio of Z_A to Z_B will vary also. Now as we change the frequency throughout the spectrum, we find that no attenuation will result while the ratio $Z_A/4Z_B$ will become negative—that is, it will be be between 0 and −1, but that outside this region there will be attenuation. (The figure of −1 is possible because the reactances combine vectorially.) This relationship is fundamental to all filter design.

The simplest and most common type of filter section is that in which impedances Z_A and Z_B have an inverse relationship; that is, their product is a constant. The usual mathematical equation given for this relationship is:

$$k^2 = Z_A \times Z_B \tag{20}$$

From this formula is derived the name of this type of filter—the *constant-k* network. Under these conditions the values of Z_I and $Z_{I'}$ are determined from the following equations:

$$Z_I = k \sqrt{1 + \left(\frac{Z_A}{2k}\right)^2} \tag{21}$$

$$Z_{I'} = \frac{k}{\sqrt{1 + \left(\frac{Z_A}{2k}\right)^2}} \tag{22}$$

From equations (21) and (22) you will note that when the frequency is such that the quantity $Z_A/2k$ is small, then Z_I and $Z_{I'}$ are almost purely resistive and of a value nearly equal to k.

Filter design

The simplest low-pass filter will be a single constant-k L section, with an inductance in the series arm and a capacitance in the shunt arm, the same basic network we employed in Chapter 4 when designing a power supply filter. The same principles are involved here: the inductance is highly reactive at the upper frequencies and thus will tend to oppose them, while the capacitance will have very low reactance at the high end and will therefore tend to act as a short circuit to these frequencies. In this case

$$Z_A = X_L = 2\pi fL \qquad\qquad Z_B = X_c = \frac{1}{2\pi fC}$$

From this we can now say:

$$Z_A \times Z_B = X_L \times X_c = \frac{2\pi fL}{2\pi fC} = \frac{L}{C}$$

However, we already know that $Z_A \times Z_B = k^2$ so we can substitute k^2 in the last equation with the result that:

$$k^2 = \frac{L}{C} \tag{23}$$

At the cutoff frequency $Z_A/4Z_B = 2\pi f_c L/(4/2\pi f_c)$. Also, as previously discussed $Z_A/4Z_B$ can equal -1. From this relationship we now get the following equation:

$$-\frac{2\pi f_c L}{\frac{(4/2\pi fC)}{2\pi fC}} = -1 \qquad\qquad \frac{(2\pi f_c)^2 LC}{4} = 1$$

$$(2\pi f_c)^2 LC = 4 \qquad 2\pi f_c \sqrt{LC} = 2 \qquad 2\pi f_c = \frac{2}{\sqrt{LC}}$$

$$f_c = \frac{1}{\pi\sqrt{LC}} \tag{24}$$

(Since we have learned that the quantity $Z_A/4Z_B$ can lie between 0 and -1, note that another value of f_c can also be obtained. But f_c is then also zero and of no practical significance.) Now that we have the relationship between L, C, f_c and k^2, we can solve for the unknowns L and C. Solving for L, we get:

$$\sqrt{L} = \frac{1}{\pi f_c \sqrt{C}} \qquad\qquad L = \frac{\sqrt{L}}{\pi f_c \sqrt{C}}$$

However, since $\sqrt{L}/\sqrt{C} = k$ or $L/C = k^2$ (equation 23), the value of L is given as:

$$L = \frac{k}{\pi f_c} \tag{25}$$

Solving for C, we get:

$$\sqrt{C} = \frac{1}{\pi f_c \sqrt{L}} \qquad\qquad C = \frac{\sqrt{C}}{\pi f_c \sqrt{L}}$$

Referring once again to equation (23), let us now utilize its reciprocal, that is, $\sqrt{C}/\sqrt{L} = 1/k$ or $C/L = 1/k^2$. From this relationship the value of C is now found to be:

$$C = \frac{1}{\pi f_c k} \qquad (26)$$

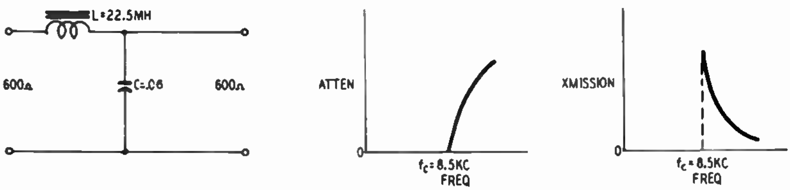


Fig. 612. *Single constant-k L section showing attenuation and transmission characteristics.*

Since the filter must work into and out of its own image impedance, k is assigned the value of the impedance of the circuit in which the filter is working. In actual practice both the source and the load will usually be around 500 or 600 ohms. Fig. 612 shows a typical filter of this type and its attenuation and transmission characteristics.

Let us determine how this network was designed. The problem was to construct an L section constant-k low-pass filter having a cutoff frequency of 8,500 c.p.s. and working between impedances of 600 ohms. Substituting these values in equations (25) and (26), we can now determine the actual values of L and C, respectively, as follows:

$$L = \frac{600}{3.14 \times 8{,}500} = 22.5 \text{ mh}$$

$$C = \frac{1}{3.14 \times 8{,}500 \times 600} = 0.062\ \mu f$$

Fig. 613 shows the constant-k high-pass filter network, along with its characteristics. Employing the same mathematical develop-

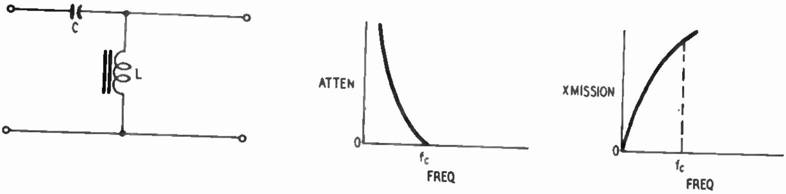


Fig. 613. *Single constant-k high-pass filter network.*

ment for this as we used for the low-pass filter, the values of L and C are found to be:

$$L = \frac{k}{4\pi f_c} \qquad (27)$$

$$C = \frac{1}{4\pi f_c k} \qquad (28)$$

If we combine a capacitor in combination with an inductance in the series arm, we will have a circuit which will be resonant at some frequency. Similarly, if we also connect a coil and capacitor

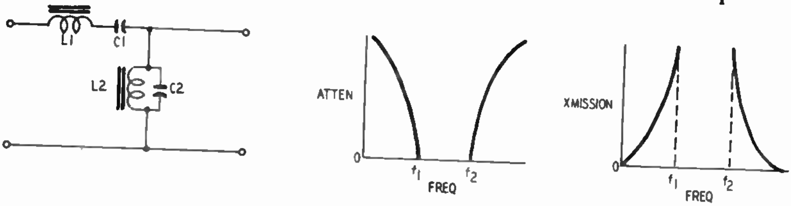


Fig. 614. *Bandpass filter contains series and parallel (fixed-tuned) circuits.*

in parallel in the shunt arm, we will have a circuit which will be anti-resonant at some frequency. What we have actually done here is to combine a high-pass and a low-pass filter so as to form a band-pass arrangement, as shown in Fig. 614. We now have two cutoff frequencies to consider, and these values, along with the constant-k, are employed to calculate the values of the components as follows:

$$L1 = \frac{k}{\pi(f_2 - f_1)} \qquad (29)$$

$$L2 = \frac{k\,(f_2 - f_1)}{4\pi\, f_1\, f_2} \qquad (30)$$

$$C1 = \frac{f_2 - f_1}{4\pi\, f_1\, f_2\, k} \qquad (31)$$

$$C2 = \frac{1}{\pi k\,(f_2 - f_1)} \qquad (32)$$

wherein f_1 is the lower cutoff frequency and f_2 the upper cutoff frequency.

When we invert the two arms of this structure, so that the series

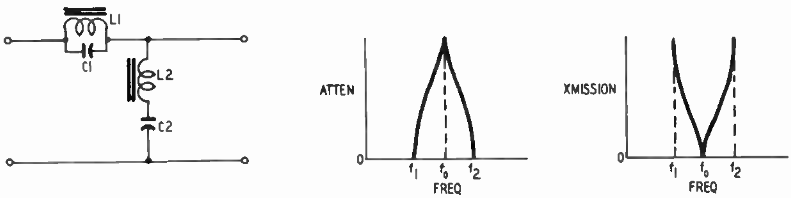


Fig. 615. *Band elimination filter and its characteristics.*

arm is anti-resonant and the shunt arm resonant, we have the band-elimination filter shown in Fig. 615. Formulas for the calculation of these circuit constants are:

$$L1 = \frac{k\,(f_2 - f_1)}{\pi\, f_1\, f_2} \qquad (33)$$

$$L2 = \frac{k}{4\pi\,(f_2 - f_1)} \qquad (34)$$

$$C1 = \frac{1}{4\pi\, k\,(f_2 - f_1)} \qquad (35)$$

$$C2 = \frac{f_2 - f_1}{\pi\, f_1\, f_2\, k} \qquad (36)$$

M-derived filters

We implied earlier that there are applications where the characteristics of the constant-k are regarded as inadequate. Reference to the characteristic curves of Figs. 612, 613, 614 and 615 will show why. Although f_c is designated as the cutoff frequency, strictly speaking there is no true cutoff at this point. The shape of this curve could perhaps be better described as a tapering off of the transmission. When sharper cutoff is required, a refinement of the constant-k network is used, known as the *m-derived* section. This has additional impedance elements in either the shunt or the series arms, resulting in an attenuation which is infinite at some frequency beyond cutoff.

Referring to our basic constant-k L section of Fig. 611, suppose we were to connect a simple switch across output terminals 3 and 4. When the switch is closed and the terminals short-circuited, we will designate the impedance appearing across the input terminals 1 and 2 as Z_{sc}. Similarly, when the switch is opened, the impedance at the input will be called Z_{oc}. Now the mid-series image impedance of the constant-k half-section will be

$$Z_1 = \sqrt{Z_{sc} \times Z_{oc}} \qquad (37)$$

Now suppose that we construct a new L section in which the short-circuit impedance is Z_{sc} times a certain constant m and the open-circuit impedance is Z_{oc} times $1/m$. Then in accordance with the formula just developed:

$$Z_1 = \sqrt{mZ_{sc}\left(\frac{Z_{oc}}{m}\right)} \qquad (38)$$

From this last equation it is apparent that m and $1/m$ will cancel with the result that the Z_1 formula remains the same. This means that the new section will have the same mid-series impedance as the constant-k network whence it was derived. It will also have the same cutoff frequency, but the mid-shunt impedance and the attenuation characteristics will be different.

Referring once again to the basic L of Fig. 611-a, we can see that the following relationships exist:

$$Z_{sc} = \frac{Z_A}{2}$$

$$Z_{oc} = \frac{Z_A}{2} + 2Z_B$$

Then the m-derived section will have these relationships:

$$Z_{scm} = \frac{mZ_A}{2} \qquad Z_{ocm} = \frac{1}{m\left(\frac{Z_A}{2} + 2Z_B\right)}$$

The series arm of the new section will be equivalent to Z_{scm} or $mZ_A/2$. Then the impedanbe of the elements of the shunt arm

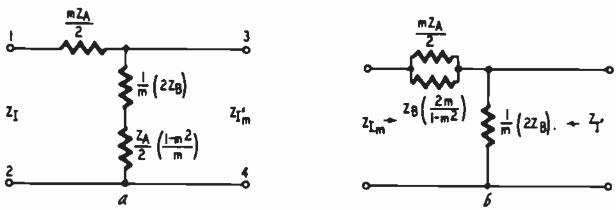


Fig. 616-a, -b. *Basic series m-derived half section.*

can be found simply by subtracting Z_{scm} from Z_{ocm}, in which case the shunt arm equals

$$\frac{1}{m}(2Z_B) + \frac{Z_A}{2}\left(\frac{1-m^2}{m}\right)$$

We now have all of the facts required to develop the basic series m-derived half-section shown in Fig. 616-a. The network is obviously more complicated than the constant-k prototype in that it requires two impedances in its series arm. As we have said, the mid-shunt impedance also differs from that of the constant-k, and this new impedance is conventionally designated as $Z_{Im'}$.

It is possible, however, to have an m-derived filter whose mid-shunt impedance is the same as that of the constant-k configuration. In this case the mid-series impedances will differ, and the network is called a shunt m-derived section. Starting with the basic proposition that $Z_I = Z_{sc} \times Z_{oc}$, the mathematical development is just as tedious as it was for the series-derived section, and so we won't go through that again. The results of such computations, however, may be found in Fig. 616-b. Since the mid-series impedance differs from that of the constant-k, it is now designated as Z_{Im}. It is interesting to note here that an inverse relationship exists between Z_{Im} and $Z_{Im'}$; that is, their product is equal to k^2.

One of our definitions of the m-derived filter was that it has a definite frequency of infinite attenuation. The mathematical relationship between this frequency f_{∞}, the cutoff frequency f_c and the constant m, is expressed for the low-pass network as:

$$f_{\infty} = \frac{f_c}{\sqrt{1 - m^2}} \tag{39}$$

For the high-pass network the relationship becomes:

$$f_{\infty} = f_c \sqrt{1 - m^2} \tag{40}$$

Circuit diagrams and formulas for all the basic m-derived sections are given in Figs. 617 through 620. You will see that *all of the*

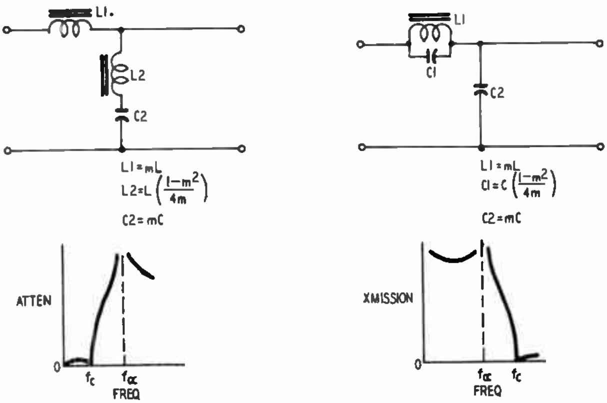


Fig. 617. *M-derived half section. M is usually taken as 0.6.*

calculations are based upon the formulas already given for the constant-k prototypes.

M-derived filter characteristics

The shape of the characteristic curves of the m-derived section, and hence the sharpness of cutoff, is determined by the value of m, which must lie somewhere between 0 and 1. The sharpness of cutoff increases as m approaches 0, and it is obvious from the design equations that, when m = 1, the circuit becomes a constant-k section once again.

For the correct impedance relationships, the ideal filter would be one having throughout the frequency range a self-impedance always equal to k. But no such filter has yet been devised, and all types exhibit some change in impedance with varying frequency. The best possible condition of impedance stability is obtained when the m-derived filter is used and when m = 0.6.

Filters may comprise any number of sections from a single L network up to perhaps a half-dozen full sections. When sections of the same type are connected in tandem, the filter is said to be *uniform*. When dissimilar types are used together, the unit is called a *composite*. Multisection filters are employed to obtain a desired characteristic, the amount of attenuation in the rejected

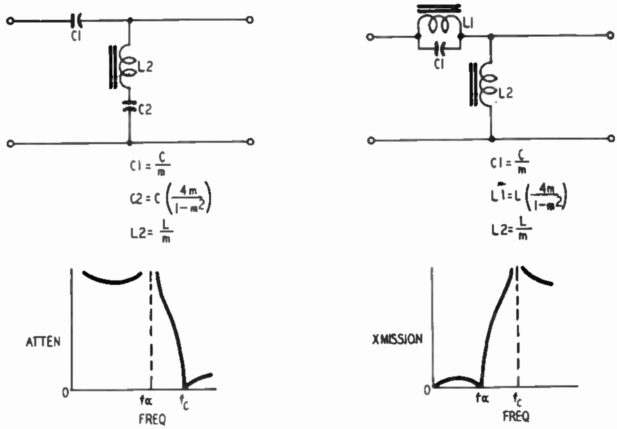


Fig. 618. *Here we have still another m-derived half section.*

bands being determined by the number of sections and the actual shape of the transmission curve depending upon the types of those sections. Reference to the characteristic curves will show that the m-derived section has a sharper cutoff but less attenuation beyond cutoff, while the constant-k unit has a more gradual cutoff but increasing attenuation beyond that point. Thus a composite filter comprising both types will afford the most desirable characteristics of each. In such cases it is customary to employ m-derived sections as both the input and output terminations, because of their more desirable impedance characteristics, with the constant-k sections sandwiched in between.

From examination of the characteristic curves, you can see that a band-pass arrangement could be obtained simply by using two separate filters in series. As long as the cutoff of the low-pass filter is at least twice the frequency of the high-pass unit, the band between these two cutoffs will be freely transmitted. In situations where the filter is not regarded as a permanent part of the installation, this arrangement is preferable because of its flexibility.

The need for equalizers

In the perfect audio system, filters and equalizers would not only be totally unnecessary, they would be quite undesirable. The perfect system exists only in theory. Even the human ear, as we learned in Chapter 1, fails to measure up to this theoretical perfection, for its response varys both with frequency and with ampli-

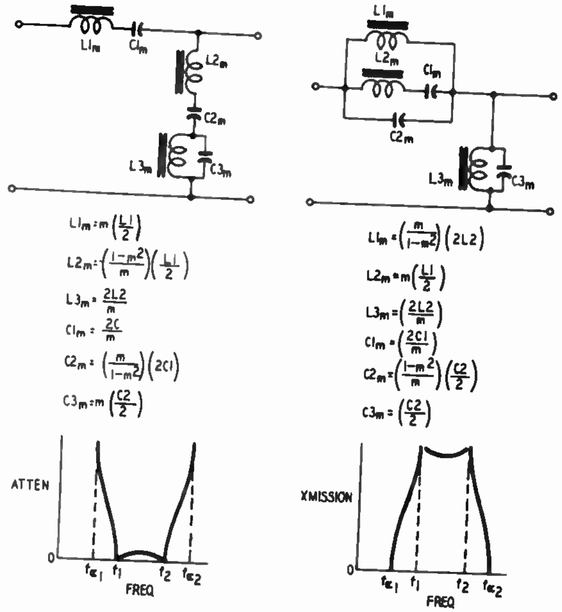


Fig. 619. *M-derived half section used as a bandpass filter.*

tude. Recorded and broadcast signals, too, are often short of perfection, due to extraneous noises or poor acoustic balance. Often the transmitting equipment is incapable of passing even the ideal signal without some alteration. And even the performing artist, alas, will sometimes err and expect engineering magic to make up the deficit.

Some of these shortcomings are corrected by the use of volume limiting, compression and expansion, which were discussed in Chapter 3. In a general way, what these devices do for the dynamics, filters and equalizers do for the frequency range.

In disc recording, the frequency response at the upper end of the spectrum is deliberately accentuated and the bass region attenuated by a fixed amount, giving a transmission curve known as the *recording characteristic*. The high-frequency "tip up" provides a better signal-to-noise ratio in that region and the bass "roll off" prevents overcutting between adjacent grooves on the powerful low-frequency passages. Alteration of the response curve of this sort is known as *pre-equalization,* since it occurs at the time the record is made. To play back such a record correctly, it is necessary to provide *post-equalization* at the reproducing system, precisely

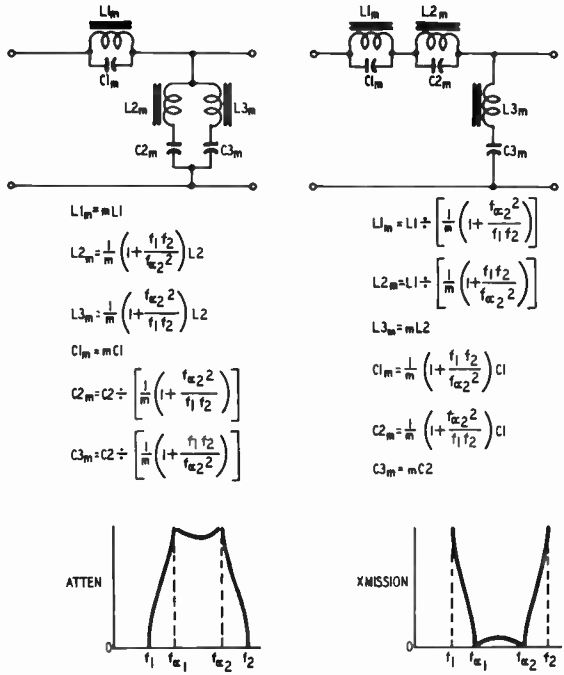


Fig. 620. *M-derived half section used as a band elimination filter.*

complementary to that of the recording characteristic. That is, where the recording is tipped up the reproducer must roll off by exactly the same amount and in the rolled-off bass region the reproducer must boost accordingly.

Similarly, pre-emphasis and de-emphasis are included in all magnetic tape recorders. Since each playback system is equalized to match its own recording characteristic, no problems are encountered nor are any adjustments necessary as long as a tape is reproduced on the same sort of machine on which it was made. But if one attempts to play back a tape on a machine of different make from that on which it was recorded, the results may not always be entirely satisfactory. Standardization of characteristics in tape recorders, and in phonograph records as well, is not all that it might be.

The ordinary tone control, common to most radios and record players, is also a simple form of equalizer. It is used here to correct for variations in individual taste and hearing and for the acoustical conditions surrounding the reproduction. Fixed pre-emphasis and de-emphasis are standard practice in FM broadcasting. While the FM system inherently discriminates against extraneous noise, this is still not sufficient to provide perfectly noise-free operation up to 15,000 c.p.s. All FM broadcast transmissions, therefore, are tipped up at the high end in accordance with the standard characteristic shown in Fig. 621, and each receiver has a built-in de-emphasizer which precisely complements this curve.

All telephone lines employed for radio program transmission will, if not equalized, deliver at their remote end a signal deficient in the high frequencies. The line has a certain amount of distributed capacitance which varies directly with the length of the circuit. Since the reactance of this distributed capacitance will decrease with increasing frequency, it will tend to act as a short circuit for the upper end of the spectrum. This defect is equalized by the insertion of losses at the low frequencies in direct proportion to those at the high end. This equalizer is installed across the line at the receiving end and so adjusted that the frequency characteristics at the output of the equalizer are perfectly flat throughout the desired audio range.

When picking up dialogue on a television or motion-picture stage with a flat audio system, the speech will seem unnatural and in extreme cases even inarticulate. This is due to the fact that the acoustics of the average sound stage actually seem to accentuate the lower frequencies. A more accurate explanation of this phenomenon lies in the fact that the materials of which the set is constructed usually have less absorption at the lower frequencies. This means that the important speech intelligibility range of 500 to 5,000 c.p.s. is somewhat attenuated, while the lower range remains strong and introduces a degree of masking. When the sound is reproduced, it is usually played back at a much higher level than originally spoken, which further aggravates the situation. This is especially true in theaters and is due to the varying response of the ear, which is much more sensitive to the bass frequencies at higher levels. Furthermore, the very high frequency components of speech add little to the intelligibility but often sound harsh and unpleasant to the ear. It is often customary to roll off the high end slightly as well. This, then, is dialogue equalization, using a response characteristic somewhat rolled off at both the high and low ends so that the intelligibility region is in effect boosted.

In the recording on film of the master sound track, which contains music and sound effects as well as dialogue, it is customary to pre-emphasize the high end for better signal-to-noise ratio and improved high-frequency reproduction. This is necessary to overcome the various losses in recording and reproduction, including recorder slit loss, laboratory processing and printing losses, pro-

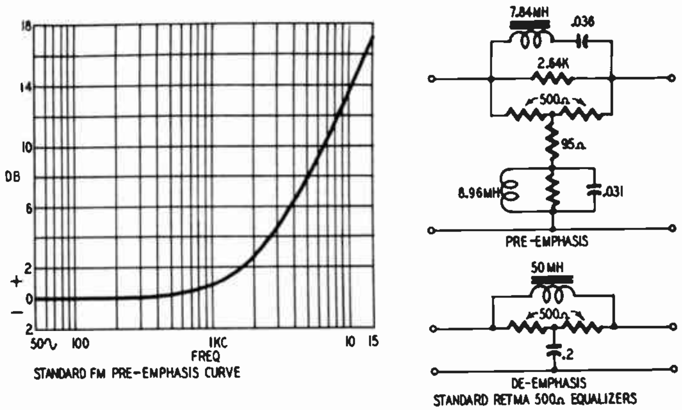


Fig. 621. *Typical pre-emphasis and de-emphasis networks. The pre-emphasis curve is used for FM broadcasting.*

jector optical-system losses, and high-frequency sound absorption by the motion-picture screen and theater acoustics.

In all of the audio production arts today it is quite common to employ elaborate variable equalizers for correction and special effects. Some of these devices accentuate or attenuate anywhere in the audible spectrum to provide any shape response curve desired. In this way it is possible, within limits, to provide much better acoustic balance between various sections of an orchestra and even to add presence and "pull out" a vocal solo previously covered over by orchestral accompaniment.

Types of equalizers

From the foregoing it should be quite obvious that the ideally flat frequency response is not always so ideal at all. The importance of equalizers can therefore hardly be exaggerated, and for this reason it will be well to discuss briefly their basic design principles. In practice an equalizer is usually designed empirically by reference to an elaborate set of charts and curves which are beyond the scope of this volume. We can, however, show the various types of circuits employed and attempt to work out a typical design problem.

Innumerable circuit configurations are possible in the design of attenuation equalizers, but the great majority will fall into one of the following basic categories:

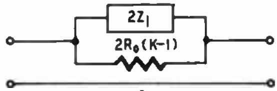
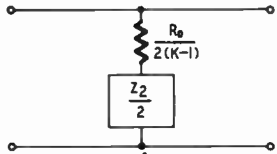
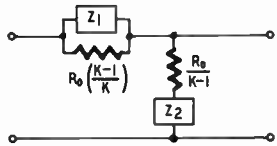
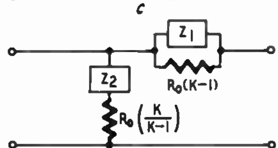
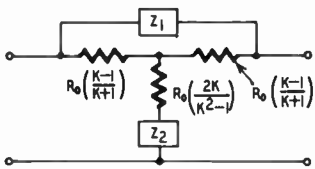
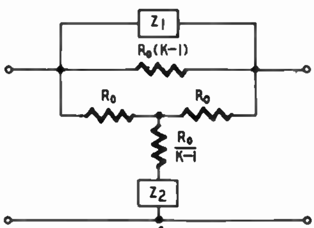
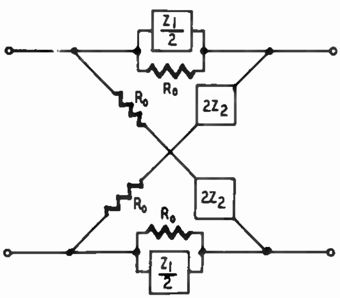
1. Series impedance
2. Shunt impedance
3. Full series
4. Full shunt
5. T
6. Bridged-T
7. Lattice

These seven basic equalizer circuits, together with some of their important characteristics, are illustrated in Fig. 622. One characteristic common to all of them is the inverse relationship between impedances Z_1 and Z_2. In each case $Z_1Z_2 = R_0$. It is desirable that the proper impedance relationships be maintained between the source, the load and the input and output terminals of the equalizer. But we note from the illustration that this does not occur with the first four types, and therefore their usefulness is quite limited. The simplest of the equalizers which does maintain this constant-impedance relationship regardless of frequency is the T type.

Fig. 622 also shows us that the T equalizer structure is basically a T-pad or attenuation network, with impedance elements in parallel with its series arm and in series with its shunt arm. The form of these impedance elements will determine the attenuation characteristics of the equalizer, and the basic T structure can therefore be further subdivided into the four types of Fig. 623. Now let us see how we arrived at the characteristic curves shown.

Referring to the basic T configuration (Fig. 622-e) it is apparent that the impedance values of Z_1 and Z_2 will vary with frequency and inversely with each other. That is, if Z_1 increases with frequency, Z_2 must simultaneously decrease to satisfy the condition $Z_1Z_2 = R_0$. Now consider the situation when Z_1 increases to infinity and Z_2 simultaneously decreases to zero. Since Z_1 is then effectively an open circuit, all that remains in the series arm of the circuit are the two T-pad resistors. And since Z_2 is now in effect a short circuit, all that remains in the shunt arm is the single T resistor. Under these conditions the impedance elements are there-

Fig. 622. *Here we have seven basic equalizer circuits and some of their characteristics.*

TYPE	NETWORK	INPUT IMP	OUTPUT IMP	INSERTION LOSS
SERIES IMP	 *a*	NOT CONSTANT	NOT CONSTANT	$20\log\frac{R_0+Z_1}{R_0}$
SHUNT IMP	 *b*	" "	" "	$20\log\frac{R_0+Z_2}{Z_2}$
FULL SERIES	 *c*	R_0	" "	$20\log\frac{R_0+Z_1}{R_0}$
FULL SHUNT	 *d*	"	" "	$20\log\frac{R_0+Z_1}{R_0}$
T	 *e*	"	R_0	$20\log\frac{R_0+Z_1}{R_0}$
BRIDGED T	 *f*	"	"	$20\log\frac{R_0+Z_1}{R_0}$
LATTICE	 *g*	"	"	$20\log\frac{R_0+Z_1}{R_0}$

fore inconsequential and the circuit then becomes nothing more than a simple T attenuator. At the other extreme, when Z_1 decreases to zero and Z_2 increases to infinity, there is effectively a short circuit in the series arm and an open in the shunt arm. Under these conditions the circuit simplifies down to nothing more than the two conductors. From this we can generalize that this equalizer circuit introduces a loss which varies from zero to the characteristic attenuation of the pad alone.

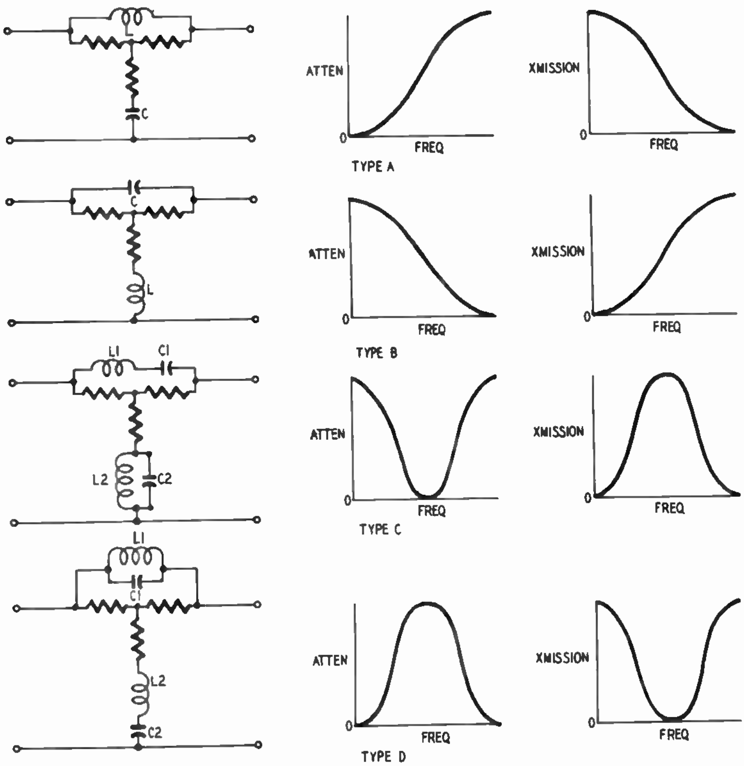


Fig. 623. *Equalizer networks and their attenuation and transmission characteristics.*

If we make Z_1 a pure inductance and Z_2 a pure capacitance, at zero frequency Z_1 will be zero and Z_2 infinite, while at infinity c.p.s. Z_1 will be infinite and Z_2 zero. These are the conditions shown for the type A network of Fig. 623. When we exchange these two impedances, making Z_1 capacitive and Z_2 inductive, quite the opposite characteristics are obtained. Then at 0 c.p.s $Z_1 = \infty$ and $Z_2 = 0$, while at $\propto$ c.p.s. $Z_1 = 0$ and $Z_2 = \infty$. This characteristic is typical of the type B equalizer of Fig. 623.

If Z_1 is designed as a resonant circuit, then Z_2 must be anti-resonant to satisfy the constant inverse condition of $Z_1Z_2 = R_0$. Similarly, if Z_2 is resonant, then Z_1 must be anti-resonant. In either case it is obvious that both circuits must resonate at the same frequency. Networks of this type will have points of maximum or minimum attenuation somewhere within the spectrum, rather than at zero or infinite frequency as in types A and B. These configurations and their characteristics are shown as types C and D of Fig. 623.

Equalizer design

In the design of T type equalizers, then, the impedance element Z_1 can take any of four forms: capacitance, inductance, a parallel-resonant circuit or a series-resonant circuit. Having determined the form of Z_1, the character of Z_2 will be automatically dictated by the Z_1Z_2 inverse relationship. When these impedances are purely reactive, the impedance Z_1 may be expressed as an equation:

$$Z_1 = R_0 \sqrt{\frac{c^{0.23a} - 1}{\left(\frac{C+1}{2}\right)^2 - \left[\left(\frac{C-1}{2}\right)^2 c^{0.23a}\right]}} \qquad (41)$$

wherein a is the attenuation loss of the equalizer in decibels. Now assuming a number of different values for a, a set of curves may be developed as in Fig. 624, in which R_0 was taken as 600 ohms. It is now possible to use these curves as the basis for a set of calculations in a typical equalizer design problem.

Suppose we have a recording system essentially flat except for an undesirable 4-db peak at 3,000 c.p.s., with the peak dropping to half (2 db) an octave away at 1,500 and 6,000 c.p.s. The system response in this region would then appear as in Fig. 625. Now, to flatten this peak we will require an equalizer which has a transmission characteristic the direct complement of this one and an attenuation characteristic with the precise shape of the curve of Fig. 625. Then from Fig. 623 we readily determine that the filter answering these requirements will be of type D. We also must know the operating impedance, and we will assume that in this application we find the most convenient operating point for the equalizer to be in a 200-ohm circuit.

To begin we see that the maximum loss required of the network is 4 db, which means that we must first design a 4-db T (or

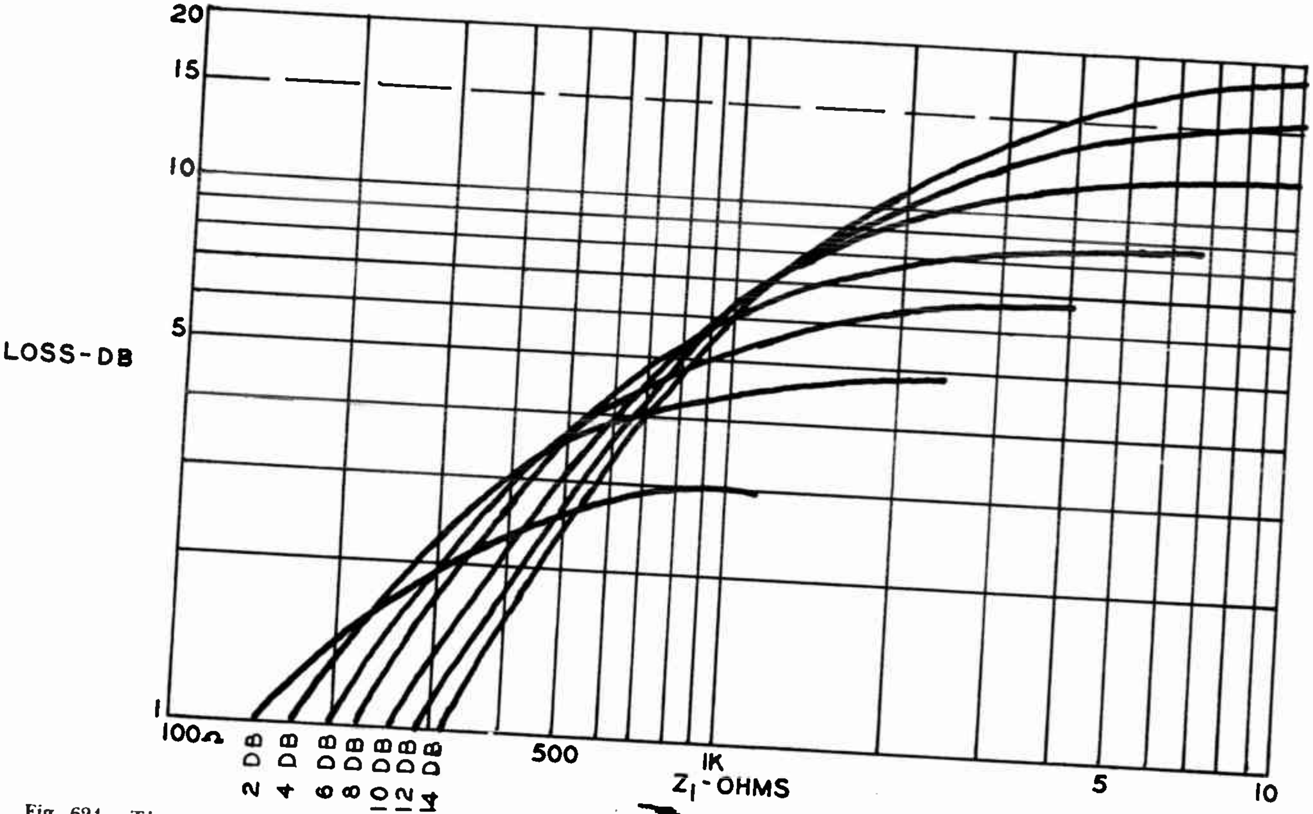


Fig. 624. *The curves shown in the illustration above can be used to aid in the solution of equalizer design problems.*

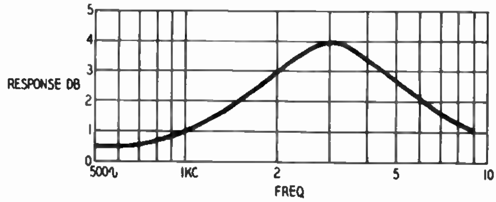
bridged-T) pad. Referring to the chart of Fig. 603, we find that these figures are all predicated upon an operating impedance of 600 ohms. We can still adapt them to our purposes, however, simply by multiplying the given values by $R_0/600 = 1/3$. Then for a 4-db T-pad at 200 ohms:

$$R1 = 135.8 \times 1/3 = 45.27 \text{ ohms}$$

$$R2 = 1{,}249 \times 1/3 = 416.3 \text{ ohms}$$

Now since the impedance elements are effectively inoperative at the resonance frequency of 3,000 c.p.s., only the pad remains in

Fig. 625. *The curve shown in the illustration at the right has a peak at about 3,000 cycles. The peak can be flattened by an equalizer whose response curve is the complement of this one.*



the circuit. We will proceed to determine the values of the impedances at some other frequency, such as one of the half-loss frequencies of 1,500 or 6,000 c.p.s. At these points the attenuation loss must be 2 db. The appropriate curve of Fig. 624 shows us that, for a 4-db 600-ohm pad, a loss of 2 db is obtained when Z_1 is about 270 ohms. Since ours is a 200-ohm circuit, we must apply the correction factor of 200/600, or 1/3, in which case Z_1 becomes 90 ohms. Two other pieces of information are necessary, and these may be obtained from any handbook of radio engineering. The first is the fact that the L-C constant for a circuit resonating at 3,000 c.p.s. is 28.145×10^{-10}. The second is that the impedance of a parallel L-C circuit at any frequency is given by the equation:

$$Z = \frac{2\pi fL}{1 - (2\pi f)^2 LC} \qquad (42)$$

Substituting in this equation the values given, we get:

$$Z_1 = \frac{6.28 \times 1{,}500 \times L1}{1 - [(6.28 \times 1{,}500)^2 \times 28.145 \times 10^{-10}]} = 90$$

Transposing this equation to solve for L1 gives the following result:

$$L1 = 90 \, \frac{1 - [(6.28 \times 1{,}500)^2 \times 28.145 \times 10^{-10}]}{6.28 \times 1{,}500} = 7.2 \text{ mh}$$

Now to find the value of C1, we simply divide the L-C constant by L1:

$$C1 = \frac{28.145 \times 10^{-10}}{7.2 \times 10^{-3}} = 0.39\ \mu f$$

Solving for the elements of the shunt arm in like fashion, we find that L2 = 15.6 mh and C2 = 0.18 μf.

This completes the design of the equalizer, which is shown in Fig. 626. We know now that our equalizer has the correct attenuation at two points, 1,500 and 3,000 c.p.s. Having tentatively established the circuit constants, we should now calculate Z_1 at several

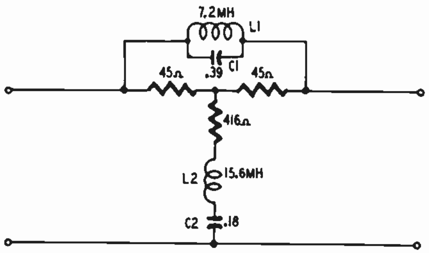


Fig. 626. *Here we have the final circuit of the equalizer whose design was calculated and described for you in the text.*

other frequencies to prove that the equalizer provides the exact attenuation characteristics we require.

Variable equalizers

In corrective re-recording for motion pictures or phonograph records it is often desirable to have available equalization which is variable, both in amplitude and in frequency. This can be accomplished by making the equalizer components variable through a ganged set of tap switches. If it is necessary to adjust only the amplitude while the resonant frequency remains constant, this can be done simply by replacing the fixed pad with a variable attenuator. A typical family of response curves resulting from this operation is shown in Fig. 627-a, wherein the peak attenuation is altered in 2-db steps. If a mathematical equation were to be developed for the ideal family of curves, it would be found that the equation would in each case be the same; that is, the shape of each curve would be identical.

In this case, however, the results as heard are not proportional. That is, the amount of available equalization becomes more and more crowded around f_0 as the peak amount is decreased. For example, consider the 8-db curve of Fig. 627-a. When the equalizer is adjusted to this condition, with a loss of 8 db at f_0, then the frequency at which the loss is half the maximum, or 4 db, is found to be about $0.5f_0$. But the half-loss frequency of the 6-db curve is $0.6f_0$, that of the 4-db curve $0.7f_0$ and of the 2-db curve $0.8f_0$.

A modification of the basic equalizer, in which the half-loss frequency remains constant for any insertion loss, is known as the *constant-B* network. A family of curves for this type of equalizer is

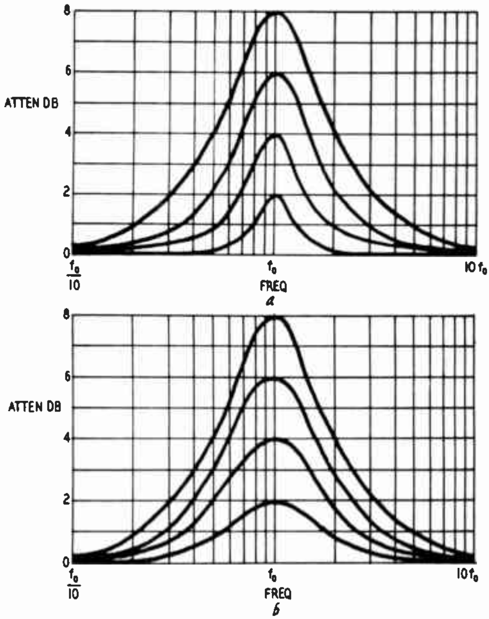


Fig. 627-a, -b. *Families of curves for equalizer circuits.*

presented in Fig. 627-b, where we can see that the lower half-loss frequency is in each case equal to $0.5f_0$.

To accomplish this characteristic, we begin with a conventional constant-resistance bridged-T attenuator and then add additional resistance elements in series with Z_1 and in parallel with Z_2. These additional resistances serve to control the shape of the attenuation curve by maintaining a constant db ratio between maximum and minimum insertion loss for any frequency. The basic circuit for the constant-B equalizer is shown in Fig. 628. Another interesting characteristic of this circuit is the fact that, so long as the maximum insertion loss doesn't exceed about 8 db, it is possible to obtain a perfect set of complementary curves simply by interchanging Z_1 and Z_2. For example a type A equalizer can be switched to a type B having attenuation characteristics identical to the transmission characteristics of type A, and so with types C and D.

Apparently all of these desirable characteristics could be obtained with the basic network if it were possible to adjust the values of the impedance elements. The particular advantage of the constant-B system lies in the fact that it is accomplished solely through the adjustment of resistances. When it is necessary to alter the resonant frequency, then the reactive components are varied. Thus this combination of adjustments will provide almost any amount of either emphasis or attenuation in any desired region of the audio spectrum.

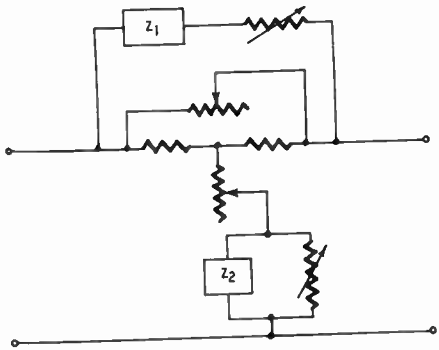


Fig. 628. *Basic arrangement of a constant-B equalizer circuit.*

Equalizer - amplifiers

Up to this point all of our equalizers have employed the use of lumped reactances, sometimes called "passive" elements. But it is possible to design an equalizer – amplifier which employs the reactive characteristics of a vacuum tube. There is normally considered to be a 180° phase shift between plate voltage and grid voltage as a signal passes through a vacuum tube, but it is possible to design the circuit so that this shift approaches 90°, with the plate current either leading or lagging the plate voltage. In this case the tube is behaving very much like an inductance or capacitance and is known as a *reactance tube.*

chapter 7

loudspeaker systems

The loudspeaker is the final link in the component chain of audio facilities, the part of the system whose function it is to bring the sequence of events around full circle, to re-create the waves of sound in air which are the starting point of all audio operations. This is indeed a formidable task, involving as it does a frequency range of 10 octaves and a dynamic range of 60 db or more.

The speaker system, taken as a whole, is an *electroacoustic transducer* which transforms electrical energy from an audio amplifier into a mechanical motion which generates acoustic energy. The problem in the design and application of speaker systems, therefore, is that of providing as efficient a coupling as possible between the electrical signal source and the air acoustical load.

The speaker load

When we refer to air as the *load* on a loudspeaker, this can be taken quite literally, as is easily demonstrated by recalling the tremendous effect of *air friction*, more commonly called "wind resistance," in dissipating the power of a fast-moving vehicle. We so often employ the expression "light as air," that we may think of air as having little or no weight at all. But the truth is that the weight of the air in a fairly large living room, 24 by 30 feet with a 14-foot ceiling, is nearly *750 pounds*. This is the sort of load which the loudspeaker must push around as it re-creates the sounds originally fed into an audio system.

The atmosphere which surrounds us is really a mixture of several well-known gases, consisting of about 78% nitrogen and 21% oxygen. The remaining 1% includes very small quantities of other gases such as argon, carbon dioxide, helium, hydrogen, krypton, ozone and xenon. Now let us see just what happens in this atmosphere when audible sounds are produced.

The atomic particles of the various gases in the mixture average about 1.5 hundred-millionths of an inch in diameter. Although there are a tremendous number (44×10^{19}) of these particles in a cubic inch of air, any one of them can move around 2.5 millionths of an inch in any direction before it will strike one of its neighbors. This void in which the particle can travel is known as its *mean free path*. And while it seems extremely tiny, it is really huge

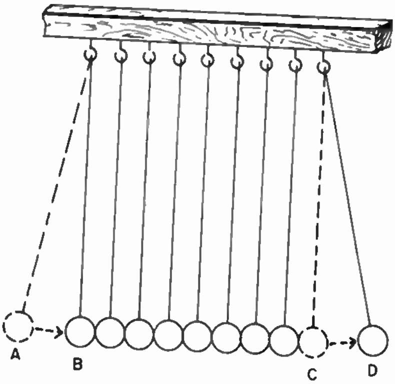


Fig. 701. *Mechanical illustration of the transmission of sound in air.*

compared to the diameter of the particle itself, the ratio being about 160 to 1.

The atmosphere is really a great deal of open space and a number of particles of several gases. These particles are constantly moving around at random, traveling in all directions and at a number of speeds, the average being about 0.5 mile per second. The particles are constantly colliding with one another, and each collision may alter the direction and velocity of the particle motion. This is the "rest" condition of the homogeneous elastic medium through which it is possible to propagate sound.

Sound in air

The propagation of a sound wave in air is closely analogous to the action of the suspended steel balls of Fig. 701. When one of the end balls is displaced from its rest position at B and swung out to position A, it will of course tend to swing back to B as soon as it is released and thereby to strike its neighbor. Almost at once and as a direct result of the A–B motion, the ball at the other end will swing out from point C to point D. But meanwhile the interven-ing balls will hardly have moved at all. Thus the *motion* A–B has resulted in a similar *motion* C–D although between these two extremities there has been practically no motion at all and there has been *nothing* which has been physically transported from A–B to C–D.

Bearing this analogy in mind, let us consider the arrangement of Fig. 702. In the center of this apparatus we have a permanent

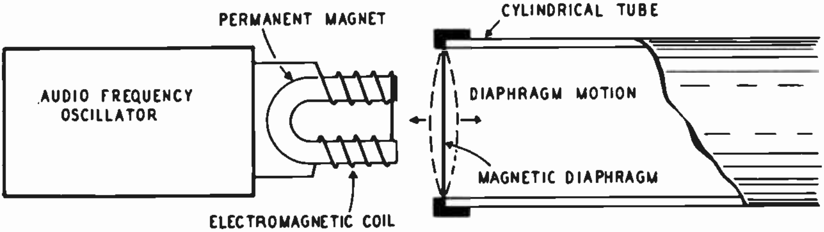


Fig. 702. *Movement of the diaphragm, caused by the varying current through the coil, produces sound.*

magnet, an electromagnet and a magnetic diaphragm. Since the diaphragm is within the field of the permanent magnet, there will be a certain force of attraction between these two units. When cur-rent passes through the coil, it will set up a magnetic field which either aids or opposes that of the permanent magnet and the attrac-tive force between magnet and diaphragm will be either increased or decreased, depending upon the direction of the current. If the diaphragm is thin and flexible enough, it will have a certain amount of curvature due to its tendency to be attracted by the magnet. The amount of this curvature will vary with the strength of the magnetic field.

If an alternating current is passed through the coil, as from the audio-frequency oscillator shown, the diaphragm will tend to vibrate at a rate determined by the frequency of the current pass-ing through the coil and the amount of diaphragm displacement at any given instant will be proportional to the current intensity. This action is the principle of the magnetic-diaphragm receiver, the type most commonly employed in low-cost headsets.

Piston effect

Let us now attach a cylindrical tube to our receiver in such a way as to cause the diaphragm to disturb the air particles in the tube. The diaphragm now really becomes a *piston,* although it has a very short stroke, working against the fluid pressure of the air in the tube. Before the audio signal is applied, the diaphragm is in a fixed position and the air particles in the tube are in their customary random-motion condition known as *equilibrium.*

When the current in the coil is such as to oppose that of the permanent magnet, the pull on the diaphragm will be weakened and it will move in towards the cylinder. In this case the particles of air which are immediately in front of the diaphragm have less space to occupy than formerly and therefore more of them will collide with particles in the space ahead of the compressed area.

The ultimate result of these additional collisions is that the effect of the original compression at the diaphragm is transmitted to each adjacent group of particles, so that the disturbance travels right on down the tube. It should be understood that the particles originally compressed at the diaphragm do not themselves move down to the end of the tube, but the effect of their motion does, due to the fact that there is a particle-to-particle interaction propagated from point to point along the entire length. This is known as the *compression* half of the sound cycle.

When the diaphragm moves back, there is a larger space for the nearby particles with consequently fewer collisions among the particles than under the equilibrium condition. This condition is transmitted down the tube in similar fashion, due to a forward-traveling segment of the air which has fewer than average particles. This is known as the expansion or *rarefaction* half of the sound cycle.

As the diaphragm vibrates back and forth at an audio rate these compression and rarefaction effects alternately move along the length of the tube, but the actual air particles themselves do not move any appreciable distance from their normal random positions. Thus, you can see that the process of sound propagation is very similar to the electron theory of current flow in which random-moving free electrons are caused to collide with their neighbors when a voltage is applied. These effects are felt along the entire length of the conductor, although the distance traversed by any single electron is very slight. It is for this reason that much electrical theory can be made to apply to the mechanical problems of loudspeaker system design.

Like other electroacoustic and electromechanical devices, such as microphones, disc recorders and pickups, the loudspeaker has available a number of methods for converting electrical energy into mechanical energy. Various schemes have been tried employing the piezoelectric effects of crystals, electrostatic effects of charged plates and several electromagnetic systems. Although each of these systems has its supporters, only one method is in general use today — the loudspeaker shown in Fig. 703.

The dynamic speaker

This speaker consists of a paper cone which has fastened to its apex a *voice coil,* which in turn is mounted in the intense field of a powerful permanent magnet. The cone has a flexible suspension

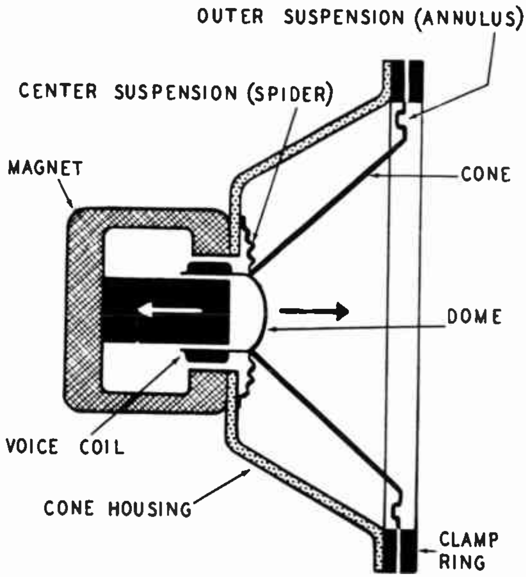


Fig. 703. *Cross-sectional view of a dynamic speaker. Most speakers are of the permanent-magnet type.*

mounting both at its apex and around its outer edge. The entire coil and cone assembly is therefore free to move as a unit, up to the mechanical limitations of the suspensions.

When an audio-frequency current passes through the voice coil, it establishes a magnetic field which alternately aids and opposes that of the field magnet. There will, therefore, be a tendency for the voice coil (which has now become an electromagnet) to move, either further into the field or out of it, as determined by the direction of the audio current at any instant.

Since the cone is rigidly attached to the voice coil, it too will be forced to move as the coil moves. This motion will set up compression and rarefaction regions in the air particles before it, in much the same manner as did our piston diaphragm of Fig. 702. These disturbances of the air particles may be perceived as sound. If they represent a true reproduction of the electrical signal causing them and if the electrical signal is a faithful representation of the original sound waves, then the re-created sound should present to the air the identical aural impression that the ear would have received had it been present at the original production of the sound. As we stated in Chapter 1, no such degree of perfection has yet been attained. Now let us examine in greater detail the problems posed by such requirements in the design of practical loudspeaker systems.

Voice-coil force

The force which the voice coil exerts on the cone in causing it to move is the product of three factors and expressed by the equation:

$$F = Bli \tag{43}$$

where F is equal to the force in dynes, B the flux density of the air gap, l the conductor length of the voice coil in centimeters and i the current flow through the voice coil in amperes.

When this force is exerted on the cone, the movement of the cone will be opposed, not only by its own mass and inertia, but also by the load of the air to which it is coupled. The velocity of the movement of the cone is therefore expressed by the equation:

$$v = \frac{F}{Z_M + Z_A} \tag{44}$$

where v equals the velocity in centimeters per second (cm/s), Z_M the impedance of the speaker mechanism in mechanical ohms and Z_A the radiation impedance of the air load in mechanical ohms.

The cone acts to couple the air-load impedance Z_A to the driving force of the loudspeaker motor. In this type of speaker, commonly called a *direct radiator,* the original force generated by the motor is localized to the voice coil area but it must be transmitted to the air load which normally extends over an exceedingly large area. To do this effectively while holding Z_M to a minimum, the cone must be as light and as stiff as possible. But while a hard, stiff cone has the greatest efficiency it also has the highest transient distortion. A soft feltlike cone, on the other hand, will have a smoother response curve with better transient response but its efficiency and high-frequency response will be poorer.

At the lower frequencies, up to between 700 to 1,400 c.p.s., the cone acts somewhat like a piston diaphragm whose diameter is that of the cone itself. In this region all parts of the cone move in phase to displace air particles. The frequency at which the piston action ceases is known as a *mode of vibration.* Above this point the cone vibrations are more like waves which travel outward from the apex. The center part of the cone vibrates with considerably greater intensity than the outer edges, this effect becoming more pronounced with increasing frequency. As a result the effective mass is reduced and the cone acts as if it were becoming steadily smaller as the frequency is increased. The practical effect of this is to provide a high-frequency response much better than it would be if the cone were to behave as a piston throughout the spectrum.

Cone resonance

The coil and cone assembly unit will have a definite resonant period usually designed to be at the lower extremity of the frequency range being reproduced. At this point the motional impedance Z_M becomes very small, with a resulting tendency for the velocity and amplitude of the radiated power to become quite large. There is generated, however, a counter-e.m.f. which is like that in any motor in that it tends to oppose the current flow caused by the applied voltage. By proper design this back-e.m.f. can be controlled so that the response at resonance either peaks or rolls off.

Advantage is often taken of this fact in cheap speakers to introduce deliberately a strong resonant peak which is highly distorted and therefore rich in harmonics. This increases the overall loudness at the lower frequencies considerably. But it is, of course, only a trick to produce false bass and it has attendant with it very high orders of intermodulation and transient distortion. It is consequently avoided in high-quality speaker design wherein the response may have only a slight resonant peak or none at all. Below resonance the impedance increases rapidly and so the response drops off rapidly.

Speaker enclosures

An enclosure or baffle is an essential element of any direct-radiator loudspeaker system. Whenever the front of the speaker cone is compressing or rarefying the particles of air before it, the rear surface of the cone is doing much the same thing, except that the two actions are exactly 180° out of phase with each other. If these rear pressures are permitted to affect those in front, there will be a partial cancellation between them, this effect being more troublesome at the long-wavelength bass end of the spectrum. Thus, for good efficiency and to avoid an attenuated bass characteristic, it is necessary to provide some means of preventing this back-to-front cancellation.

The device which controls the speaker back wave is known as a *baffle* and its diameter should be at least a half-wavelength for the lowest frequency to be reproduced without attenuation. In its simplest form the baffle is nothing more than a board of the proper

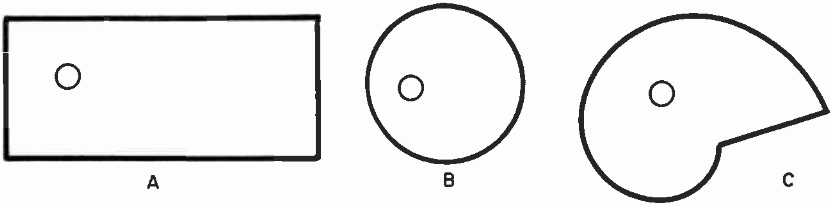


Fig. 704. *Irregularly shaped baffles and off-centering of speakers avoids a single cut-off frequency.*

dimensions having a suitable hole against which the speaker is mounted. With this arrangement the back wave must travel out to the end of the baffle and then turn around to the front before it can cause any serious trouble. Since by then it has been largely dissipated in the air load at the rear, its effect is of little consequence except at the cutoff frequency. To distribute over a considerable range the destructive interference effects at cutoff, the baffle is shaped irregularly so that there are in effect a number of cutoff frequencies rather than one critical frequency. Examples of such shapes are shown in Fig. 704.

To be perfectly effective down to the lowest frequencies the flat baffle must be of very large size. This condition is approached in practice when the speaker is mounted in a wall between two relatively large rooms. This method is rather simple, free from cabinet resonance, and saves the space required for a large bulky enclosure. The simplest type of enclosure, of course, is the flat baffle to which have been attached sides, top and bottom, thus forming an open-backed box. This is the familiar construction so often found in the cabinets of radio and television sets. It is not very satisfactory, however, as the dimensions are usually such as to preclude adequate bass response and resonant peaks degrade the frequency response.

When a back is attached to this cabinet it becomes a total enclosure, often incorrectly referred to as an infinite baffle. This is likewise an undesirable method of baffling, for it reacts upon the speaker in such a way as to raise its resonant frequency. This may cause the resonance peak to be considerably more pronounced than before and it will certainly impair the overall bass response for, as we know, the output is rapidly attenuated below resonance. As

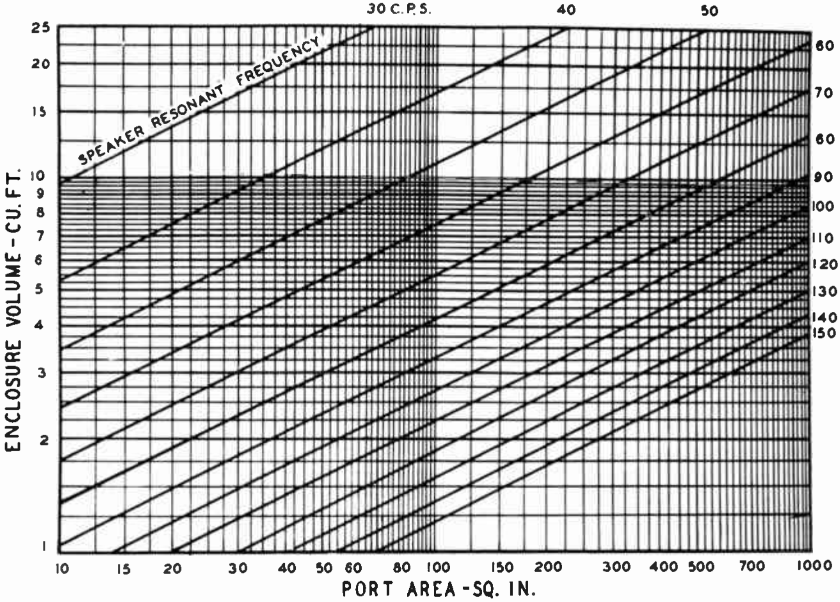


Fig. 705. *This chart gives the relationship between the volume, resonant frequency and port area of a bass-reflex enclosure.*

a practical matter we can say that the efficiency is cut almost in half in the bass region when a total enclosure is employed.

Bass reflex

An enclosure which not only does not attenuate the bass, but actually reinforces it, is the *acoustic phase inverter*, popularly known as the bass-reflex cabinet. Almost a total enclosure, it differs by having a port opening in the front of the cabinet. When this opening is properly designed, it will cause the speaker back wave to emerge in phase with the front wave and thus actually add to it. It cannot do this throughout the spectrum, but only in a region of two or three octaves and this reinforcement band is usually chosen to be down at the lowest frequencies. Sound-absorbing material is placed within the cabinet to absorb the upper frequencies so that they cannot emerge out of phase and cause interference and cancellation.

The port is usually tuned to a frequency just a little below the free-air resonance of the speaker. This results in a bass response curve somewhat smoother and with a lower limit. Tuning is accomplished by altering the area of the port, whose shape may be either round or rectangular, but whose *size* is rather critical and is given by the equation:

$$A = \frac{\pi V^2}{4} \left(\frac{2\pi f}{c} \right)^4 \qquad (45)$$

where A equals the area of the port, V the volume of the enclosure, f the free-air resonant frequency of the speaker and c the velocity of sound in air. A useful chart showing the relationship between these various quantities is given in Fig. 705.

Since it performs extremely well for a very small size unit and is a proven design of many years' standing, it still enjoys widespread popularity. The construction is rather simple, but a few hints should make the work go more smoothly.

Enclosure construction

The overall dimensions of the enclosure are not critical, but better results will be obtained with 12-inch and larger speakers if the cabinet volume is at least 4 cubic feet. The depth of the box should be at least 1 foot. All boundary surfaces should be of heavy, seasoned lumber, preferably of plywood ¾ inch or more in thickness. The front panel in particular should be absolutely flat so that no undue stresses are placed on the speaker when it is mounted.

All joints must be airtight and well secured by screws and glue (except for the back panel). Bracing inside corners of the joints with glue blocks is recommended for avoiding noises at cabinet resonance. Cleats should be provided for attaching the removable back to the sides, and a very close fit is desirable for a joint as airtight as possible. The back should be secured to the cleats by screws at each corner and at 4-inch intervals along each edge. The inner surfaces of the cabinet should be lined with a material which is sound-absortive, particularly at the high frequencies, such as 2-inch-thick Fiberglas or rock wool. This can also be obtained with paper on one side only, and is mounted with the paper against the cabinet. It should be installed before the speaker is mounted to prevent any fragments from getting into the speaker mechanism. A further precaution is to cover all of the padding with a couple of layers of cheesecloth. All of the back should be padded along with about half of each of the remaining surfaces. The choice of grille cloth is important also, for it could be responsible for considerable high-frequency attenuation. If a grille cloth is necessary

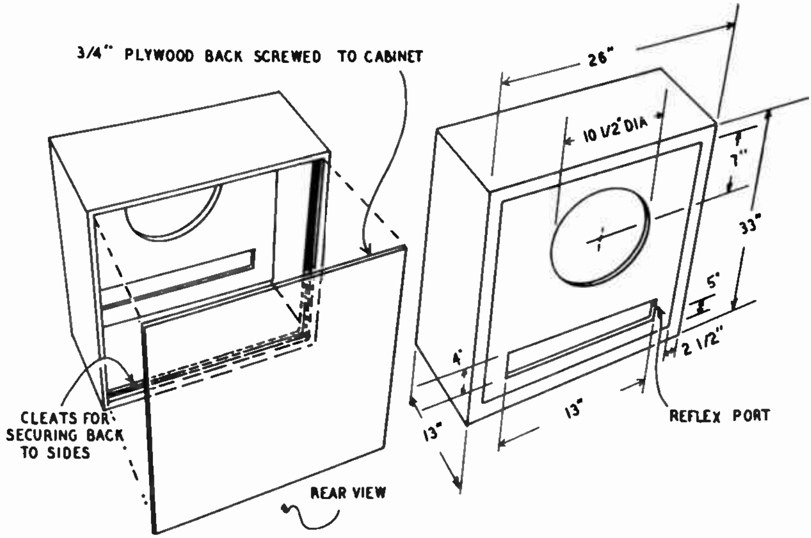


Fig. 706. *Constructional details of a bass-reflex enclosure.*

for appearance sake, then it should be of a rather coarse mesh and not too soft. To get a rough idea of the sound transmission characteristic of a given cloth it may be held up to the light. If it appears quite dense, it would be better not to use it. A typical bass-reflex enclosure design is shown in Fig. 706.

Another type of phase inverter is the *acoustic labyrinth,* whose cross-section is shown in Fig. 707. This operates on the same general principles as the bass reflex, except that tuning is determined by the length of the air column into which the rear of the cone works. This system usually has somewhat better bass response than the bass reflex, sometimes by as much as an additional full octave. Its main drawback is the difficulty of construction.

Helmholtz resonator

The principle of phase inversion is employed in another way in the *Helmholtz resonator* type of enclosure, for which is also claimed an additional octave of bass response over that of the bass-reflex system. This principle, commonly employed in the design of the bodies of string instruments, can best be understood by considering first what occurs when it is used in this application.

The body of a guitar, for example, is a wooden box, fully enclosed except for a sound hole. When a string is plucked, particles

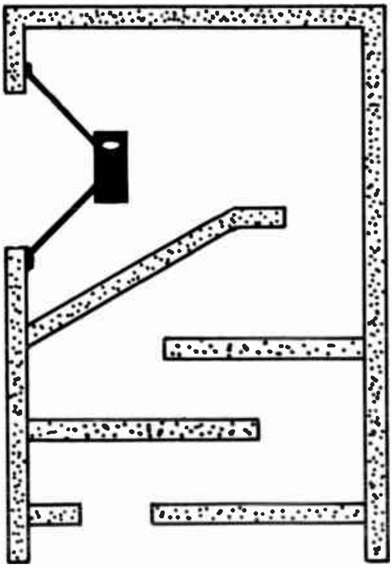


Fig. 707. *Cross-sectional view of an acoustic labyrinth.*

of air are set into motion in the vicinity of the vibrating source, including the air over the sound hole. The compression and rarefaction of the particles in this area are transmitted to the captive body of air within the box. For this air to return to an equilibrium condition it must return these pressure effects back out of the same hole through which they entered.

Thus the pushing up and down on this mass of air is rather analogous to pushing on a spring. When a trampoline performer in the circus, for example, jumps up and down on his springy mat, he is doing much the same thing. The secret of the tremendous lift he gets from his apparatus lies in the fact that he times his up-and-down motion to a point where it is nearly self-sustaining. He has in fact reached the resonant frequency of his mechanical system.

In the same manner the Helmholtz resonator, as its name implies, has a resonant period, and it is one with a very high Q. This means that the air pressures developed at resonance will be very much stronger than those even at nearby frequencies. This is hardly a desirable condition for an enclosure which should have a broad and flat response, but the Helmholtz principle does offer the advantage of considerable efficiency for a very small size. To adapt this to a speaker enclosure, one must retain a measure of this efficiency while flattening the resonant peaks.

A practical form of Helmholtz resonator for loudspeakers is shown in Fig. 708. Note that the speaker is mounted within the cabinet on a baffle board somewhat smaller than the inner dimensions of the box itself. The front of the cabinet has a rectangular

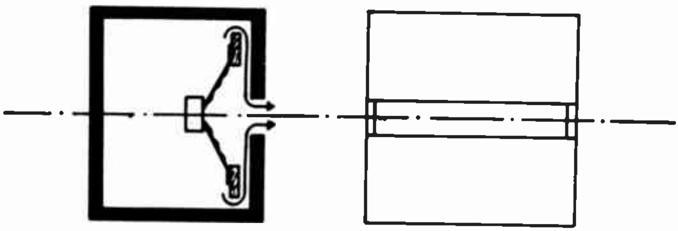


Fig. 708. *The Helmholtz resonator has a strong resonance point.*

or oval port, and the separate speaker mounting board is installed an inch or so behind it. Thus, there is a duct through which the back wave can travel to join the front wave. If the unit is properly dimensioned, the two waves will emerge through an opening in phase at the bass frequencies. Furthermore, when the port has the correct size and shape, the resonance peak is flattened considerably, being now of such magnitude that it is useful for bass boost without excessive distortion. The interior is padded as was the bass reflex for absorption of the upper spectrum which is not to be boosted. The outstanding characteristic of this enclosure is the diminutive size required for such performance. The unit of Fig. 708 is only 20 inches square and 16 inches deep.

The musical-instrument approach is also employed in the enclosure of Fig. 709 but the basis here is the *closed pipe,* a device which is used for some of the tone generators in pipe organs. This pipe is stopped at only one end and, when it has a high ratio of length to width, it will have a resonant frequency whose wavelength is four times the length of the pipe itself. The element is then also analogous to a closed-end quarter-wave section of transmission line.

This pipe will characteristically resonate at all of the odd har-

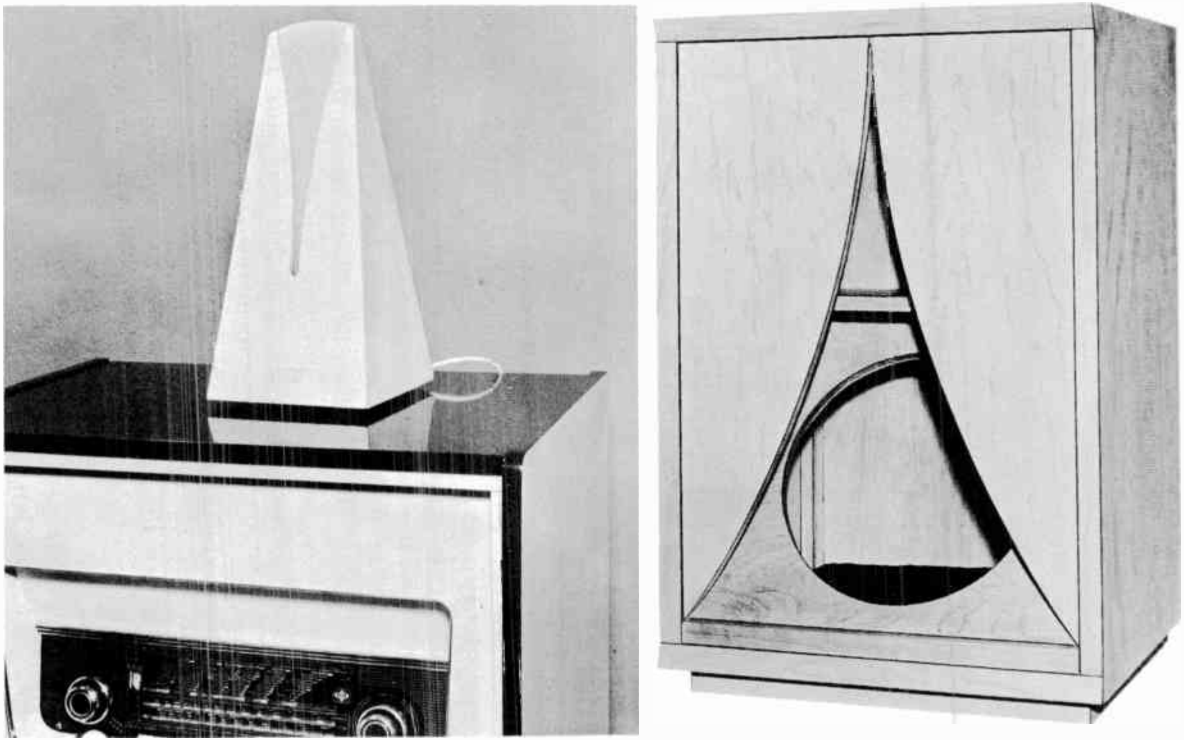


Fig. 709. *The unfinished cabinet on the right shows the speaker board mounted behind the tapered slot. A port is behind the narrow end of the slot. On the left a miniature version using the same principle boasts a smooth response from 80 cycles to beyond 15 kc with a 4-watt, 5-inch, 8-ohm speaker. This 4-pound unit may also be hung on the wall in an inverted position.* (Karlson Associates, Inc.)

monics of its resonant frequency or, when it is used to produce a tone, the sound will be rich in the odd harmonics of the fundamental. If a tapered slot is cut into the open end of the pipe, the resonant peaks will be considerably broadened and if the slot is finally extended to a distance greater than two-thirds of the overall length of the pipe, then the resonant peaks will just about disappear.

With the correct dimensioning of the pipe and a slot which follows an exponential taper, the transmission characteristics of the device become uniform in the audible range but its abilities as a quarter-wave acoustic-matching transformer remain unimpaired.

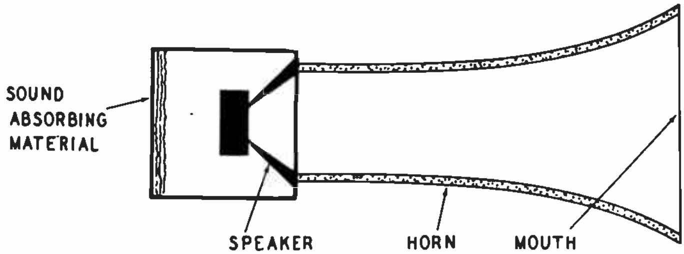


Fig. 710. *Typical horn-type loudspeaker.*

Adherents of this design claim superior performance in terms of minimum distortion and a better overall directional pattern.

In some applications, where space is not at a premium, a conventional direct-radiator loudspeaker may be used in conjunction with a *directional baffle* or horn, as shown in Fig. 710. Such a horn is considerably shorter than the true horn type loudspeaker which is yet to be discussed. It has a throat area equal to or slightly less than the cone size and has a fully enclosed box for capturing the back wave. The box is lined with sound-absorbing material to minimize resonance effects.

With a horn of optimum throat size the mechanical impedance of the air load presented to the cone will be appreciably greater, especially at the lower frequencies, than that of conventional plane baffles. This increases the efficiency to several times that of the ordinary baffle and in addition there is the desirable concentration of energy due to the directivity of the horn baffle.

The folded horn

For indoor applications where space must be conserved, the directional baffle or horn may be *folded* as in Fig. 711-a. Despite the labyrinthine curvature of this unit, it is nonetheless a horn whose rate of expansion follows an exponential taper. It is designed to be installed in a corner, so that the walls of the room act as an extension of the walls of the horn.

While the low frequencies can be made to follow the sharp

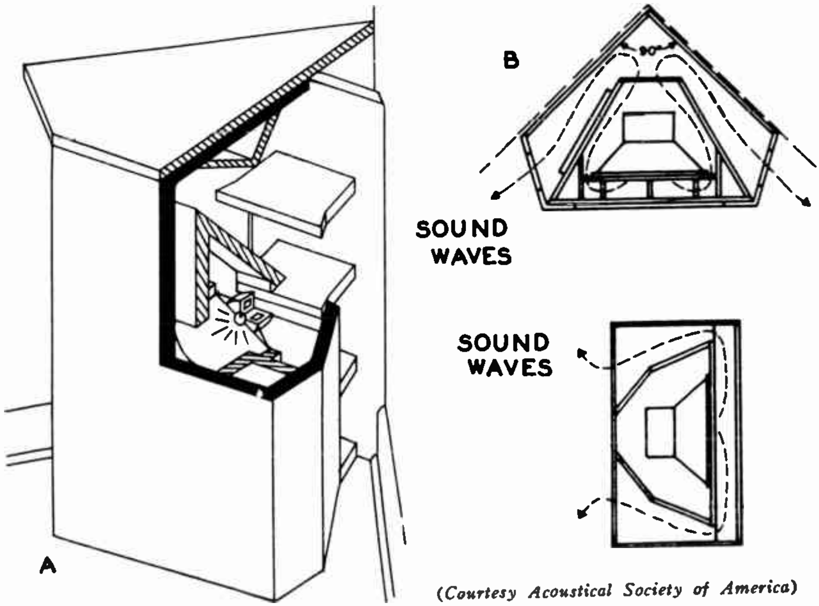


Fig. 711-a, -b. *Example of the construction of a folded horn.*

turns of the sound path shown in Fig. 711-b, the higher frequencies will be dissipated. The folded corner horn thus cuts off at an upper limit of about 1,500 c.p.s. and no effort is made to use it for reproduction above this point. Additional high-frequency speaker elements must therefore be employed to complete a wide-range system.

Horn-type speakers

Referring to Fig. 702, the straight pipe is unsuitable as a sound radiator except where very sharp resonance effects are desired. But suppose we consider the action occurring when we replace the pipe with a horn as in Fig. 712. Now each successive compression and rarefaction of the air is slightly greater in volume due to the flare of the horn.

The horn now is acting as an acoustic transformer, coupling the relatively heavy vibrating surface at the throat of the horn to the relatively light air load at the horn mouth. This may be visualized by reference to Fig. 713. For purposes of discussion it is assumed that the horn is of very large size relative to the wavelength of the transmitted sound, and there are no losses occurring in the horn.

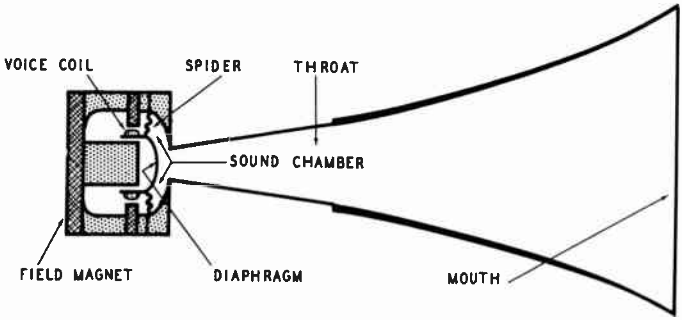


Fig. 712. *The horn acts as an acoustic transformer.*

The magnitude of the acoustic power varies directly with the cross-sectional area of the horn and with the square of the air

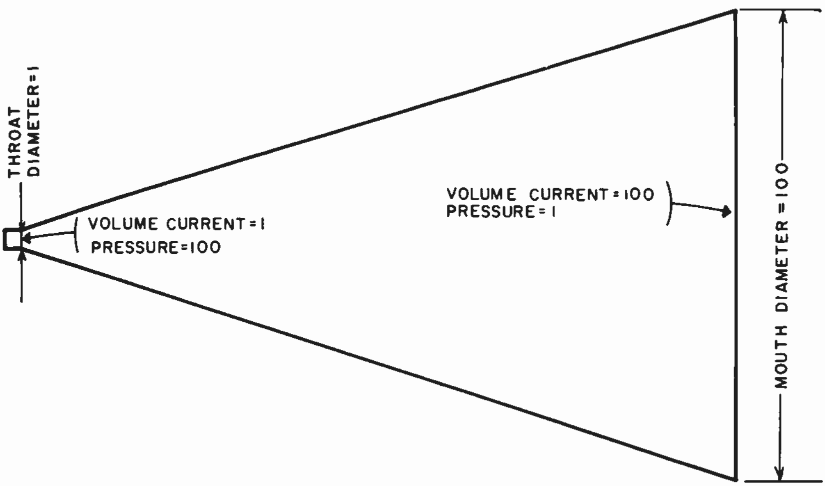


Fig. 713. *Relationship of throat and mouth pressures in a horn.*

pressure. Since the area of a circle is in turn proportional to the square of the radius, there is an inverse relationship between the diameters and corresponding pressures at each end of the horn. Thus if the ratio of mouth and throat diameters is 100:1, then the ratio of throat to mouth pressure is similarly 100:1.

Just as in an electrical circuit the power equals the voltage times the current, so in acoustics the power equals the pressure times the rate of air flow, known as *volume current*. Since in this horn we have assumed a 1:1 power ratio between throat and mouth, when we decrease the pressure to one-hundredth, to maintain the same power level we must increase the volume current 100 times. This is directly analogous to an electrical transformer having a turns-and-voltage ratio of 100:1 and a current ratio of 1:100 and explains why the correctly designed horn behaves as an acoustic transformer.

Horn design

The mouth of the horn terminates in the acoustic medium

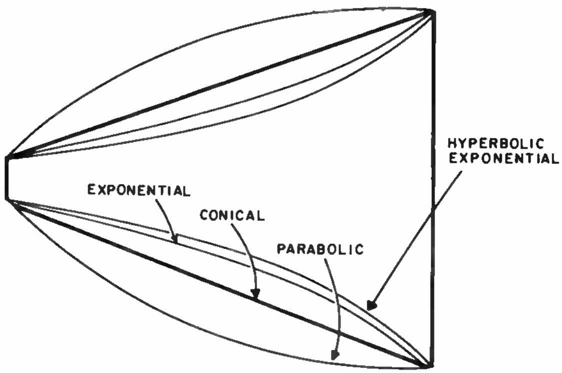


Fig. 714. *This illustration shows shapes of typical horns.*

which is air. The size of the horn mouth with respect to the wavelength of the lowest frequency to be reproduced is critical, for this will determine the smoothness of the bass response and freedom from resonance. The lower cutoff frequency of a horn will be that at which the diameter of the mouth is about a third-wavelength. For frequencies below cutoff most of the sound will be reflected into the horn as soon as it reaches the mouth.

The shape of the horn should be such that it produces a smooth and continuous increase in cross-sectional area as it expands from throat to mouth. A wide variety of shapes have been tried experimentally, including those of Fig. 714, but most commonly used is that which follows a mathematical equation based upon the *exponential law*.

The exponential horn is simply one in which the cross-sectional area doubles for each unit of its length. But nothing is said about the size of these units, so long as they are all the same. Thus it is possible to make long narrow horns with a slow taper or short wide horns with a rapid flare, and yet each may follow the exponential law. Obviously the length of the intervals between each area doubling will determine the horn shape and its *expansion ratio.*

The expansion ratio is the other factor which determines the bass response, the lower cutoff frequency being inversely proportional to the rate of expansion. This relationship is reduced to some practical values in the table of Fig. 715. We can also quickly find the optimum mouth diameter in inches simply by dividing the lower cutoff frequency into 4,000.

LOWER CUTOFF FREQUENCY	16 C.P.S.	32	64	128	256
INTERVALS OF LENGTH FOR AREA DOUBLING	48 INCHES	24	12	6	3

Fig. 715. *Relationship of horn area and lower cutoff frequency.*

With these facts at hand, we can proceed to design a practical exponential horn. Suppose we wish to build a horn which will reproduce effectively down to 32 c.p.s., using a driver with a 1-inch throat diameter. Then the diameter of the mouth will equal 4,000/32, or 125 inches. Now the area of the throat equals the diameter squared times 0.7854, or in this case 0.7854 square inch. Since from Fig. 715 we learn that the area doubles for every 2 feet, at a distance 2 feet from the throat the area must be $0.7854 \times 2 = 1.5708$ square inches. At 4 feet it is $1.5708 \times 2 = 3.1416$. Continuing out to the limit of the mouth as previously calculated, we arrive at the design of Fig. 716.

This gives us the horn areas at 2-foot intervals throughout its length. If the horn is to be circular we must determine the diameters at each point by the formula $\sqrt{A/0.7854}$. The horn may also be square with approximately equivalent results. In this case the width and height may be found simply by extracting the square root of the area at each point.

Since our design was predicated on a mouth diameter of 125 inches, its area would be 12,271.875 square inches, which is a little less than that at the 28-foot length. Thus the horn could be slightly shorter and still provide a smooth response down to 32 c.p.s. as required. In any case this is a huge thing and so the horns are often coiled as in Fig. 717 or folded as in Fig. 711. This is entirely permissible, provided that the bends are not too sharp and the rate of expansion follows the appropriate mathematical laws.

Fig. 712 shows that the horn is coupled to the diaphragm of the driver unit through an air space known as the *sound chamber.* The purpose of this is to provide the best possible coupling between these two elements. Since the throat area is usually smaller than the diaphragm area, the diaphragm works into a load at low frequencies which is the product of the throat impedance and the square of the ratio of the diaphragm area to the throat area. In this way the diaphragm imparts a much higher velocity to the air in the throat than it would if the diaphragm were located in the

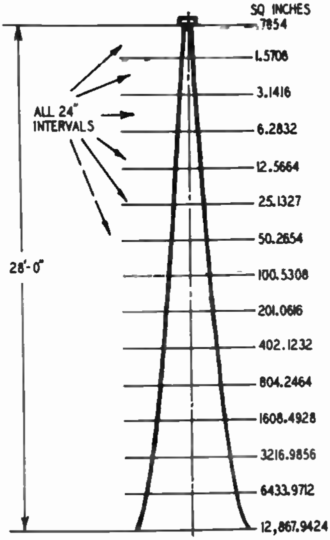


Fig. 716. *Design of a horn.*

throat and drove it directly at its own velocity. The result of all this is an efficiency which may be 10 times as great as with an equivalent direct radiator having the same power input.

Multiple speakers

Whenever cost and space permit, multiple smaller loudspeakers all of which cover the same frequency range are preferable to a single larger speaker. In some applications this is usually mandatory, as for sound reinforcement in auditoriums or stadiums, but it is also good practice even in home installations for the following reasons:

1. A greater efficiency and frequency response due to lighter cones.

2. A better directional pattern due to variable placement of individual speakers.

3. Less power dissipation as heat losses in the coils.

When the speakers are connected in multiple, a new impedance will be presented to the source depending upon the method of connection. A series connection of a large group of speakers is not recommended, as transient voltages high enough to cause arcing in the driver units may be developed. Parallel or series-parallel connection is therefore much to be preferred.

If only two identical speakers are used simultaneously, the

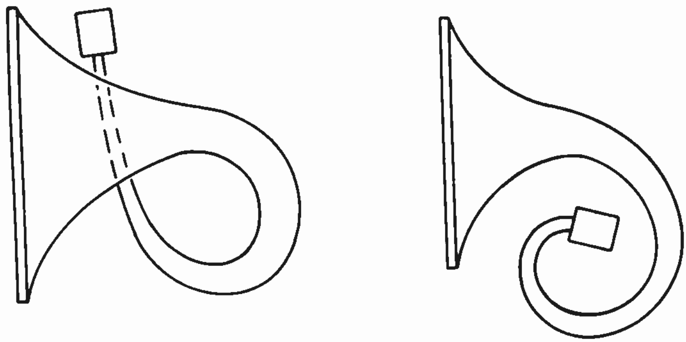


Fig. 717. *Horns are often coiled or folded.*

choice of connection is not critical. Whichever method provides the most convenient match to the available transformer taps may be used. The total impedance of two identical speakers connected in series will be twice the voice coil impedance of either one alone. Similarly when two identical voice coils are connected in parallel the resultant impedance of the combination will be one-half the impedance of either voice coil alone.

It is especially important when multiple speakers are operated close together that they be properly phased. This means that the several cones must all set up a compression alternation simultaneously and must also cause rarefaction simultaneously. If they do not, destructive interference and partial cancellation may result. This means that the cones must move in the same directions simultaneously, and this direction of movement will depend upon the direction of current flow through the voice coils, all other factors being equal.

The solution to the problem, then, lies simply in connecting the common signal to the several voice coils so that the direction of current flow is the same in each coil at any given time. For series operation of two speakers, two unlike terminals must serve as the junction between the units. Then the other unlike pair will be connected across the source as shown in Fig. 718-a. For parallel operation of two units each pair of like terminals is joined across the source as in Fig. 718-b.

Complementary systems

An even more widely used multiple-speaker arrangement is one in which the several units cover complementary frequency ranges. This system has all the advantages already cited for multiple identical arrangements, plus the following:

1. Less intermodulation distortion,
2. Less transient distortion,
3. Less frequency-modulation distortion.

Since the cone type direct radiator is not very efficient at the upper frequencies, it is normally used only for the low-frequency element, commonly known as the *woofer*. The high-frequency elements are almost invariably some form of horn, commonly referred to as a *tweeter*. In three-element systems the mid-range speaker is sometimes called a *squawker*.

The biggest problem with all types of high-frequency speakers is their extremely sharp directivity. To make the high-frequency radiation pattern approximate that of the lower frequencies, various modifications of the basic exponential taper are used. These include *multicellular* and *multisectional mouths*, so-called *optical slits* and *acoustic lens*, and *reciprocating flares*. Each of these methods has its earnest proponents, but the effectiveness of one over another will hinge largely on personal taste as well as the performance of the remainder of the loudspeaker system.

The several elements may be mounted at somewhat separated points, but repeated listening tests would seem to confirm that they should be as close together as possible, so that the sound will appear to be emerging from a single point source. Sometimes they are nested together on a common axis and in some cases the various units are integral parts of a single *coaxial* speaker assembly.

Crossover networks

Since it is not possible to design a loudspeaker which will reproduce a given band of frequencies faithfully while at the same time sharply attenuating all frequencies outside its assigned band, it is necessary to install an electrical network into the circuit, so arranged that each speaker receives only those signals which lie within its predetermined range. The electrical circuits are known as *crossover* or *dividing* networks. Multiple-speaker systems today may have four or more radiating elements, but a simple two-speaker combination of woofer and tweeter is by far the most widely used. Our discussion of dividing networks will therefore be confined to those for two-way speaker systems. It is possible, of course, that power requirements are such that several identical woofers and several identical tweeters are employed, but this would still be only a two-way system with only one crossover frequency.

The dividing network is usually placed between the amplifier

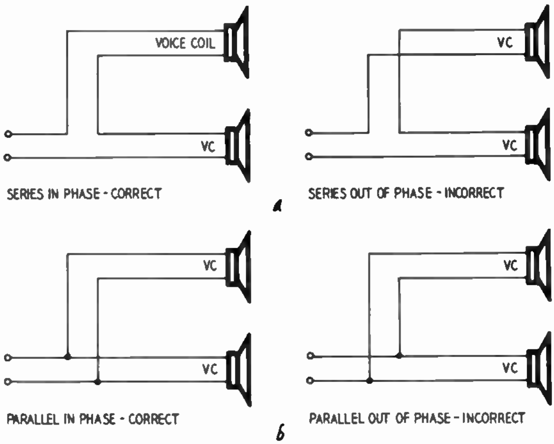


Fig. 718-a, -b. *Proper phasing of speakers is important.*

output transformer and the voice coils of the speakers, and it is for this type of operation that design data will be given. It is possible, however, to construct networks which are to operate elsewhere in the circuit. One such location is between the final amplifier stage and its load circuit. An example of a design for this class of operation is shown in Fig. 719. This has the advantage that, since the network is in a high-impedance circuit, smaller values of capacitance and inductance are required than in the more conventional use. Its greatest disadvantage lies in the fact that two separate output transformers are required, one for each speaker. The transformers need not be of exceptionally high quality, however, as neither of them is required to pass the full spectrum.

It has also been suggested that frequency division occur even farther toward the system input. In this case two separate amplifying channels are required following the dividing network. Such a system would provide greater stability and power losses in the network would be very much less, but the duplication of amplifiers would raise the cost considerably.

Crossover frequency

The selection of a crossover frequency will depend upon several factors. If the frequency is too low, exceedingly large and expensive capacitors will be required as well as a larger tweeter element.

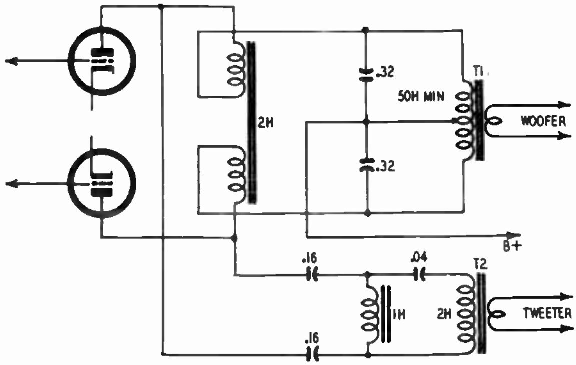


Fig. 719. *Representative crossover network.*

This will practically defeat one of the purposes of the two-way system.

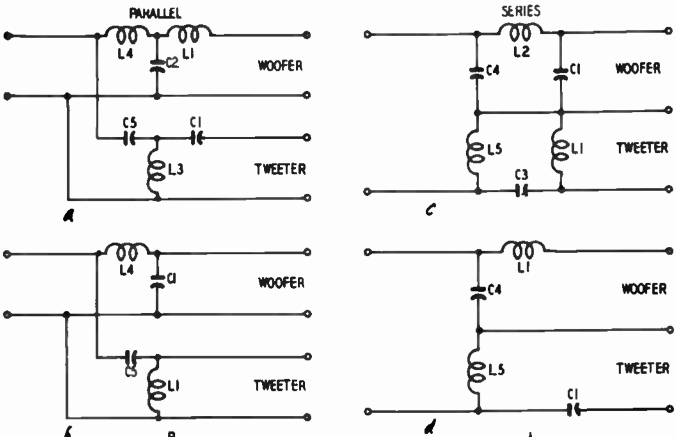
If the crossover frequency is too high, there is the possibility of encountering the characteristic dip in response which all large cone speakers seem to exhibit. And if the system will be operating often at voice frequencies, as in motion-picture theater applications, too high a crossover will result in splitting the fundamental speech energy between the two speaker units, a condition resulting in poor presence.

Where economy of investment and space are factors of uppermost importance, a fairly high crossover frequency, say around 800 c.p.s., would be satisfactory. But where equal consideration is given to each of the several factors mentioned, probably the best compromise will be a crossover in the neighborhood of 400 to 500 c.p.s.

Attenuation rate

The rate of attenuation at the crossover frequency is also dependent upon several factors. A very sharp cutoff is theoretically desirable for preventing excessive low-frequency energy from getting into the tweeter unit. It would also help to minimize objectionable irregularities due to response peaks in the speakers at frequencies beyond cutoff.

A sharp cutoff requires more filter sections and consequently more expense. It also causes greater power losses in the network



$$L1 = \frac{R_0}{2\pi f_c} \qquad C1 = \frac{1}{2\pi f_c R_0}$$

$$L2 = \frac{2R_0}{2\pi f_c} \qquad C2 = \frac{2}{2\pi f_c R_0}$$

$$L3 = \frac{R_0}{4\pi f_c} \qquad C3 = \frac{1}{4\pi f_c R_0}$$

$$L4 = (1+m)\frac{R_0}{2\pi f_c} \qquad C4 = (1+m)\frac{1}{2\pi f_c R_0}$$

$$L5 = \left(\frac{1}{1+m}\right)\frac{R_0}{2\pi f_c} \qquad C5 = \left(\frac{1}{1+m}\right)\frac{1}{2\pi f_c R_0}$$

INDUCTANCE VALUES IN HENRIES   CAP VALUES IN FARADS

Fig. 720-a, -b, -c, -d. *Design circuits and formulas for crossover networks.*

which must be compensated for by a larger amplifier. And finally, a sharp cutoff will increase the transient distortion. For these reasons it is generally accepted that an attenuation rate of at least 12 db per octave is desirable but that anything over 18 db per octave is excessive.

All two-way dividing networks are really combinations of a high-pass filter which feeds a tweeter and a low-pass filter which feeds a woofer, the two filters operating from a common source at their input terminals. These inputs may be connected either in series or in parallel.

Two basic methods of crossover network design are commonly employed. One utilizes a modification of conventional filter design while the other maintains a somewhat better impedance relationship and is therefore known as the *constant-resistance* method. Neither is overwhelmingly better than the other, and so our discussion will cover both of them.

Filter networks

When employing the filter method, it is necessary to follow special design procedures only for the first half-section of each filter. The succeeding sections (if any) are designed and constructed by the conventional methods described in Chapter 6.

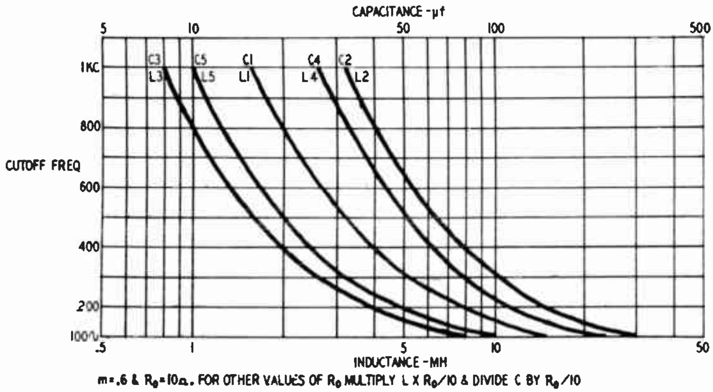


Fig. 721. *Family of characteristic curves of crossover networks.*

For the first half-section, it is essential that the m-derived type be used. As is customary for this configuration, m is assigned a value of 0.6. When the filters are connected in parallel they will have mid-shunt terminations, and when they are connected in series they will have mid-series terminations. Typical configurations for both full- and half-section dividing networks for both series and parallel operation are shown in Fig. 720. Design formulas are also shown for each of the filter components. When R_0 is assumed to be 10 ohms and m is 0.6, solving the equations for a set of values for the circuit constants will provide the curves of Fig. 721. When R_0 is some other value than 10 ohms, it is a simple matter to employ these curves and adapt the figures. To obtain corrected inductance values in this case, it is simply necessary to multiply each L by $R_0/10$, and for corrected capacitance to divide each C by $R_0/10$.

You will note from Fig. 720 that networks a and b are standard filter configurations, with a low-pass and high-pass filter in parallel in each case. Network a consists of full-T sections while network b comprises two L half-sections. Network c has two full π sections in a series configuration, while network d comprises two L half-sections.

When an attenuation rate of about 12 db per octave is satisfactory, the L-section filter is sufficient. When a rate of 18 db per octave is required, a full section will provide this. Thus networks a and c of Fig. 720 have attenuation rates of 18 db per octave,

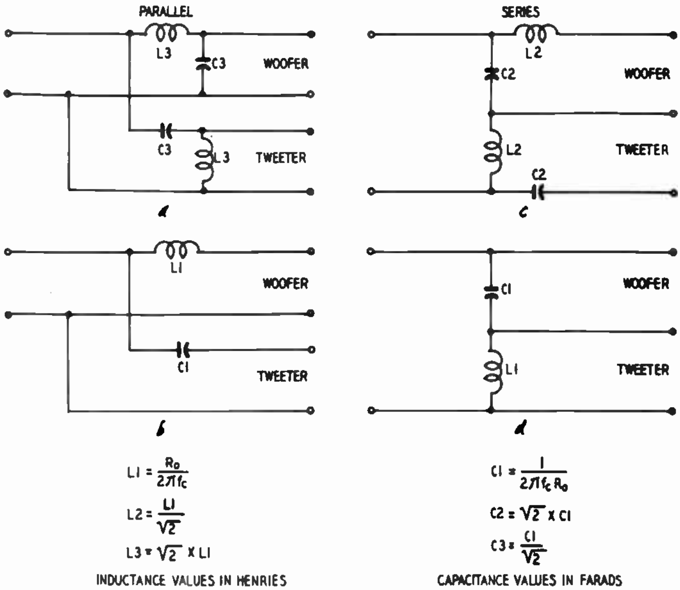


Fig. 722-a, -b, -c, -d. *Constant-resistance dividing networks.*

while networks b and d will attenuate at the rate of 12 db per octave beyond their cutoff frequencies.

Constant-resistance networks

The constant-resistance network is so named because it supposedly maintains a constant resistance characteristic at its input terminals. This is not strictly true, however, as the variation of its loudspeaker load impedance with frequency will reflect to its input, and in actual practice the input resistance is very little improved over that of the filter type.

Typical circuits of constant-resistance dividing networks are presented in Fig. 722. Note that the configuration of network a is the same as that of Fig. 720-b and that network c resembles that of Fig. 720-d. But it will also be noted from the design data that both the method and the result differ somewhat. A set of design curves developed from the formulas is given in Fig. 723.

A convenience in designing and constructing this type of network will be evident from examination of Fig. 722, wherein we see that all inductances in a given circuit are equal in value, as are all capacitances. The attenuation characteristics of this network method are not as steep on the average as those of the filter method. Networks b and d of Fig. 722 will attenuate at a rate of only 6 db per octave, while those at a and c will slope at 12 db per octave. No steeper rates are possible.

Resistance in the network

The curves of Figs. 721 and 723 are predicated upon pure reac-

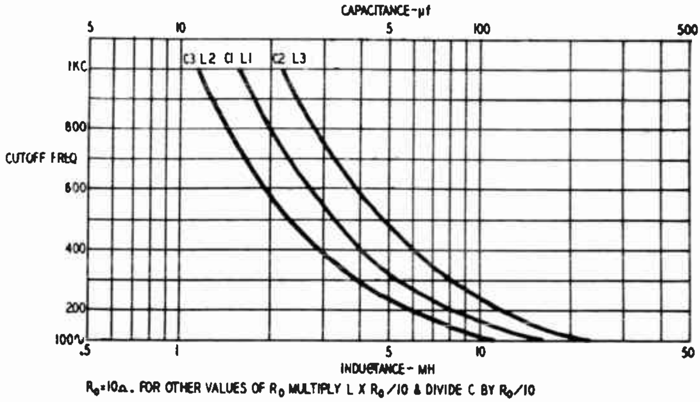


Fig. 723. *Design curves based on the networks of Fig. 722.*

tances, with no resistance whatever in the network. Such a condition does not occur in practice, of course, so the curves will be slightly inaccurate. This is particularly true with respect to the crossover frequency, which may be shifted a bit.

This is a minor difficulty however, compared to the relatively large amounts of power which are dissipated in the resistance of the network. In a 100-watt system, for example, a loss in the transmission band of only 1 db will result in a power loss of 21 watts. In practice the network loss can be held down to about 0.5 db, this being the minimum practicable value.

chapter 8

microphones

The microphone may well be regarded as the very fountainhead of the audio system, for it is at this point that sound first becomes electricity. If there is any considerable discrepancy in this translation from acoustical to electrical energy, the system is nearly defeated at the outset for any subsequent corrective measures that may be attempted will be only partially successful.

The modern microphone is an amazingly efficient and accurate device, but the fact remains that the perfect microphone is yet to be built. When it is, it will exhibit a perfectly flat frequency response over a spectrum which exceeds somewhat the range of the human ear. Throughout this range absolutely no distortion of any sort will occur. It will provide this distortion-free operation over a dynamic range of at least 60 db. If it is to be omnidirectional, it will exhibit no discrimination of any sort at any point around its polar axis. But if it is intended to be unidirectional, it will present exactly the same discrimination anywhere in its range. It will have a difference in front-to-back sensitivity of at least 18 db. It will generate sufficient output so that reasonable amounts of amplification will bring up its level to usable values. Its physical size and shape will be such that they will have a negligible effect in breaking up the sound waves which strike it. This, then, is the perfect microphone of the future. We don't have it yet, but we have had numerous approaches to it in the history of audio. Let's see what they are.

One of the oldest microphones, and the one probably most familiar, is the ordinary telephone transmitter, which is in reality a carbon microphone. It is of historical interest, however, that Alexander Graham Bell's original telephone of 1875 was *not* a carbon instrument but one which operated on magnetic principles very similar to those of today's dynamic microphone. It was the Blake transmitter, invented a year later, which is the basis of the *carbon microphone* utilized down to the present day.

The carbon microphone

This microphone operates on the principle that the vibration of its diaphragm can vary the intensity of an already existing electric current. Thus it does not generate a voltage of itself, as do

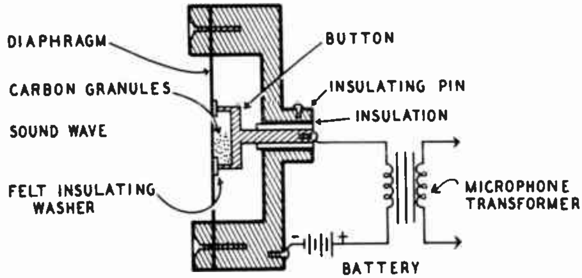


Fig. 801. *Cross-sectional view of carbon microphone.*

many microphones, but it is instead really a current modulator. The operation may be understood by reference to Fig. 801. The basic circuit components are a d.c. voltage source, usually provided by a battery, a cup of carbon granules known as a "button," a metal diaphragm and a transformer. The negative terminal of the battery connects to the diaphragm through its mounting ring. The diaphragm, in turn, rests against one side of the button; the other side is connected to the output. The circuit is then completed through the transformer primary to the positive terminal of the battery.

The rate of current flow through the microphone circuit will depend upon the battery voltage (usually 6 volts), the resistance of the transformer (very low) and the resistance of the carbon button (around 100 ohms). When no sound waves hit against the diaphragm, it remains stationary, the resistance of the button remains constant and a steady value of d.c. flows through the circuit. But when sound waves strike the diaphragm, it will vibrate in accordance with the variations in their intensity and frequency. This vibration causes a varying pressure to be exerted on the carbon granules, changing the state of their compression. As the compression increases, the resistance of the granules decreases, and the current therefore rises. Conversely, a decrease in compression will result in an increase in resistance, resulting in a decrease in current. The varying current is pulsating d.c., produced by the superimposition of the audio a.c. component upon the steady d.c. supplied by the battery.

Double-button carbon microphone

A more sensitive microphone operating on these principles is the *double-button* type, whose circuit is shown in Fig. 802. In this case one button becomes more conductive at the same time the other becomes more resistive. The output is thus effectively doubled when combined in a center-tapped output transformer. This is really a push-pull arrangement and as such has the characteristic discrimination against even-harmonic distortion. But despite these and other improvements, the carbon microphone still exhibits carbon resistance noise and for this reason is now seldom used except for telephone and communications applications.

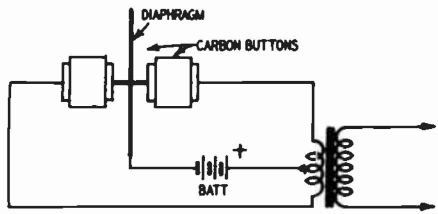


Fig. 802. *The double-button carbon microphone is characterized by a higher sensitivity than the single-button unit shown in Fig. 801.*

The crystal microphone

The *crystal* microphone does actually generate a voltage of its own, based upon the familiar piezoelectric effect. This phenomenon is exhibited as the tendency of certain piezoactive materials to distort physically when a voltage is applied across them and, conversely, to develop a difference of potential between their two faces when subjected to mechanical stress. Crystal microphones are generally of two types: the diaphragm unit and the *sound-cell* type. The diaphragm crystal microphone is shown in cross-section in Fig. 803. The crystal material normally used is Rochelle salt, and not the quartz crystal usually found in r.f. oscillators.

Two slabs, known as a *bimorph* element, are cut and assembled in such a manner that they will respond to sound-wave pressures.

Foil electrodes are applied to the outer faces of the element and between the slabs as well. When sound waves strike the diaphragm, the connecting pin causes the elements to bend and a voltage is thereby developed between the center and outer electrodes.

The sound cell unit employs two bimorph elements. These are mounted back to back, enclosed in a rectangular plastic frame and sealed in by two pliant membranes. In this system the sound waves strike the cell directly, thus eliminating the diaphragm and driving mechanism. By making this assembly small, its resonant frequency will be above the limits of audibility. The device then is pressure-actuated throughout the entire range, and with a little equalization the response can be made substantially uniform up to 15,000 c.p.s.

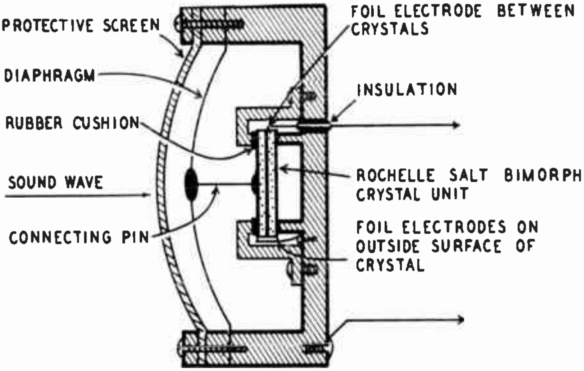


Fig. 803. *Crystal microphone in cross-section.*

The output impedance of the crystal microphone is quite high, which means that it will match directly the grid input circuit of the amplifier. This high impedance is sometimes a disadvantage in quality work for long runs of mike cable will be very susceptible to induced noise and an attenuated high-frequency response due to distributed capacitance between the conductors. The crystal is also extremely sensitive to mechanical shock and high temperatures, becoming completely inoperative at 130°F.

The dynamic microphone

The *dynamic* microphone, which has been with us since Bell's experiments in the 1870's, has been doing a workmanlike job for years and is now probably more popular than ever. This microphone, also known as the moving-coil type, utilizes a diaphragm, a coil and a permanent magnet, constructed as shown in Fig. 804.

The coil consists of a large number of turns of extremely thin insulated aluminum ribbon rigidly attached to the diaphragm.

The coil is arranged in a manner similar to that of a voice coil on a loudspeaker so that it will move between the poles of the powerful permanent magnet. When the diaphragm is actuated by sound waves, magnetic lines of force will be cut and a voltage induced in the coil. Since elementary a.c. theory tells us that an induced e.m.f. is proportional to the rate of motion through lines of force, we know that the output of the moving coil will be determined by its velocity. This, in turn, will be proportional to the pressure exerted on the diaphragm.

The velocity response curve of a basic dynamic microphone is far from flat, however, having a peak of 30 db or more around

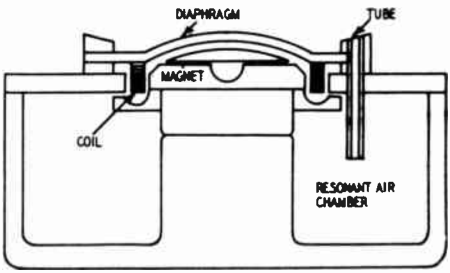


Fig. 804. *Dynamic or moving-coil microphone.*

1,000 c.p.s. This is flattened by the use of coupled acoustic networks which act as equalizers in the same manner as electrical networks. The size and shape of the several air pockets behind the diaphragm and inside the case are all deliberately designed as acoustic equalizers. There is also an acoustic tube whose opening meets the same wave fronts which strike the diaphragm. The movement of the diaphragm is then modified by the varying air pressures striking it front and back, with the result that the low-frequency response is extended another two or three octaves. When dialogue type equalization (bass rolloff) is desired, it may be obtained simply by plugging up this hole.

The moving-coil impedance of this microphone is quite low, which means that a matching transformer is necessary to couple it to the amplifier input. This has the advantage, however, that long runs of shielded cable may be used without adversely affecting the frequency response or picking up noise.

The modern dynamic microphone is considerably reduced in size as compared with its ancestors and as shown in Fig. 805-a, has excellent frequency response except for a slight droop at the low end. A typical response curve and directivity pattern of a leading commercial model are shown in Figs. 805-b-c. This unit employs a nonmetallic Acoustalloy diaphragm which, in addition to providing wide frequency response, will withstand high humidity,

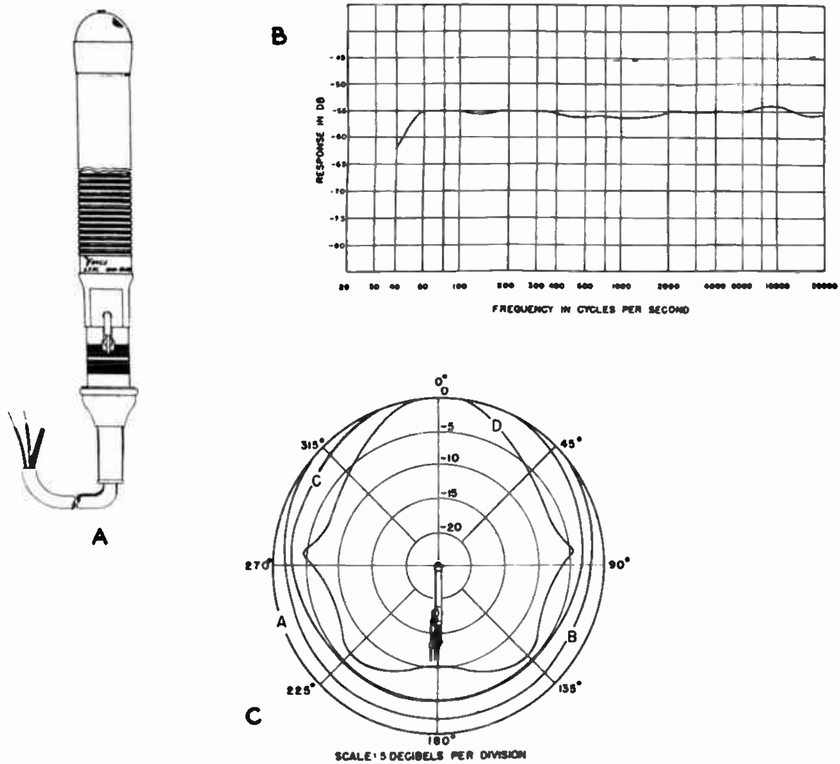


Fig. 805-a, -b, -c. *Dynamic microphone (a) has excellent frequency response (b). The graph (c) shows the directivity pattern.*

extremes of temperature, corrosive effects of salt air and severe mechanical shock. Its ruggedness and superior performance make it ideal for all but the most exacting applications.

Velocity microphone

An old standby of the broadcast and recording industry for high-quality musical pickups is the well-known *velocity* or pressure-gradient microphone. Unlike most other microphones, the velocity mike has no diaphragm. In this instrument the moving element is a thin corrugated Duralumin ribbon, suspended edgewise so as to be able to vibrate freely between the poles of a permanent magnet. The internal construction of this microphone is shown in Fig. 806.

Because of the lightness of the ribbon, its motion corresponds to the velocity of the air particles making up the sound waves. Since it has no diaphragm and is open in construction so that air flows freely through it, the velocity microphone is free from cavity resonance, diaphragm resonance and pressure doubling, all of which can cause undesirable peaks in microphone response.

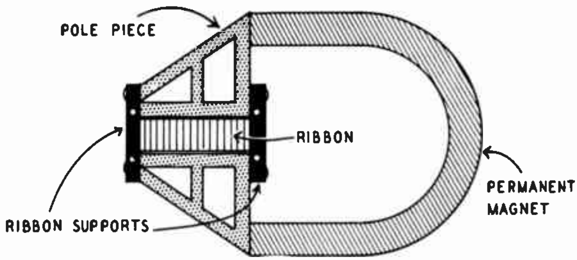


Fig. 806. *Velocity or pressure-gradient microphone.*

The ribbon does have a natural period of vibration, of course, but when this is designed to be lower than the lowest frequency to be reproduced, then the response of the ribbon is substantially independent of frequency. This is true until the frequency gets so high that the length of the path from the front to the rear of the ribbon approaches a quarter-wavelength, at which point phasing effects enter into the situation and the response tends to fall off. This can be seen in Fig. 807 which shows that the high end begins to droop slightly around 2,500 c.p.s. but this droop does not really become significant until around 12,000 c.p.s.

The ribbon is actuated by sound waves which strike either face, that is, which arrive at right angles to its surface. Voltages are generated when the wave is such as to form a pressure area on one side of the ribbon and a partial vacuum on the other. Thus, it will not respond to sound waves which arrive from a direction parallel to the plane of the ribbon. In this case equal and opposite pressures will be exerted on both faces of the ribbon simultaneously and it will remain immobile. Since the microphone responds to waves from front and back, but not from either side, its directivity pattern will be that of a *figure eight.*

Note that in the frequency response curves of Fig. 807 that two bass rolloff positions are available for speech equalization. This is possible because a coupling transformer is an integral part of this system, and a variation in frequency response is accomplished by altering the turns ratio with a shorting strap. The impedance of the transformer may be any value, but it is usually low (50 and 250 ohms) so that low-impedance cable runs can be made.

Although the ribbon is quite delicate and may be permanently misshapen by being dropped or hit by a heavy blast of air, the velocity microphone is otherwise rather rugged with its perform-

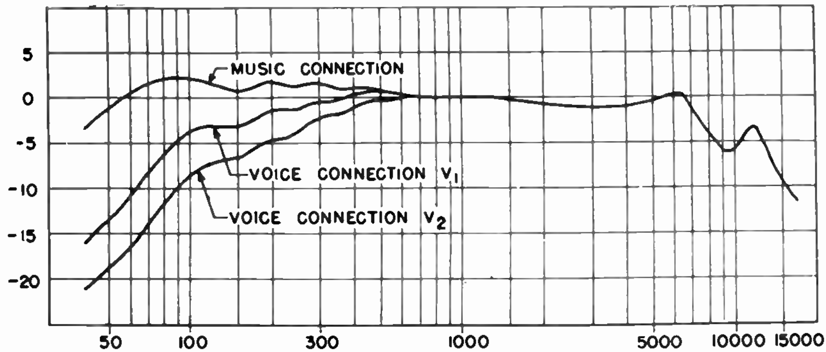


Fig. 807. *Frequency response curves of a typical velocity microphone.*

ance unaffected by temperature, humidity or variations in atmospheric pressure.

If it were possible to combine into a single microphone the perfectly nondirectional characteristics of one microphone and the perfect bidirectional (figure-eight) characteristic of another, the resultant directivity curve of the combination would be the unidirectional pattern of Fig. 808. Putting it another way, the geometric sum of a perfect circle and a perfect figure eight is a perfect *cardioid* or *heart shape*.

One type of microphone which achieves this characteristic is shown in Fig. 809. In this case a single ribbon is employed, but it is divided into two sections. One section, open to the air on both faces, operates as a typical bidirectional velocity unit. The other section has its back connected to an acoustic labyrinth and acts as an omnidirectional pressure unit.

A refinement of this system permits adjustment of the directivity characteristic to circular, figure eight, cardioid or anything in between. We see from Fig. 809 that the pressure unit is coupled to its acoustic labyrinth by a connector tube. When an aperture, placed in the labyrinth connector directly behind the pressure ribbon, is made variable in size by means of a rotating shutter, the directional characteristics of the microphone can be controlled.

When the aperture is so large that the back of the ribbon is effectively open to the atmosphere, the acoustic impedance of the labyrinth has been reduced to zero. Then this part of the ribbon is

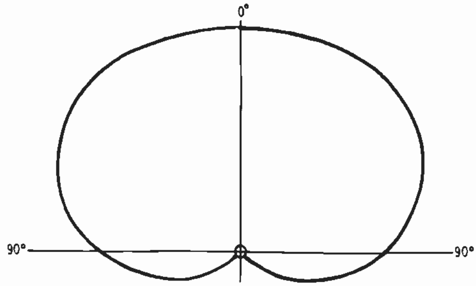


Fig. 808. *Graph of a cardioid characteristic.*

also a velocity unit and the microphone will exhibit the typical figure eight characteristic.

When the aperture is completely closed, the acoustic impedance is infinite and the typical nondirectional characteristic of a pres-

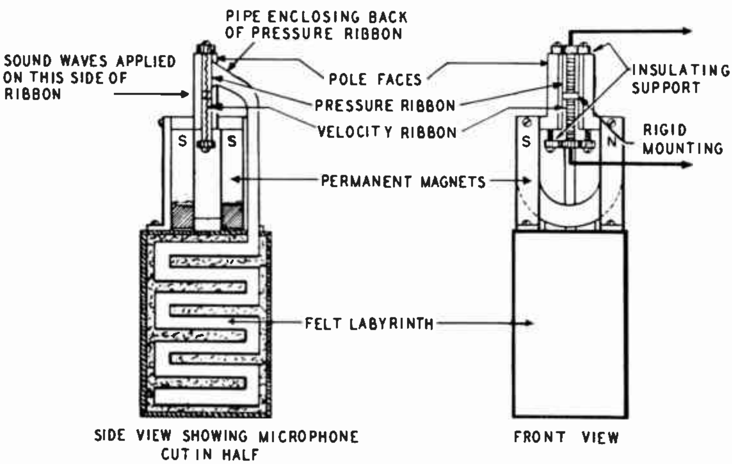


Fig. 809. *Modern microphone shown in cross section.*

sure-operated device is obtained. As the size of the aperture is varied between these two extremes, a critical value will be reached which will cause a phase shift resulting in the unidirectional cardioid characteristic. Other positions of the aperture shutter will result in patterns varying between bidirectional and nondirectional.

Frequency response characteristics of this microphone in two of its three main positions are shown in Fig. 810. Note that this instrument has two choices of dialogue equalization: one rolling off about 10 to 15 db at 60 c.p.s., the second attenuating from 17 to 22 db at that frequency.

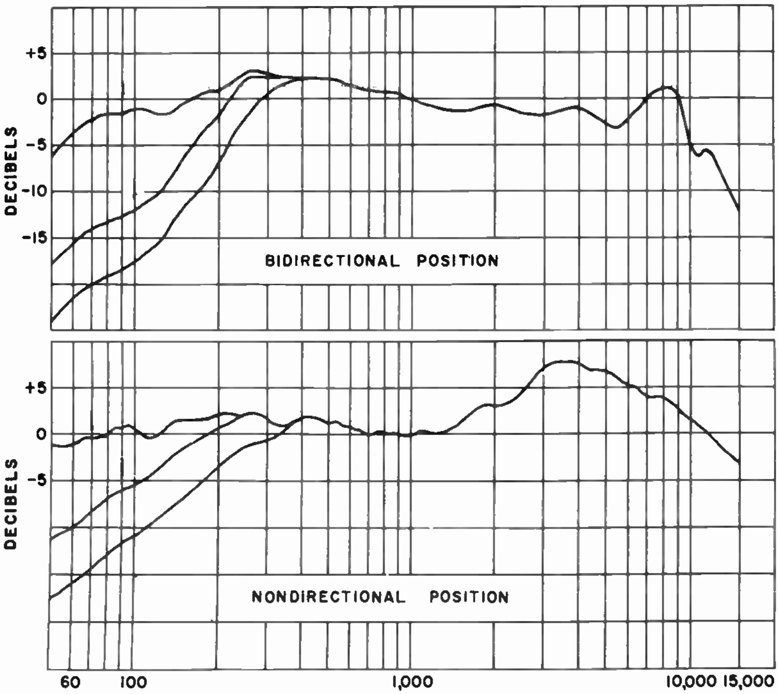


Fig. 810. *Several of the frequency response characteristics of the microphone illustrated in Fig. 809.*

Another approach to the problem of microphone directivity employs a pressure element which is a moving-coil dynamic unit, complete with diaphragm, and designed to be nondirectional at the lower frequencies but increasingly directional at the high end. This is housed with a conventional velocity-operated ribbon.

A number of characteristics (see Fig. 811) are provided simply by the setting of a selector switch. Selection in this case is accomplished by means of a potentiometer which determines the relative outputs of the two microphone components. If we designate the voltage developed developed by the pressure element as e_p and that of the velocity element as e_v, then the true cardioid is obtained when $e_p = e_v$, as in Fig. 811-c. When only the velocity unit is used, $e_v = 100\%$ and $e_p = 0$, in which case we obtain the characteristic of Fig. 811-a. At the other extreme only the pressure element is active, when $e_p = 100\%$ and $e_v = 0$, and the nondirectional characteristic of Fig. 811-b is obtained. At other proportions of e_v and e_p the characteristics of Figs. 811-d-e-f are possible.

Still another approach to unidirectional microphone design is the *mechanophase* principle, unique in that it requires only one moving unit instead of two. To develop this idea let us refer to Fig. 804 once again and imagine that the back is removed from

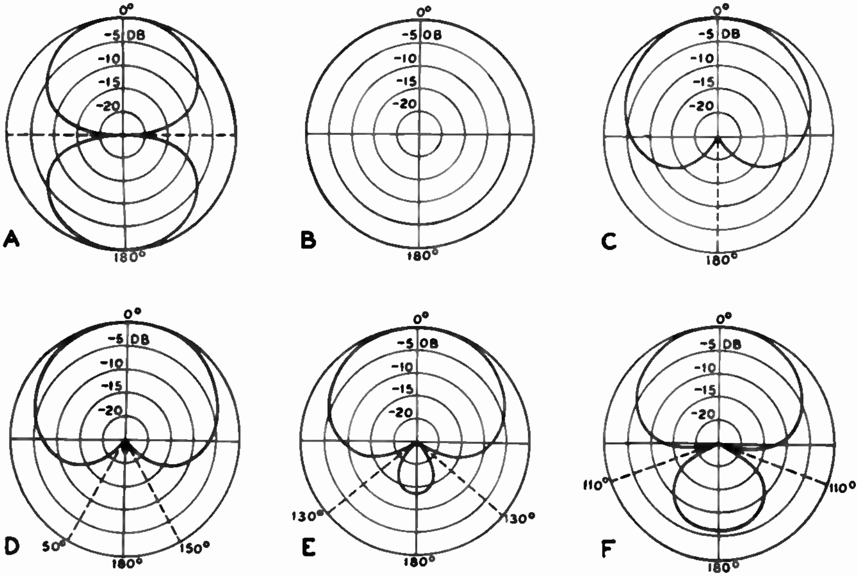


Fig. 811. *These illustrations (a to f inclusive) show the varying response characteristics that can be obtained with a modern microphone.*

that microphone in such a manner that the diaphragm is open to the air on both sides. The diaphragm can then respond to sound waves arriving from either front or back; that is, it will move when there is a difference in pressure between the air molecules at either face. Similarly when sound arrives from a direction at right angles to the plane of the diaphragm, it will impinge upon both faces with equal pressure and no movement will result. Under these conditions, therefore, the moving-coil microphone behaves exactly as a velocity or *pressure-gradient* unit.

When the back of the dynamic unit is closed in conventional fashion, then the rear face of the diaphragm is completely enclosed by the case, which presents an infinite impedance to sound. Fig. 808 again shows the cardioid result when the effects of these two responses are added in equal amounts.

Now if the acoustical impedance at the back of the diaphragm in the velocity condition is zero and if the acoustical impedance at the back of the diaphragm in the pressure condition is infinite, then the impedance for the condition of cardioid response must lie somewhere between these two extremes. The solution to this is provided by the arrangement shown in Fig. 812. The acoustic chamber has been moved from its customary place behind the diaphragm to a location below it. In the back behind the main

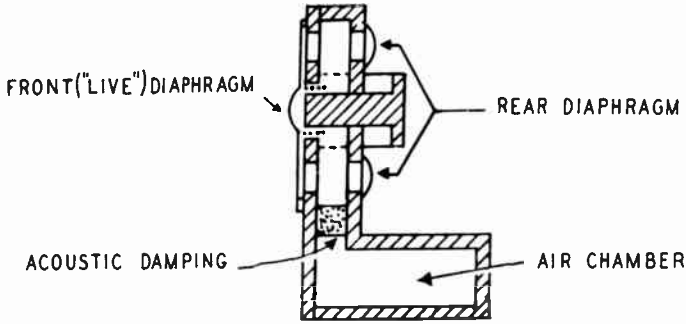


Fig. 812. *In this microphone the acoustic air chamber is below the microphone.*

"live" diaphragm are placed two small additional diaphragms.

If the two openings over which the small diaphragms are placed are closed tightly with a stiff backing, the microphone again becomes a nondirectional pressure type. And if the diaphragms are removed so that the back is open, then the microphone becomes a velocity unit. But if the mechanical impedances of the diaphragms are just right, the microphone becomes unidirectional and exhibits the characteristic cardioid response pattern.

A slightly modified application of this principle is seen in Fig. 813-a, along with its directivity characteristic in Fig. 813-b. The cardioid pattern is obtained through the use of three additional sound entrances located in the microphone case at different distances in back of the diaphragm. These three openings, each utilizing the correct acoustical impedance, combine to form one effective back entrance which varies in distance from the diaphragm inversely with frequency. The resulting phase and amplitude conditions produce a uniform cardioid pattern over a fairly wide frequency range.

The condenser microphone

The *condenser* microphone was used very extensively in the early days of broadcasting and electrical recording, then fell into disfavor and near-oblivion and now is once again riding such a wave of popularity that it is probably used more than any other type for the most exacting applications. Through the application of miniaturization techniques, many of its original shortcomings have been overcome.

In this type of microphone a diaphragm acts as one plate of a capacitor. Since capacitance is inversely proportional to the distance between the plates (thickness of the dielectric), it is obvious

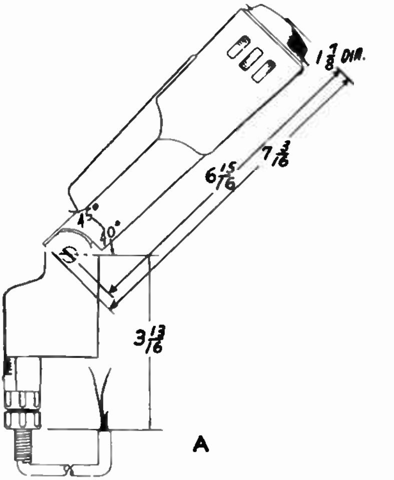
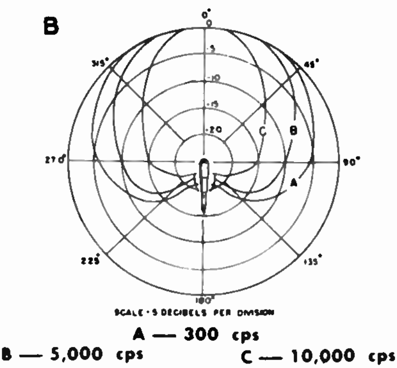


Fig. 813. *This microphone (a) has openings in the case arranged in such a manner that a cardioid characteristic (b) is obtained.*



that in an air-dielectric unit, the capacitance can be increased by moving the plates closer together. The condenser microphone depends upon this fact for its operation, developing variations in its capacitance caused by microscopic deflections of its diaphragm relative to a fixed plate. These deflections are, in turn, caused by pressure variations in the sound waves.

When a fixed voltage is applied to the plates, changes in the capacitance will cause minute charging currents to flow through the circuit from one plate around to the other. If a resistance is placed in series with the polarizing voltage, a varying drop of potential will appear across the resistor as the capacitance changes. The voltage thus produced will be a precise electrical replica of the mechanical movement of the diaphragm.

Because of the very small variations in capacitance, the size of the resistor must be very great for any useful voltage drop to be developed across it. This in turn means that the charging voltage must be even greater for a useful fixed potential across the plates. Even so, the output signal from the condenser head may be as little as −100 VU.

Due to the very minute size of the signal it is imperative that it be amplified immediately following the microphone itself. If any length of cable were to carry the signal before amplification, the distributed capacitance of the wire would become significant as compared to the microphone condenser and the signal would be literally engulfed.

In the early days of the industry problems such as these dictated a bulky, noise-susceptible unit which soon lost popularity despite its very excellent frequency response. But today we have the tiny unit of Fig. 814, which is only 3/8 long and 5/8 inch in diameter. Even with its cathode-follower preamplifier it is only about 6 inches long.

This microphone, dubbed the *Lipstik,* has a number of interesting features unique to condenser microphones. Heretofore, stretched membranes of various metals have been employed for the diaphragms, but this unit uses one having a core of glass. It is only 1/2 inch in diameter and .002 inch thick, with a conducting molecular layer of gold on one of its surfaces. No stretching is necessary because of the inherent stiffness and elasticity of the material.

Its small size creates a neglible amount of disturbance in the sound medium. The amplifier tube and all its associated circuitry (except the power supply) are contained in the small tubular case into which the condenser head is inserted. This feat is accomplished through the use of printed circuits, using the photoforming and etching process.

Which microphone?

The choice of a microphone for a given application will depend to a considerable extent upon the requirements of the job and the characteristics of the several mikes available. The inexpensive crystal microphone will be quite adequate for speech-frequency use in home recording, indoor public-address systems and amateur radio.

The dynamic microphone, due to its extreme ruggedness, is an excellent all-around mike for indoor or outdoor use in public address, sports broadcasts and all but the most exacting music pickups. Some commercial models are especially designed for close-talking speech applications, exhibiting a response from around 100 to 7,000 c.p.s. with a slightly rising characteristic.

The velocity ribbon microphone has excellent quality and is of particular value where its characteristic figure eight pattern is advantageous. It should not be exposed to winds for these will produce a low-frequency "putting" noise nor should it be placed too close to the sound source for then the bass will be overemphasized.

The unidirectional cardioid pattern of the dynamic-velocity microphone is especially useful in applications where there is a considerable amount of acoustic background noise or highly reverberant conditions. In such cases the back-wave rejection afforded by the cardioid characteristic is invaluable. It is adequate for most

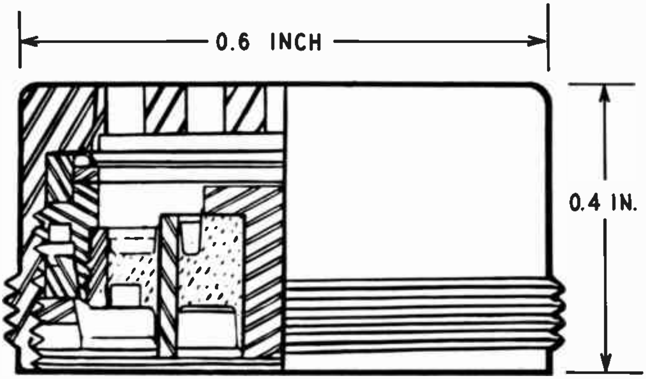


Fig. 814. *Extremely small size is unique feature of this condenser microphone.*

music pickups, especially for a solo vocalist who must be separated from the orchestra, but its overall quality doesn't quite measure up to that of the velocity instrument.

The condenser microphone exhibits the highest quality yet achieved in mikes and is used almost exclusively in highest-quality applications. Remember, however, that initial cost is considerably higher. Sometimes the only directivity characteristic available is nondirectional and an a.c. source must be available for the power supply to the condenser head and the preamplifier. Some of these microphones, however, will operate from a battery pack, although the values of the voltages are rather critical.

Microphone techniques

Not only is the choice of microphone important, but its placement with respect to the performers, and the performers with respect to each other, will make the important difference between a very good pickup and one which is just adequate, all other factors being equal.

For the highest quality results, the pickup, transmission and recording of sound from a studio, sound stage or remote location will depend upon a number of factors including:

1. The acoustical environment surrounding the sound source.
2. Characteristics of the electrical system, including microphones, amplifiers, equalizers and recorders.
3. Microphone techniques.

Factors 1 and 2 are covered elsewhere in this text, but at this moment we should discuss the very important matter of microphone usage. There appears to be considerable confusion and controversy about this subject, but the fundamental principles of a good pickup are quite straightforward and logical and when intelligently applied provide superior results.

For purposes of this discussion we are assuming the conventional monaural system and we must therefore bear in mind the fact that the system is unable to transmit the impression of the location of a sound, but can only provide certain audible clues as to the relative distance from the microphone of the various sound sources.

When sounds are emitted from a musical instrument or by the voice of a singer, the microphone is normally so located that the first sounds reaching it are those arriving directly from the source, and these in turn are followed rather quickly by reflections of the same sound which have bounced off nearby objects and the boundaries of the room. The time required for all these multiple reflections to be collected by the microphone and the values of their relative amplitudes and phases will determine the effective overall *reverberation* of the transmitted sound.

We know that all musical sounds are really very complex waves, containing many harmonics of various frequencies, phases and amplitudes. If two or more microphones are placed in the paths of these waves, no matter how these microphones are located, the waveforms of the sounds striking each of the mikes will differ somewhat due to slight discrepancies in arrival time and the fact that both sets of waves will not have traveled over identical reflection paths.

The basic idea of multiple microphone pickups is to provide greater presence, due to closer-in placement, and greater individual control over the several voices since a separate mike is often used for each section of the orchestra, with additional mikes for instrumental and vocal soloists. But no microphone yet built has the ability to pick up only the sounds desired from a given area while discriminating completely against all others. This means that some sounds from other areas will spill over to some extent into each microphone.

However, the time intervals will cause the several sounds to present waveshapes at each individual microphone which are dissimilar and which, when combined into a single composite wave in the mixer, will represent a considerably distorted version of the original waves in space. For this reason a single microphone properly placed, with a correct setup of performers in a good acoustical environment, will provide markedly superior results.

INSTRUMENT	PEAK POWER IN WATTS
FULL SYMPHONY ORCHESTRA	70
BASS DRUM OR TYMPANI	25
PIPE ORGAN	13
SNARE DRUM	12
CYMBALS	10
TROMBONE	6
PIANO	0.4
TRUMPET	0.3
BASS SAXOPHONE	0.3
TUBA	0.2
CONTRABASS	0.16
PICCOLO	.08
VIOLIN	.07
OBOE	.06
BASSOON	.06
GUITAR	.06
FLUTE	.05
CLARINET	.05
SAXOPHONE	.05
FRENCH HORN	.05
TRIANGLE	.05

Fig. 815. *Peak power in watts of musical instruments.*

It should be obvious from the foregoing that the single-microphone technique is not a cure-all, but only a readily available tool requiring intelligent usage. Assuming good acoustical conditions, it is still of primary importance that the performers be placed correctly with respect to the microphone, and the microphone in turn placed in a position which provides a pleasing relationship of direct to reverberant sound.

With small musical groups, up to a half-dozen or so members, the problem is a relatively simple one. Usually a circular grouping around the mike is best with a nondirectional microphone, with the distance from the mike to each performer determined and adjusted after aural checking on a high-quality monitoring system.

Fig. 815, which indicates the peak powers of a number of musical instruments, is useful as a starting point for the orchestral setup. We can begin with a very basic assumption that the instruments should be placed with respect to the mike at distances inversely proportional to their powers. This is true only in a very general way, however, for it is modified by other considerations.

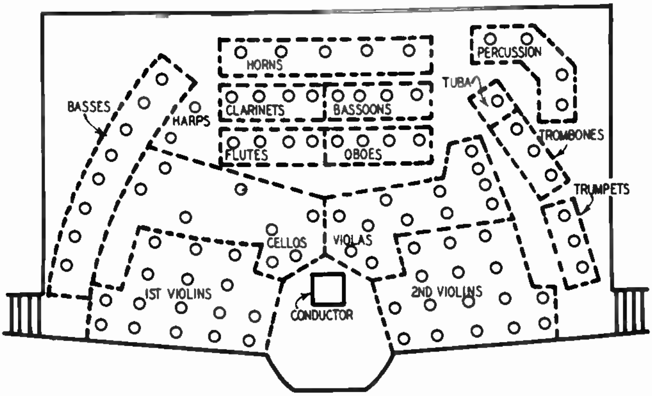


Fig. 816. *Arrangement of instruments in an orchestra.*

The range of some instruments is such that they have a more striking effect on the ear; that is, the ear is most sensitive in the range at which these instruments normally play. We must also consider the number of players in any given section, for the table of Fig. 815 tells us the peak outputs only of single instruments. It also is necessary to consider the characteristics of the music itself and the relative importance of the various voices. Since the violins most often carry the important line in conventional orchestra scoring, they will usually require more prominence than the power chart would seem to indicate. And finally we must consider the directivity of the instruments themselves as producers of sound. The strings, woodwinds and percussion instruments are practically omnidirectional, while the brasses project strongly in the direction of their bells.

Referring to Fig. 816, let us see how these principles have been applied in a high-quality professional setup. You will see that this arrangement develops quite logically from the principles just discussed, with the possible exception of the seating of the French horns, particularly since their bells will be facing downstage. But if this is the location which provides a balance most pleasing to the conductor and his listeners, then it is the correct one.

chapter 9

sound recording

Three basic methods of sound recording are in common use today. The modulated-groove engraving was invented by Thomas Edison in 1877. His original recordings were made on a cylindrical medium, but in 1888 Emile Berliner introduced the round flat disc which was the prototype of all present phonograph records.

In the early 1920's Dr. Lee de Forest conceived the idea of converting sound waves to variations in light intensity, which are then recorded photographically. This is the basis of the sound-on-film system which has been the standard of the motion-picture industry for many years.

The first magnetic sound recorder was patented by the Danish scientist, Valdemar Poulsen, in 1898. It is therefore not the newest of the recording methods, but it was the last to achieve widespread use and now appears to be the one holding the greatest promise for the future. Poulsen's medium was wire, which was not very satisfactory and which was probably largely responsible for the slow progress of magnetic recording. A most important advance, however, was made in Germany around the time of World War II, when magnetically coated tape became the vastly improved medium. Now in addition to wire and tape we also have magnetically coated motion picture film and even magnetic discs.

Disc recording

Disc recording ordinarily involves the engraving of a spiral groove on a soft material. Originally this medium was a specially compounded wax, but now the blank discs normally have a coating of a soft lacquer. Its basic ingredient may be either acetyl cellulose, ethyl cellulose or nitro cellulose. The first two are sometimes used for amateur recording discs, this fact having given rise to the misnomer "acetate" when referring to any instantaneous recording. Nitrocellulose is by far the best type recording lacquer yet discovered, but it is also highly inflammable when in the form of the fine thread "chip" removed in the process of cutting a record. For this reason special precautions are required in its handling and these improved discs are usually found only in professional installations.

The engraved groove on a record is a perfect spiral when no signal is being applied to the cutter, but when sound is actually being recorded the path of the groove will vary either from side to side or up and down. The latter motion is known as *vertical recording,* sometimes dubbed "hill-and-dale," and was the method originally employed by Edison. It has the advantage that overcutting from one groove into an adjacent one is much less likely, but processing for mass duplication has always been extremely difficult. It is therefore not used for commercial phonograph records, although some electrical transcriptions for broadcast station use still are cut vertically. The side-to-side motion of the recorded groove constitutes the *lateral* system, which is used almost exclusively at present.

The drive mechanism

All recording systems require that the recording medium itself be in motion, and a drive mechanism capable of imparting the desired movement is therefore an integral part of the system. In disc recording, the motive power is provided through a turntable which supports the disc and is driven by an electric motor. The quality of the motor and its associated drive mechanism can vary considerably, depending upon whether the turntable is of the very cheapest type common in portable phonographs or of the finest construction found in professional disc-recording machines and high-fidelity reproducing equipment.

Representative types of turntable motors are shown in Fig. 901. The two-pole motor has very poor speed regulation, causing wow and flutter in reproduction. It also radiates a powerful hum field. The four-pole motor is a considerable improvement, but it still exhibits substantial low-frequency rumble. The very best motor available today for recording applications is the hysteresis synchronous unit, which has a very large number of poles. This motor is inherently much smoother running than any other type known, having imperceptible wow, flutter and rumble.

In the better types of equipment a great deal of care is employed in obtaining isolation from vibration and in machining all moving parts to extremely fine tolerances, thereby reducing the possibility of speed variations. The weight of the turntable is usually supported by a circular shaft resting on a thrust bearing. This is a

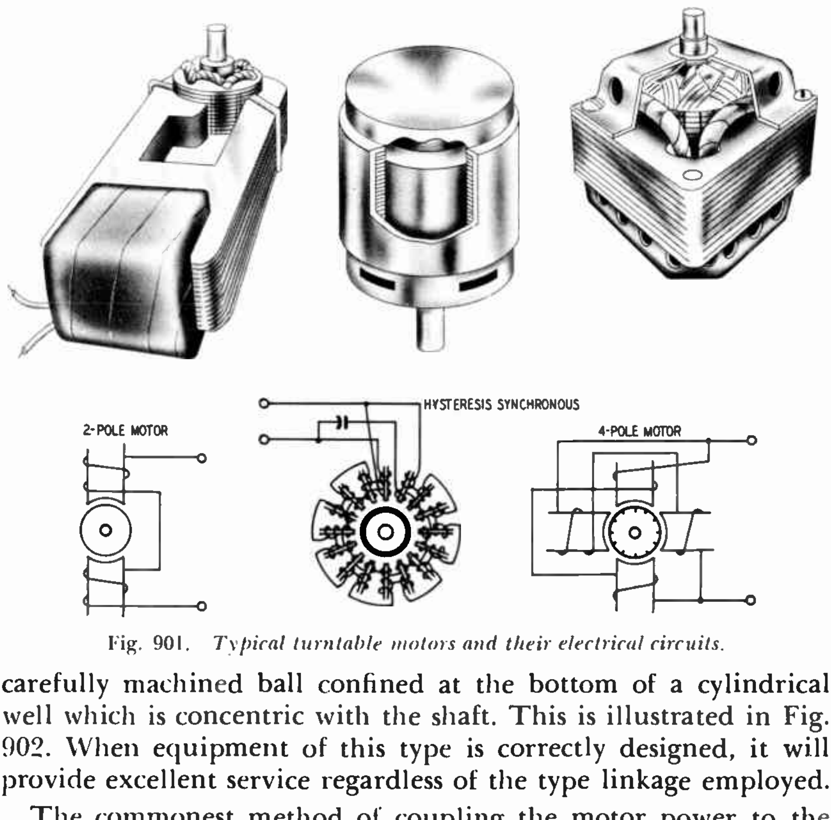


Fig. 901. *Typical turntable motors and their electrical circuits.*

carefully machined ball confined at the bottom of a cylindrical well which is concentric with the shaft. This is illustrated in Fig. 902. When equipment of this type is correctly designed, it will provide excellent service regardless of the type linkage employed.

The commonest method of coupling the motor power to the turntable in recording machines is shown in Fig. 903-a, and is known as *direct drive*. Since the amount of drag created by a cutting stylus as it plows a groove in a blank recording disc is many times greater than that of a playback stylus and pickup, it is important that the drive system be extremely powerful and not subject to slippage. These characteristics are possible through direct drive having one or more sets of reduction gears depending upon the number of speeds at which the recorder is to operate and in series with a mechanical vibration filter. This filter acts just like an electrical lowpass filter in that it attenuates all vibration at frequencies above the once-per-revolution rate of turntable speed. It consists of an arrangement of springs, dashpot damping and rubber coupling, which are analogous in their action to the appropriate values of L, C and R in an electrical filter.

The commonest type of coupling for reproducing turntables is the friction drive of Fig. 903-b. An idler wheel, usually of rubber,

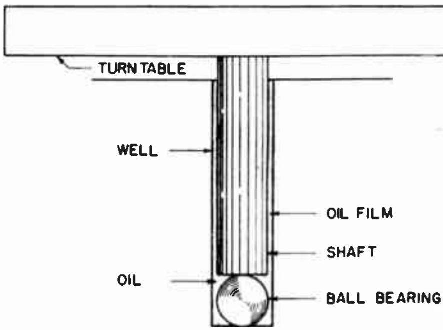


Fig. 902. *Turntable shaft rests on a ball bearing.*

transmits the motor rotation to that of the turntable rim. Since the power requirements for playback are not nearly so severe, this method is capable of excellent performance for its wow and flutter content can be held to exceptionally low values. In better equipment provision is made for disengaging the idler from the motor and turntable when not in operation. This prevents the rubber idler wheel from acquiring concave or flat spots on its perimeter, and thus a potential source of speed irregularity is eliminated before it can cause trouble.

The belt-drive system of Fig. 903-c also provides very good isolation between motor and turntable. Rather cumbersome for home installations, its use is confined to a few high-quality recording equipments.

The turntable itself, if correctly designed, will help in providing a smooth rotational movement for it can act as a flywheel and oppose any tendency toward speed fluctuation. A cheap turntable may be simply a flat flanged disc stamped out of a piece of sheet metal, but a high-quality table is made from a heavy casting, precision-machined for perfect balance and smooth fit of its moving parts. It may be constructed of iron, steel, aluminum, brass or an alloy. If it is to be used with cutters or pickups which operate magnetically, then the table itself should be of a nonmagnetic material. Otherwise the attraction between the table and the magnets in the cutter or pickup will seriously affect both the recording and reproducing operations.

The feed mechanism

The record cutting device is usually moved across the blank

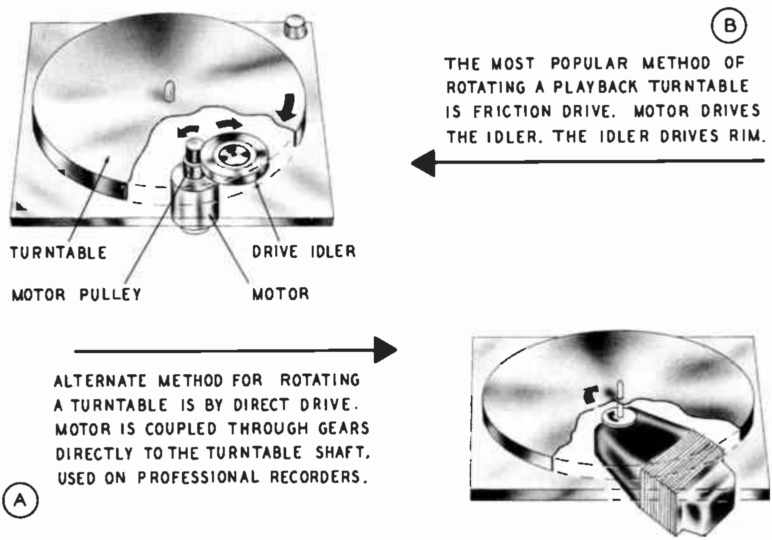


Fig. 903-a, -b. *Direct drive (a) and friction drive (b) systems.*

disc by means of a lathe mechanism in which the cutter is moved toward the center of the record at a predetermined rate by means of a lead screw. The number of grooves cut for a given part of the disc diameter is known as the *pitch* and is usually expressed in terms of lines per inch. This pitch is determined by the gear ratio in the linkage as well as the pitch of the lead screw itself. The pitch to be used when making any given recording will depend upon the playing time, the selected pitch being one which will utilize the maximum amount of useful disc surface. Practical values of groove spacing for 78-r.p.m. records will vary between 90 and 110 lines per inch. For 33⅓-r.p.m. transcriptions we sometimes go a little finer, perhaps up to 120 l.p.i. Microgroove recordings (LP and 45) may go up to 300 l.p.i. or even higher.

Tracking error

Since the cutting head moves across the disc in a perfectly straight line while the playback device moves in an arc, there is a certain amount of *tracking error*, as illustrated in Fig. 904. This is a source of frequency-modulation distortion and the generation of a large number of spurious sidebands. It is also a cause of improper seating of the stylus in the groove, resulting in excessive record wear. It can be minimized by the correct choice of an offset angle between the pickup and tone arm and by causing the playback stylus to pass over the record at a point somewhat beyond the center shaft. The distance between the center and the stylus path is known as "overhang." Optimum conditions of offset angle

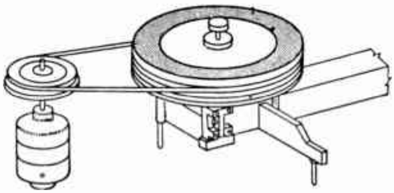


Fig. 903-c. *In the illustration at the left we have a simplified drawing of a belt-driven turntable. The advantage of such an arrangement is that the motor is isolated from the turntable.*

and overhang may be determined mathematically by use of a pair of rather involved equations. But since in any commercial tone arm the offset angle will be determined in advance by the manufacturer, it becomes a simple matter to follow his recommendations concerning mounting for the correct amount of overhang.

Recording and reproducing styli

Most better-quality recording styli are made of sapphire, while the best reproducing styli are diamond. It would be desirable to use diamond for cutting as well, but manufacturing a diamond cutting stylus is far more complex than making a reproducing needle. This can be readily appreciated by reference to Fig. 905 where we can see the differences in construction between recording and playback styli. While the shaping and polishing of the reproducing stylus is quite a formidable task, the making of a high-quality recording stylus is infinitely more difficult, demanding the utmost in lapidary skill. The overall size of the cutting portion is exceptionally small and the dimensions and angles of each facet are extremely critical for they determine the shape of the groove, as well as being an important factor in its noise characteristics.

The very small flat surface at the tip of the cutting stylus of Fig. 906 is known as a *burnishing facet*. This is very important in a stylus designed to cut lacquer for it is expected to polish the groove to smoothness as it cuts it. The older styli used for cutting wax were not dulled at the tip in this manner, the cutting face and

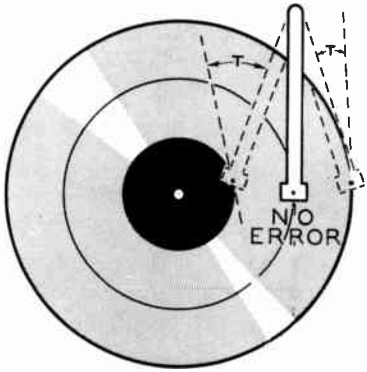


Fig. 904. *Error produced by angular motion of pickup arm.*

the heel converging sharply to form *the feather-edge* stylus. In current professional installations it is quite common to heat the stylus, making the lacquer much softer and more readily engravable. The feather-edge stylus is often used in this case.

There are no standards of groove size and shape but the grooves of Fig. 907, along with the reproducing styli to be used with them,

Fig. 905. *Difference in construction between recording and playback styli.*

are fairly representative of current practice. It is very important that the proper reproducing stylus be used for each size groove. A correctly seated stylus will be supported by the side walls of the groove, as shown in Fig. 908, riding at a level about midway between the top and bottom. When the point is so small that it drags on the bottom of the groove or so large that it rides on the top edge, much greater noise and distortion will appear in the reproduction.

All dimensions of Fig. 907 are given in inches, which means that the usual stylus for 78-r.p.m. record reproduction at *c* has a tip radius of .003 inch (3/1,000 inch or 3 *mils*). This is therefore commonly called a 3-mil stylus, while the microgroove stylus at *a* is referred to as a 1-mil point. The 2.5-mil tip is often found in

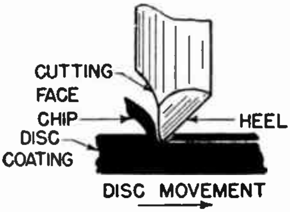


Fig. 906. *Drawing at the left illustrates a cutting (or recording) stylus. These are generally made of sapphire while the best reproducing styli are diamond.*

broadcast stations for use in the reproduction of electrical transcriptions but a pickup with 1.0- and 3.0-mil styli is quite adequate for most purposes.

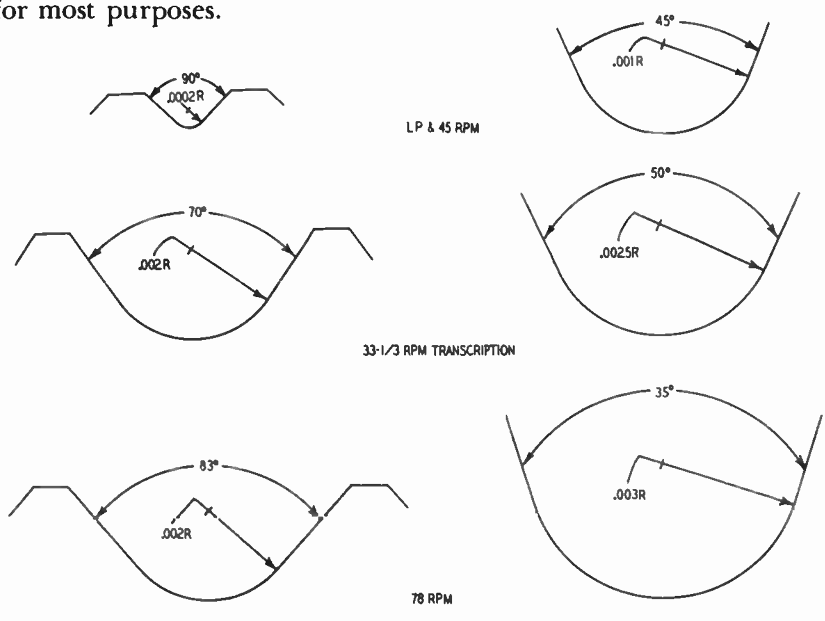


Fig. 907. *Record groove dimensions (left) and stylus dimensions (right). These drawings are not to scale, and are exaggerated for the sake of clarity.*

The cutter

The disc recording head transforms audio voltages into corresponding movement of a cutting stylus, which in turn engraves a modulated groove on the blank disc. Both magnetic and crystal cutters are used although the magnetic is far more common, particularly in professional applications.

The principles of operation of the basic magnetic cutter are illustrated in Fig. 909. A soft iron armature is pivoted with a cutting stylus attached to its lower end. The upper end is mounted between damping blocks of rubber, Viscaloid or some similar material which provides the restoring force to return the armature to a dead-center rest position. The audio signal is fed to the coil which surrounds the armature but is insulated from it. When current flows through the coil, the armature is temporarily magnetized with a polarity which will depend at any instant upon the

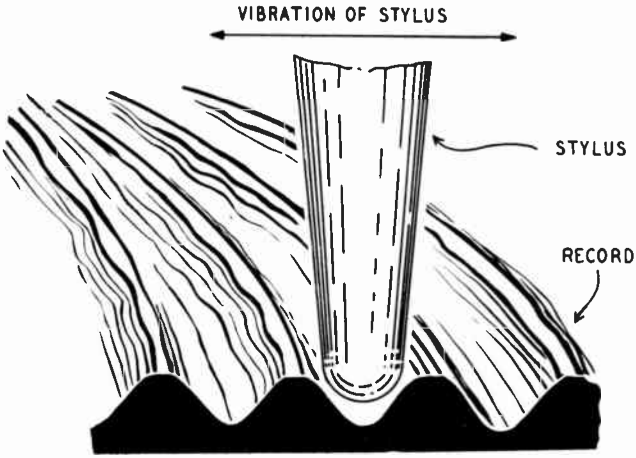


Fig. 908. *Stylus is supported by walls of the groove.*

direction of the current flow. There will then be a reaction between the upper pole of the armature and the field of the permanent magnet such that the armature will tend to rotate and move the stylus from side to side. Thus a laterally modulated groove will be cut in the disc, since the cutter armature behaves exactly like an electric motor which is repeatedly reversing direction. This motor analogy of cutter action is very important to our forthcoming discussion of recording characteristics and the meaning of the terms *constant velocity* and *constant amplitude.*

You will recall from our discussion of the crystal microphone that the so-called piezoelectric effect is actually a pair of complementary effects. In the first place, the physical distortion of a crystal will cause an e.m.f. to be developed between its faces. This principle is the basis of the crystal pickup and crystal microphone. Conversely, when a voltage is applied across the faces of a crystal, it will tend to alter its shape. This fact forms the basis of the operation of a crystal cutter.

The crystal recorder consists of a conventional Rochelle-salt bimorph element fastened into its mounting by a clamp at one end. To the free end is attached the stylus. Damping blocks are clamped against the sides of the crystal to reduce the effects of resonance.

When an audio voltage is applied across the cutter, it will tend to twist at its free end and thereby swing the stylus from side to side. The direction and amount of displacement of the stylus at any given instant will depend upon the polarity and the magnitude of the driving voltage.

This means that, as the voltage is increased, the stylus displacement will also increase in direct proportion, regardless of the signal frequency. Thus if a given signal voltage at 1,000 c.p.s. will move the stylus from side to side for a given distance, the same amount of voltage will produce the same amount of displacement whether the frequency be 2,000, 5,000 or 10,000 c.p.s. In such a system, where the amount of stylus displacement is directly proportional to the magnitude of the driving voltage, the cutter is said to have a *constant-amplitude* characteristic. Now if the cutter must swing the same distance for a given voltage, regardless of frequency, as the frequency increases the stylus must obviously go through many more reversals of direction in a given period.

Constant-amplitude recording

In other words, as we apply a constant voltage to a crystal cutter, it will cut a groove which undulates from side to side by a definite amount. Then as we increase the frequency of the exciting voltage, the stylus will still swing the same distance from side to side, but it must move from peak to peak more times in every second. With increasing frequency, then, the overall length of the path described by the stylus also increases. Now since distance equals rate multiplied by time, it is obvious that a greater distance in a given time can result only from an increased rate of speed. And so it is with the crystal cutter; while it has a constant-amplitude characteristic as frequency varies, the *velocity* must necessarily increase proportionately with frequency.

To reproduce a signal inscribed in a constant-amplitude groove, we must have a device whose output voltage is directly proportional to stylus displacement, regardless of the frequency. The perfect crystal pickup would meet these requirements, but a typical magnetic cartridge definitely would not. Let us consider why this is so.

You will recall that we compared the magnetic cutter to an electric motor. Let us analyze this a little further and determine just how the operations are similar. When we apply a battery voltage to a d.c. electric motor the motor will rotate in a given direction and at a fixed speed. If we insert another battery in series with the first, thus increasing the voltage, we also increase the rate of speed of the motor. It is, therefore, the motor *velocity* which is directly proportional to the magnitude of the applied voltage.

Constant–velocity recording

The direction in which the motor rotates will depend upon the polarity of the voltage, for if we reverse the leads from the battery,

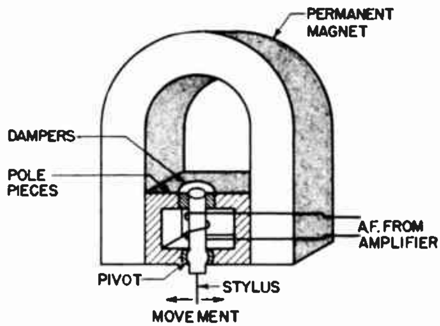


Fig. 909. *Chief components of a magnetic cutter.*

the motor will also reverse direction. Now we know the factors which determine velocity and direction, but you will note that absolutely no mention has been made of *amplitude,* which is the distance of the movement. It is obvious, however, that a motor will continue to move just as long as a voltage is applied to it. The same is true of a magnetic cutter, at least up to the point where further movement is physically blocked by dampers or pole pieces. Thus we can say that the stylus displacement in a magnetic cutter will continue right up to the point of the mechanical limit of the cutter or until the polarity of the voltage reverses or until the voltage is removed altogether. In actual practice, of course, the signal level will be so adjusted that reversals in polarity will be reached before the armature strikes its mechanical limit. And since the speed at which this displacement occurs is directly proportional to the magnitude of the applied signal voltage, the magnetic cutter is said to be inherently a *constant-velocity* device.

Now since the velocity remains constant for a given signal level, regardless of the frequency of that signal, it is obvious that the stylus will be unable to travel as far when the armature is excited by a high-frequency signal as it will when the voltage is of comparatively low frequency. From this we can infer that *under constant-velocity conditions, the amplitude of stylus displacement is inversely proportional to the magnitude of the signal voltage.* This condition would be quite acceptable so long as our reproducer had an identical constant-velocity characteristic. That is, the constant-velocity reproducer would provide an output voltage which was determined by the rate of cutting magnetic lines of force. It would therefore be acting like a generator and its output amplitude would be directly proportional to the velocity of stylus movement.

But consider what happens when we attempt to reproduce with a constant-amplitude playback cartridge a record made under constant-velocity conditions. The output signal of this reproducer will be dependent upon the amplitude of stylus displacement. But we have already noted that stylus displacement in a constant-velocity recording system is inversely proportional to the signal voltage. Thus the output signal of a constant-amplitude reproducer will decrease as the frequency is increased. And since there is an inverse relationship between the two, doubling the frequency will halve the displacement and therefore the signal level as well. From this we can generalize that, when a constant-velocity recording is reproduced with a constant-amplitude pickup, for each increase in signal frequency of one octave the signal voltage at the output terminals will be reduced by one-half.

Constant amplitude vs. constant velocity

A voltage ratio of 2 to 1, however, is (under equal-impedance conditions) equal to 6 db. And this fact leads us to the conclusion that *the difference in stylus displacement for constant-amplitude and constant-velocity recording are such that, for any given pickup, the output signals from the two types of recordings will differ by a constant slope of 6 db per octave.*

This can be better understood by reference to Fig. 910, which indicates the behavior of a perfect constant-amplitude pickup in translating into voltage the groove displacements of recordings made under both constant-amplitude and constant-velocity conditions. When reproducing a recording having the constant-amplitude characteristic, this pickup will, of course, produce a constant voltage output throughout the frequency range. But if the constant-velocity groove is such that it causes the production of an equivalent voltage at 1,000 c.p.s., then the voltage an octave below at 500 c.p.s. will be 6 db higher while that at an octave above (2,000 c.p.s.) will be down 6 db. Furthermore, this 6-db-per-octave relationship will continue throughout the range.

Now let us consider these same two recordings when both are reproduced by a perfect constant-velocity pickup. (See Fig. 911.) It is the constant-velocity recording characteristic which produces a constant voltage at the output of this pickup. And the output when reproducing the constant-amplitude characteristic increases with frequency, but still at a fixed rate of 6 db per octave.

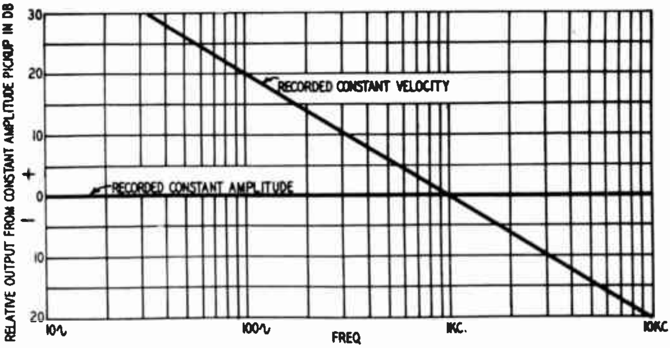


Fig. 910. *Output of constant-velocity reproducer decreases 6 db per octave when used with a record made under conditions of constant amplitude.*

Now the graphs of both Figs. 910 and 911 tell precisely the same story. They differ only by virtue of their frames of reference, the curve of Fig. 910 being predicated upon a constant-amplitude reproducing characteristic while the information of Fig. 911 is presented with reference to constant-velocity reproduction. Since either system works equally well, it is important that one or the other be accepted as standard to avoid serious confusion. The conventional method of indicating recording-system response is relative to the constant-velocity reproducing condition, as shown in Fig. 911. Any recording characteristic curve, then, will actually show the effect that such a record will have upon the output voltage of a perfect constant-velocity reproducer. This is a very important consideration as we discuss the actual recording characteristics encountered in commercial practice. These are really composite characteristics, being at some times constant velocity, and at others, constant amplitude.

Perhaps the simplest reason for the constant-velocity reference in recording characteristic curves lies in the fact that most recorders have approximately this characteristic inherently. But since in this method of recording the displacement of the stylus varies inversely with frequency, the displacement can become exceedingly large at the lower frequencies, with consequent overcutting into adjacent grooves. This situation results in distortion, pre-echo effects and simply refusal of the reproducing stylus to track the groove.

It is therefore common practice to attenuate the stylus displacement at the lower frequencies downward from between 300 to 600 c.p.s. at a rate of about 6 db per octave. Below this crossover frequency the recording characteristic is altered to something very close to constant amplitude.

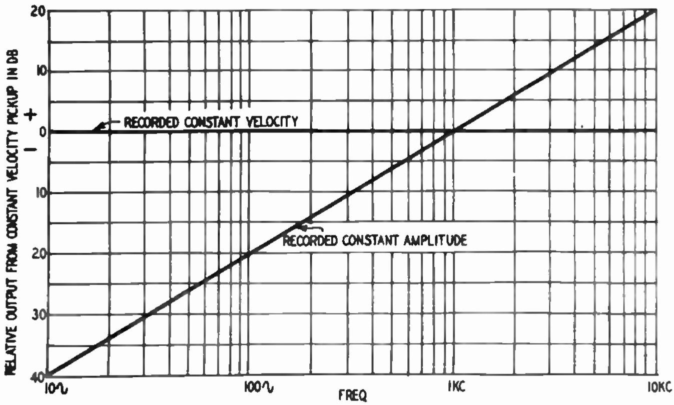


Fig. 911. *This curve also shows a 6 db slope per octave. Compare this with Fig. 910.*

Now we have a recording characteristic which is constant-amplitude from the bottom of the range up to the crossover frequency around 500 c.p.s. and then becomes constant velocity from this point to the top of the range. But the inverse relationship between displacement and frequency bobs up to plague us again at the high end, for displacement then becomes so slight that the signal-to-noise ratio is very unfavorable. To overcome this we insert a pre-emphasis "tip up" characteristic beginning around 2,000 to 5,000 c.p.s. and moving to the top of the range at a rate of about 6 db per octave. Thus we have in effect reconverted our recording characteristic to constant amplitude in the upper region.

At this point we have a characteristic essentially constant amplitude in the bass region, constant velocity in the mid-range and constant-amplitude again at the high frequencies. Typical recording characteristic curves, most of which fit this pattern, are shown in Fig. 912. Note that these curves are somewhat idealized as no cutter or equalizer will exhibit such perfect straight-line response and sharp-corner angles as are shown here. They do, however, show clearly the intended crossover frequencies and pre-emphasis characteristics of a number of commercial products.

It is interesting to note that the lower bend of the curve is always described in terms of the crossover frequency, where the characteristic changes from constant amplitude to constant velocity. But oddly enough the upper bend is defined in terms of the amount of deviation from the constant-velocity characteristic at 10,000 c.p.s. The old European characteristic, for example, has

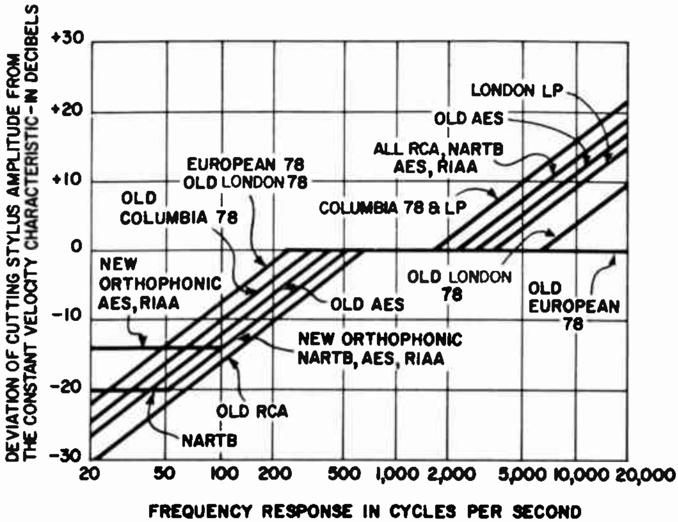


Fig. 912. *Typical recording characteristics.*

no tip up at the high end, but remains constant velocity right out to the end of the range. The NARTB curve, on the other hand, becomes constant amplitude at such a point that it will produce 14 db more signal in reproduction at 10,000 c.p.s. for the same value of voltage to the recorder. That particular curve, then, would be described as having a 500-c.p.s. crossover and a 14-db tip up.

A closer inspection of the bass region of the curves in Fig. 912 will show that a couple of these characteristics return to constant velocity once more in the region of 50 to 100 c.p.s. Since it is characteristic of constant velocity that groove displacement increases with decreasing frequency, this low-frequency "shelf" provides the equivalent of a small amount of bass boost at the lower extremity of the range. This has become necessary as the bass response of reproducing equipment has proceeded more rapidly than the design of turntables, most of which exhibit some rumble in this region. The purpose of the low-frequency shelf, then, is to improve the signal-to-noise ratio in the vicinity of the rumble.

Differences in characteristics

After considering the very many curves of Fig. 912, the question may quite naturally arise as to the reasons for such multiplicity. In the beginning, of course, there was hardly any other choice. A horn was designed to collect and record sound by purely mechanical means and whatever its response characteristics happened to be these were the standards for a given company. As elecirical

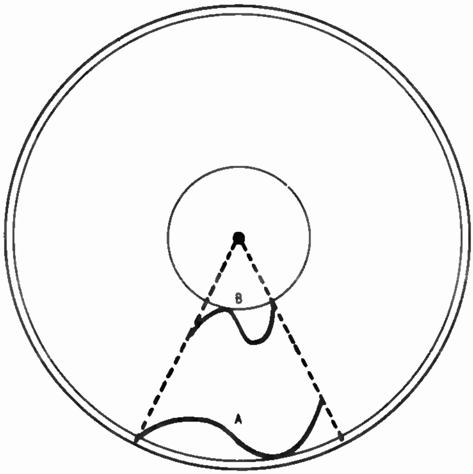


Fig. 913. *The angular distance of groove A is identical with that of groove B.*

recording achieved prominence, the recording equipment was generally of higher quality and greater flexibility than the reproduction equipment upon which records were played. During this era the recording engineers simply assumed the average characteristics of home instruments of the time and then designed a recording characteristics which would produce a record which sounded most pleasing, even though it may not have been a very accurate complement of the reproducing characteristic.

There were as many opinions about this matter as companies in the field, as may be readily deduced from Fig. 912. Furthermore, these characteristics were regarded as trade secrets to be jealously guarded from one's competitors. As a consequence it is still necessary, for the most accurate reproduction, to have a reproducing compensator on the playback system which will provide characteristic curves complementary to all of those of Fig. 912. Only recently have many of these characteristics been published, and only since then has there been any concerted effort at establishing a mutually acceptable standard. Various professional groups, including the Audio Engineering Society, National Association of Radio and Television Broadcasters and the Record Industry Association of America, have been working diligently on this problem. But it will be many years before the effects of the earlier short-thinking policies are completely dissipated.

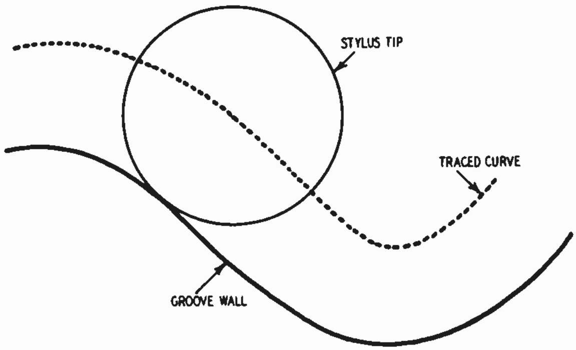


Fig. 914. *The pickup cannot always follow the exact curve of the recording stylus.*

The manner in which recorders and pickups are made to exhibit the desired characteristics will vary with the individual equipment. There are no cutters and no pickups which are perfectly constant velocity or amplitude, and they might therefore said to be somewhat equalized already due to their mechanical characteristics. The balance of the equalization must then be provided by an electrical network at the appropriate place in the audio system. It is apparent that all recording and reproducing characteristics are composites, consisting of the inherent mechanical characteristics of the cutter or pickup, to which is added the correct electrical network to provide the desired overall result. The network must be designed specifically for the cutter or pickup with which it is to work and should not be used indiscriminately with equipment of other types.

Diameter equalization

A problem peculiar to disc recording is the manner in which the frequency response varies at different diameters. This may be understood by referring to Fig. 913, which presents a highly exaggerated picture of two grooves, both of the same frequency and amplitude, but one lying at the outer edge of the record while the other is close to the lead-out groove. Under these conditions and with the disc rotating at a constant speed, the stylus will require precisely the same amount of time to trace groove *A* as it will groove *B*, since both of these grooves constitute the same part of a revolution. But the amount of distance traveled in tracing groove *A* is much greater than that in tracing groove *B*. Hence groove *A* covers a comparatively long distance and its displacement occurs rather gradually. But in groove *B* the same cycle must be traced

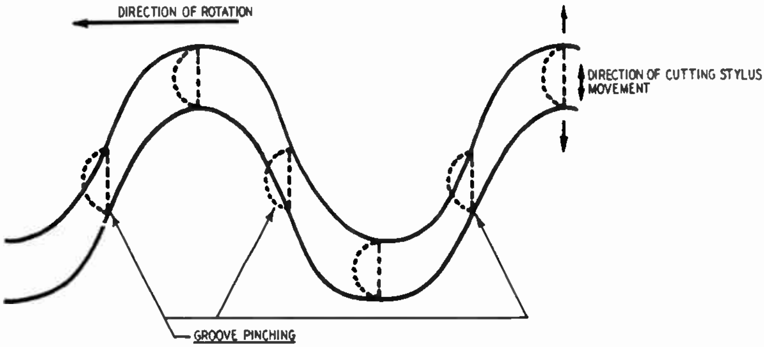


Fig. 915. *Illustration of pinch effect.*

in a much shorter distance, with much steeper undulations and much more acute bends. This condition becomes more pronounced with higher frequencies, smaller diameters and lower rotational speeds. Ultimately it reaches the point where the groove wavelength approaches the tip radius of the reproducing stylus. At this point the high-frequency response becomes rather severely attenuated.

This difficulty is sometimes partially compensated for by more alteration of the characteristic by means of a method known as *diameter equalization*. This is a variable equalizer which is coupled to the lead screw or other mechanism which drives the cutter across the disc. The equalization is automatically varied as the inner diameter is approached, with additional high-frequency accentuation to overcome the tracing losses.

Disc distortion

Other types of distortion in reproduction are due to the differences in shape between recording and reproducing styli, as shown in Fig. 905. The cutting stylus is really a *V*-shaped chisel, while the reproducing stylus tip appears as a spherical surface to the walls of the record grooves. The pickup cannot possibly trace the cutter action precisely, as may be clearly seen from Fig. 914. Since the shape of the traced curve is not quite the same as that of the groove itself, we must conclude that some distortion is present. The effect here is one of harmonic distortion and, as in the case of frequency distortion due to tracking, the problem is most pronounced at the lower speeds and inner diameters.

Since the recording stylus does not face into the groove when it is being modulated, it will tend to cut a narrower groove on steep wave fronts. This condition is known as *pinch effect,* (Fig. 915). As the groove becomes narrower, the reproducing stylus has

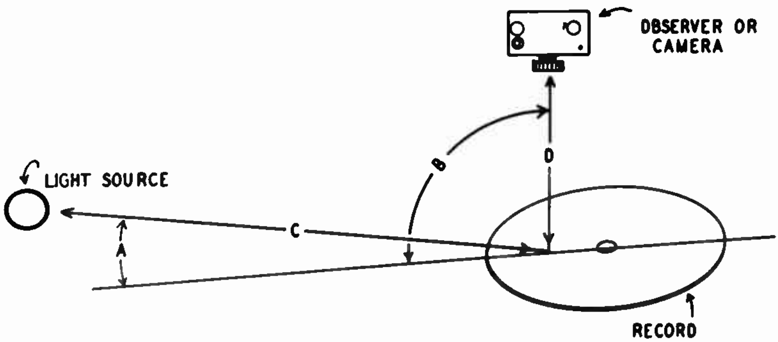


Fig. 916. *Setup for visual groove inspection.*

no choice but to rise up and then fall again as the groove widens. This vertical stylus motion will cause a voltage to be generated in most pickups; a spurious addition to the signal. Since the pinching occurs twice in each alternation, the vertical motion will be at a rate just twice that of the desired lateral motion. Hence it will produce a signal quite rich in second-harmonic distortion.

Visual groove inspection

A great deal may be learned by the practiced observer about the characteristics of a record simply by means of visual inspection. The grooves of a well-cut disc will appear shiny when observed under a single light source. If they do not, this is an immediate indication that either the stylus or the disc material is defective. In either case the disc will be noisy and generally of poor quality.

But if a good shiny groove is present, considerable information, including the recorded level, and by inference the frequency response may be obtained by observation. This information can be obtained from the light patterns reflected from the groove walls.

The simple setup employed for such examination is shown in Fig. 916. The disc is illuminated by a concentrated light source, preferably a clear glass bulb, set at a distance *C* of 6 to 8 feet. The angle *A* between the record plane and the light path is small, usually between 5° and 15°. The purpose of this is to cause the light to be reflected by the groove walls, but not to strike the bottoms of the grooves. The distance *D* between viewer and record surface is also several feet. The angle *B* between the line of sight and the record plane is that which provides the best reflected pattern and will usually be between 75° and 90°.

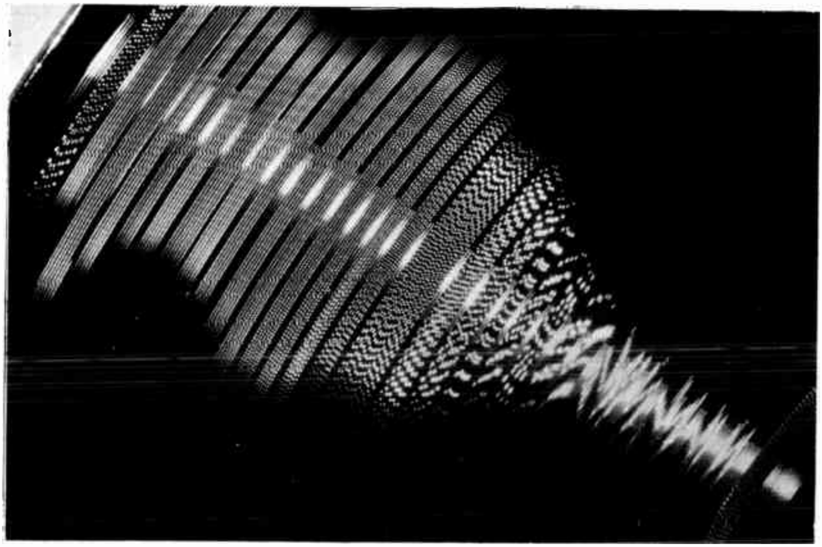


Fig. 917. *Reflected light patterns of a frequency record.*

Under these conditions a reflected light pattern such as that of Fig. 917 will be observed. This is a photograph of a disc cut with a number of grooves at frequencies ranging from 50 to 10,000 c.p.s., with the uncut blank disc between each band of modulated grooves. The rather remarkable thing about this pattern is the fact that the width of the reflection of each band is directly proportional to the velocity of the modulation. The reasons for this relationship may be understood by reference to Fig. 918.

This illustration depicts the action of the rays from the light source as they strike and are reflected by the opposite wall of a groove. The angle of reflection is of course equal to the angle of incidence, which means that the largest angle between incident and reflected rays will occur at the steepest curvature of the groove. This occurs at points *A* and *B*, where the groove path crosses the line X—X[1]. This line is the path which the groove would follow if no modulation were present and no displacement occurred. Hence the steepness of the groove slope is greatest when it is crossing the no-modulation path. This is analogous to the electrical wave whose rate of change is greatest when its curve passes through zero.

The steepness of the curve described by the modulated groove will be directly proportional to the stylus velocity. If means were available to measure this steepness, we should have a clue to the

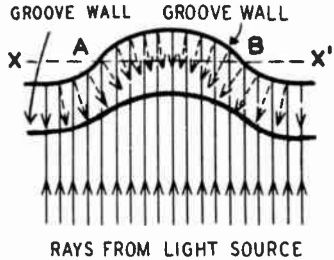


Fig. 918. *The drawing at the right helps explain the reason for the reflected light patterns of Fig. 917. The greatest angle between the rays from the light source and those that are reflected will occur at the steepest groove curvature.*

modulation velocity. Let us consider now an actual situation in which the groove, of which Fig. 918 is only a small part, is extended all the way around the record a number of times. With the light rays striking the record from a point source, only a certain definite number of them will be reflected into the line of vision of the observer. And since the greatest angle between incident and reflected light occurs at *A* and *B* and all similar points on the groove, it will be these points which outline the extremities beyond which light is not reflected to the eyes of the observer. With this pattern therefore having a definite width, it is possible actually to calculate the peak groove velocity by means of the following mathematical relationship:

$$V = \pi n W \left(1 + \frac{kR}{D} \right) \qquad (46)$$

where n is the disc speed in revolutions per second, .W is the width of the pattern, R is the radius from record center to groove, D is the distance from the record to the observer and k is equal to 1/ (1 + cos B).

Without going into specific values for a given record, it is still quite useful to analyze the performance of a cutter by visual inspection of a test record it has made. For example let us analyze the record of Fig. 917. Recalling that the width of the light pattern is proportional to the velocity, it would appear that this record has a constant-velocity characteristic from 10,000 c.p.s. down to a crossover frequency of between 500 and 1,000 c.p.s.

Below this point the velocity decreases at a rather constant rate, which would seem to indicate a typical constant-amplitude characteristic. We can also see at once that there is a slight response dip in the region of 3,000 c.p.s. and that the system begins to droop again around 9,000 to 10,000 c.p.s. All of this information is immediately available by visual inspection, without a pickup ever having been put to the disc.

Visible trouble indications

It is also possible to spot a number of troubles this way, includ-

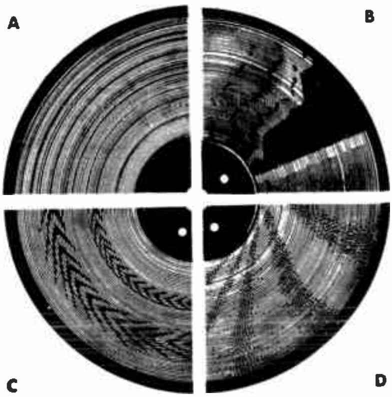


Fig. 919. *Visual inspection can reveal serious faults.*

ing those shown in Fig. 919. At *a* we observe there is considerable irregularity in the spacing of the grooves. There may be some difficulty with the lead screw or with the coupling between this screw and the cutter. A number of commercial records now appear this way, however, and for a purpose. Many recording machines now are equipped with continuously variable pitch adjustment. This permits decreasing the pitch (increasing the spacing between grooves) in anticipation of loud passages of modulation and closing down to a very fine pitch during very soft passages. This makes for much more efficient use of the recorded area, avoiding overcutting during heavy modulation and squeezing more time onto the disc when the modulation is light.

When a blank disc is warped, there will be some "cutter bounce," since the recorder is unable to follow accurately the vertical path presented by the disc. This will cause the groove to vary in depth and in the extreme condition the cutter will leave

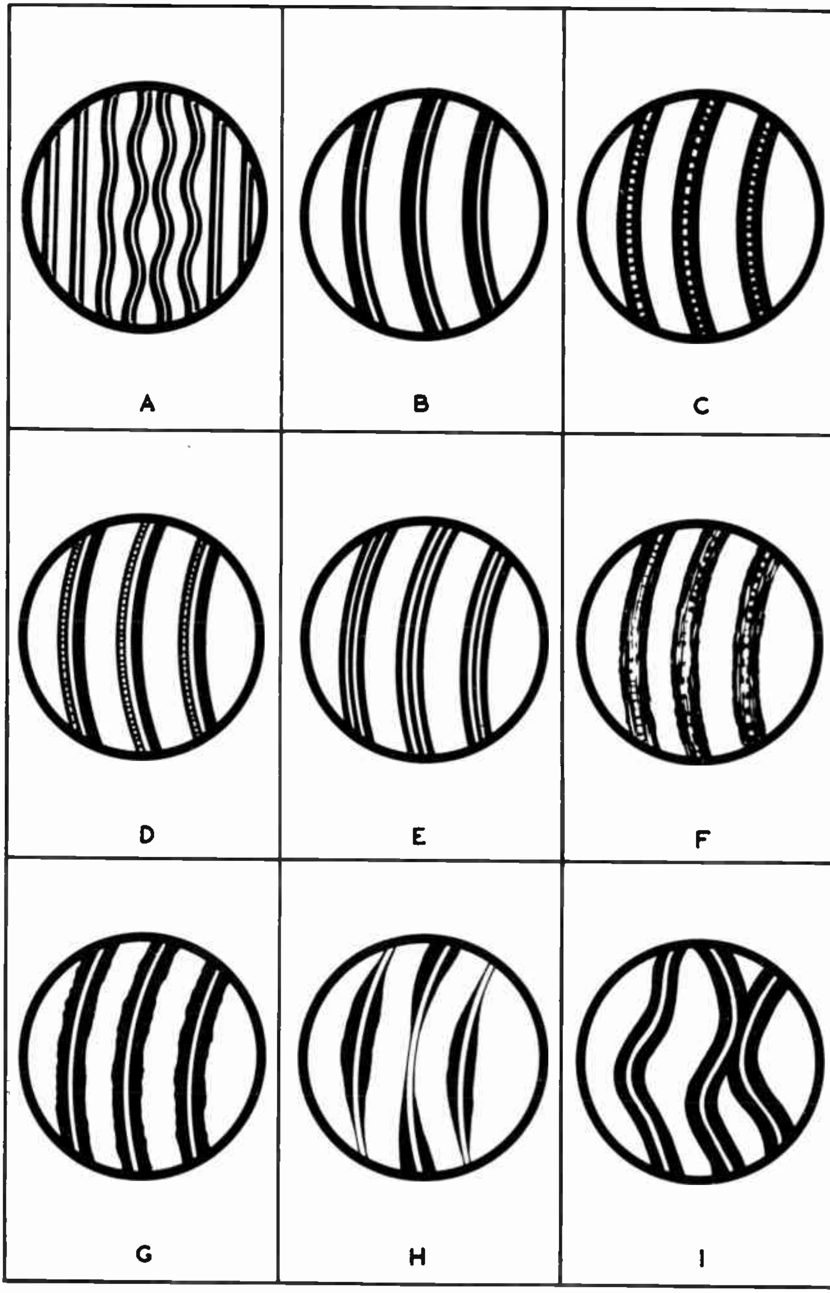


Fig. 920. *Typical faults revealed by microscopic examination of the grooves of a record.*

the disc altogether, resulting in a bare spot where no groove is cut. This is readily apparent as shown in the *b* quadrant of Fig. 919.

The moire pattern of Fig. 919-c indicates the presence of a low-frequency component such as hum or rumble. The spoke pattern of Fig. 919-d is indicative of a cyclical variation in speed such as might be caused by a flat spot on an idler. This would of course be heard in reproduction as wow or flutter.

Microscopic inspection

When record grooves are examined by reflected light under a microscope, much additional information becomes available. A well-cut groove will have smooth, shiny black walls, with a sharply defined and perfectly centered bottom, as shown in Fig. 920-a. The flat disc surface between the grooves, known as the *land,* appears very bright, as does the groove bottom.

The cutting stylus should normally be perfectly perpendicular to the surface of the disc. If it is tipped to one side or the other, so will the groove be tipped, perhaps so much that the reproducing stylus strikes bottom or jumps out of the top. This condition can easily be recognized under magnification, for the groove bottom will not be accurately centered between the walls (Fig. 920-b). Similarly, if the stylus is tipped forward or back, it may simply tear the disc material rather than engrave it. The first indication of this will occur at the groove bottom, the cutting stylus tip being unable to cut cleanly. The bottom will then take on a mottled appearance, as shown in Fig. 920-c.

A speck of dust or other foreign material may become trapped between the cutting stylus and the groove being cut with the result that the groove becomes scored as in Fig. 920-d. The audible effect of this is usually a hissing noise. The particle often escapes after only a few revolutions, which means that careful visual inspection of all grooves is necessary to spot it.

If the tip of the stylus becomes chipped, it will then probably have two or more points at its end, resulting in a groove with a multiple bottom. This is the cause of the double-bottom groove shown in Fig. 920-e. The only cure for this is to replace the stylus or have it resharpened. This is standard practice with sapphires, although the cheaper stellite points cannot practicably be repaired and must be discarded.

Other stylus troubles are indicated in Figs. 920-f-g. Here the styli are simply worn dull or else were improperly polished in the first place. At *f* both the tip and the burnishing facets appear to be badly worn, to the point where they are no longer able to cut a smoothly polished groove. The groove at *g* is quite ragged at its top edges, indicating that the trouble with this stylus is confined to the upper portion only.

In Fig. 920-h the groove depth varies widely. This will be due either to warpage of the disc, as previously discussed, or some mechanical difficulty which causes the cutter to have excessive vertical motion.

If the signal level is too high, the grooves will cross one another's paths, as shown in Fig. 920-i. Under these circumstances there will be echo effects in which the modulation of adjacent grooves carries over onto their neighbors'. At its worst, overcutting will be such that a reproducer simply cannot track and will jump grooves or even "skate" all the way across the record. The grooves will quite frequently appear to touch one another at their upper edges — this is especially true in microgroove practice — but the excessive displacement shown in this illustration could never be tolerated.

Magnetic recording

The basic notion of recording sound by magnetic means is an old one, having been originally explored by Valdemar Poulsen in Denmark prior to the turn of this century. The medium he employed was steel wire, and he received a patent on his *Telegraphone* machine in 1898. This machine was actually marketed for commercial purposes, a few of them even finding their way into the United States. This is quite remarkable in view of the fact that the available wire was very crude and the discovery of the vacuum-tube amplifier was still some years away.

The system fell into disfavor, however, and was all but forgotten for about three decades. But in the Nineteen Thirties there was some renewed interest in both the United States and in Germany. The American interests continued to use wire and steel tape, however, and it was not until Germany was occupied after World War II that the rest of the world learned of the important advances that had been made in that country in the art of magnetic recording.

The Germans had developed a plastic tape, coated with a fine layer of tiny iron-oxide particles, which far outstripped in fidelity anything previously known in any recording method. It is interesting to note also that another reason for the superiority of the German system was the application of *ultrasonic bias,* originally an American idea for which patent application was made in 1921 but which had never been applied to any commercial equipment here until the postwar years.

Most if not all professional-grade equipment manufactured today is heavily indebted to the German *Magnetophone* design. The most important development since then has been the introduction in the United States of tapes having higher coercive force characteristics, resulting in better frequency response, sensitivity and signal-to-noise ratio.

Magnetic principles

To understand the operation of a tape recorder, it would be useful to review briefly some of the basic principles of magnetics. The cause-and-effect relationship in magnetism is between the fundamental unit of magnetic force, known as the *gilbert,* and the fundamental unit of magnetic flux, known as the *maxwell.* The ability of a magnetic material to become magnetized is known as its *permeability,* while its opposition to becoming magnetized is called *reluctance.*

With these quantities it is possible to set up an equivalent to Ohm's law for magnetic circuits, in which flux is analogous to current, magnetomotive force to voltage, permeability to conductivity and reluctance to resistance. The relationships would then be expressed as follows:

$$R = \frac{\mathrm{H}}{\mathrm{B}} \qquad (47) \qquad\qquad \mathrm{M} = \frac{\mathrm{B}}{\mathrm{H}} \qquad (48)$$

where B equals the flux density in maxwells, H is the magnetomotive force in gilberts, R equals the reluctance and M is the permeability.

If we plot the relationship between B and H for a typical magnetic material, we will obtain the *normal magnetization curve* of Fig. 921. This is very reminiscent of the familiar vacuum-tube characteristic curves wherein the value of B increases slowly at first, then more rapidly and then again more gradually until the saturation point is reached. In magnetic circuits a material is said to be saturated when further increases in magnetic flux are only those due to the corresponding increase in the magnetizing force. the material itself making no further contribution to the total flux. It is important to note that the relationship is linear only in the middle of the curve, and the nonlinearity at the upper and lower bends must be regarded as a potential source of distortion.

As H is slowly decreased to zero, the path of the normal magnetization curve is not retraced for B. But when H = 0, there is still a definite positive value B_r for the magnetic flux, which is known as the *remanence*. This is shown in Fig. 922. To reduce the magnetism in the material to zero after it has been magnetized to saturation, it is necessary to apply further magnetomotive force

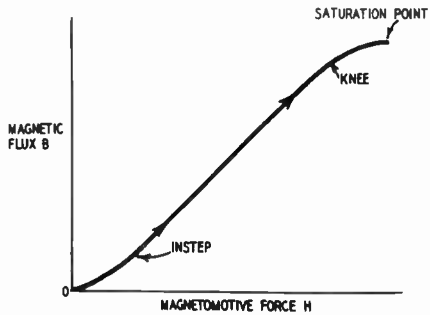


Fig. 921. *Normal magnetization curve.*

in a direction opposite to the original force. This is indicated by point H_c in Fig. 922 and is known as the *coercive force*. If H continues in the negative direction, another saturation point will be reached, and if H again decreases to zero, B will once again exhibit its remanence characteristic.

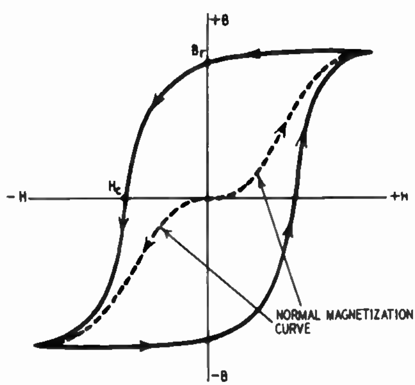


Fig. 922. *Example of hysteresis loop.*

This characteristic, which might be regarded as a lag between the magnetic flux and the magnetomotive force producing it, is known as *hysteresis*. For this reason the curve of Fig. 922 is often known as a hysteresis loop.

There is some relationship between these characteristics and the frequency characteristics of a magnetic medium. In a very general way the high-frequency response of the medium is a function of the coercive force, while the mid-range and bass sensitivity is proportional to the remanence. This is the reason for the statement that the high-coercive-force oxides developed in the United States represent an improvement over the original German style media.

Tape recording systems

To understand how these facts bear on a magnetic recording, consider Fig. 923, which shows some of the important components of a quality tape recording system. The tape is moved from supply reel to takeup reel and past the recording heads by a motor-driven system known as *capstan drive*. The tape is held against the capstan by the pressure roller. If there is no slippage and the capstan

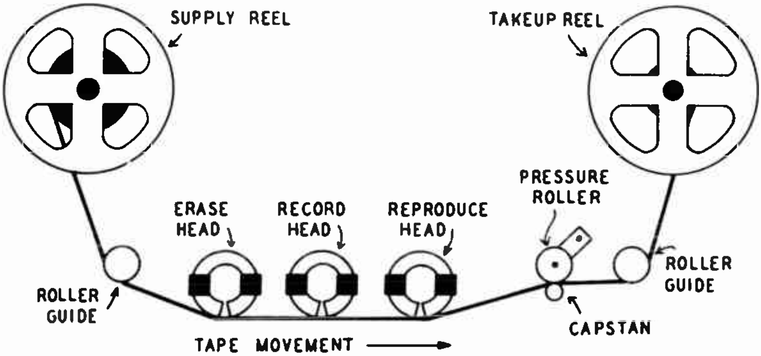


Fig. 923. *Major components of a tape recording system.*

speed is constant, the movement of the tape past the heads will be perfectly smooth.

The tape first moves past the *erase* electromagnet, which is always functioning when a tape is being recorded, even when virgin stock is used. This may seem superfluous since new tape is presumably magnetically neutral, but in fact there will usually be some residual tape noise which must be removed by the erasing process. One obvious cause of this would occur where some elements in the tape manufacturing process, such as the coating, slitting or winding machinery, become magnetized and in turn magnetize the tape. It might also be that the tape had come in contact with induction fields during transportation or storage.

Another reason for new tape noise might not be so obvious. When we speak of a magnetic material as being magnetically neutral, we mean that the near-infinite number of tiny magnets within its structure are aligned at random so that the sum of all their fields in a given length of tape is equal to zero. But when a playback head of a tape recorder is scanning a piece of tape, we are considering only an extremely short length of tape at a given time, the length being equal to the size of the head gap, which is very much less than 1 mil. Thus the head is in effect scanning a series of minutely short sections of tape successively. While the tape may on the *average* be magnetically neutral when we are considering a large enough sample, it is highly unlikely that the very short sections which pass the head successively are each of themselves entirely free of any outward display of magnetism. But

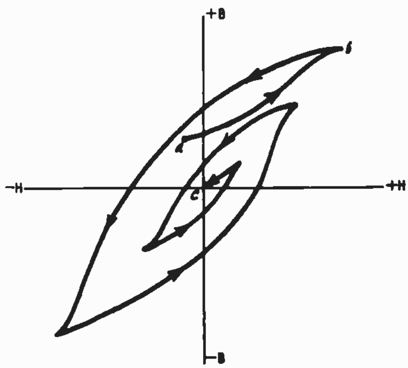


Fig. 924. *Technique for producing an a.c. erase.*

when the tape has first passed under the influence of the erase head, it is completely demagnetized and free of innate noise.

Erasing

Since the term *erase* implies rubbing out or defacing, it is perhaps a misnomer, for no mechanical or visible change takes place. In the German scientific literature the term *obliteration* is invariably used, although this is hardly more accurate. Some American authorities prefer the term *wiping,* which also leaves something to be desired. Until a better term is proposed, we shall employ the one most commonly used in the United States — *erase.* But it should be clearly understood that in magnetic recording this is in no sense a mechanical operation, but simply the means of reducing the tape to a state of magnetic neutrality. The erasing process therefore completely removes previously recorded signals and residual noise without otherwise affecting the tape in any way. Thus the magnetic recording medium is repeatedly reusable, a property which can be claimed for no other known type of recording.

There are several methods of magnetic erasing in current use, but the most satisfactory one, and the only one ever employed in high-quality professional equipment, is known as *a.c. erase*. This is based upon the fact that a magnetized material may be reduced to a state of zero magnetism by subjecting it to a progressively diminishing alternating field. The process is illustrated in Fig. 924. If the magnetic state of the tape before erasing is at point a, then

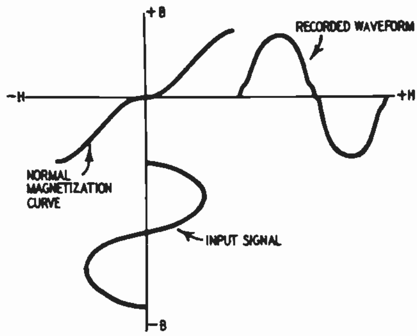


Fig. 925. *Distortion produced in recorded waveform.*

a positive field is applied up to point b, followed by a negative field and so on through a series of decreasing hysteresis loops until point c is reached, the point of zero magnetization. The value of H necessary to produce the magnetization of point b need only equal the peak value of the signal on the tape. But for an adequate safety factor it is standard practice to have the initial cycles of the field large enough to produce magnetic saturation. This will assure that the erase will remove even the most heavily overmodulated signals.

Complete demagnetization will occur only if a sufficient number of field reversals have been gone through and if the decrease in field strength has been gradual. Excellent erasing will result if a hundred or more cycles are gone through with the same rate of decrement, as for example proportional to +100, —99, +98, —97, ... —3, +2, —1, 0.

The recording process

The perfectly "clean" tape now passes to the recording electromagnet where it is magnetized in accordance with the instantaneous audio signal. But since the tape arrives in a demagnetized state, any effort to magnetize it will encounter the characteristics of the normal magnetization curve of Figs. 921 and 922. Since this curve is far from linear, the magnetized signal will also be non-linear and hence distorted, as shown in Fig. 925.

However, it would be possible to record without distortion if the signal were centered, not on the zero axis, but at the center of the linear portion of the curve. Then the extremities of the working range would be just above the instep and below the knee of Fig. 921. This could be accomplished simply by applying a d.c.

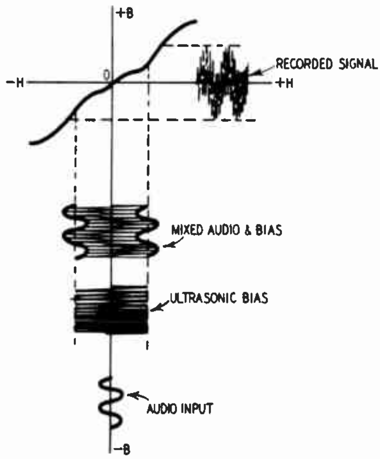


Fig. 926. *Effect of mixing ultrasonic bias and the audio signal.*

bias, either negative or positive, of such value as to arrive at the middle of either of the linear portions of the normal magnetization curve of Fig. 922. This method was actually employed in early magnetic recording systems, but it is apparent from the illustrations that the length of the linear portion of the curve is quite small and the dynamic range of the signal would therefore be severely restricted.

The ideal condition, then, would appear to be one in which *both* linear portions of the curve could be used, and this is exactly the result we achieve through the use of *ultrasonic bias*. The effect of this process becomes evident when we consider Fig. 926. The bias is mixed with the audio signal (*not* modulated by it) and in this way the audio portion of the composite signal strikes both linear portions — and only these portions — of the magnetization curve. Since the bias is ideally neither recorded nor reproduced, only the audio magnetizes the tape, but it does so without distortion. The bias is therefore somewhat analogous to a chemical catalyst in that it helps in the performance of a certain reaction without itself taking part in it.

This admittedly loose explanation is the essence of the most generally accepted theory concerning the highly complex behavior of ultrasonic bias in magnetic recording. But the entire subject is still the center of some rather lively debate, and the art is still so young that the theory just presented is far from immutable law. It appears to satisfy most of the conditions as we know them, but you should retain an open mind on the subject until more information is available.

The bias frequency should be several times the highest audio frequency to be recorded to avoid the recording of any beat frequencies. Bias frequencies in commercial practice will usually

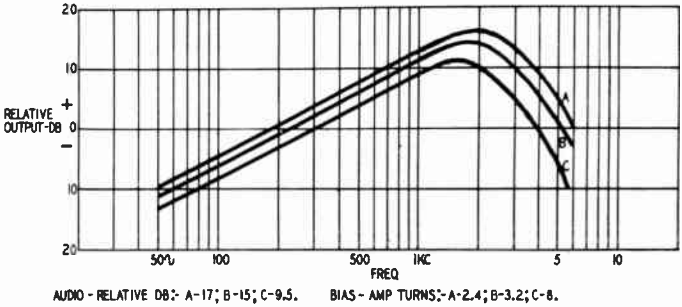


Fig. 927. *Relationship between bias and frequency response.*

range between 30 and 100 kc. Since such frequencies are quite convenient for the erase voltage, it is usual practice to permit a single oscillator provide the original signal for both applications.

The relationship between bias and frequency response is illustrated in Fig. 927. These particular curves were obtained with a constant current flowing through the windings of the recording head at all audio frequencies, and with the bias frequency fixed at 40 kc. It may be inferred from this that the higher the bias current used, the poorer the high-frequency response relative to the bass response. Pre-equalization is standard practice in magnetic recording, however, and the value of the bias voltage is not normally selected for the frequency characteristics it provides.

Optimum adjustment of bias current is predicated upon minimum distortion, and the relationship between these two factors and the output amplitude are shown in Fig. 928. When the input signal is held constant while the bias voltage is varied, at low bias values the output is also low while distortion is high. As the bias is increased the level also increases, but the distortion goes down and finally reaches a minimum. Further increasing the bias causes the distortion to rise again while the output continues its upward swing. Finally the distortion reaches a secondary peak simultaneously with the peak in output. Then, as additional bias is applied, both output and distortion decrease slowly.

The bias may be adjusted to the point of the valley in the distortion curve, but in practice this is seldom done. Such an adjust-

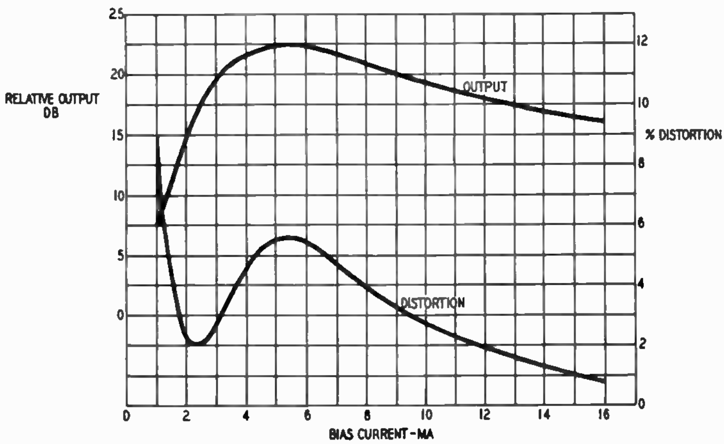


Fig. 928. *Effect of bias current on output and distortion.*

ment is exceedingly critical, and the signal-to-noise ratio is not as good as can be obtained elsewhere. Adjustment of low-speed (7.5 inches per second or less) home machines is usually made simply for peak output, despite the fact that somewhat less distortion is possible elsewhere. Since these machines exhibit a rather high order of distortion inherently, there is little reason for adjusting bias for the optimum tape condition. For professional equipment, an adjustment much closer to the ideal may be effected simply by choosing a bias current twice that required for maximum output. More refined techniques involve the use of a distortion meter or harmonic wave analyzer.

Methods of magnetization

Three distinct methods for magnetizing the tape or other material have been devised. Classified according to the direction of magnetization (that is, depending upon the orientation of the recording head gap with respect to the tape) they are known as *longitudinal, lateral* and *perpendicular* or *transverse*.

In longitudinal magnetization (Fig. 929-a) the head gap extends across the width of the tape and the direction of magnetization is along its length. This was the original method employed by Poulsen, and is still used almost exclusively in commercial machines. Its principles may be understood by recalling that any magnetic material is really a large number of tiny magnets of molecular size. In the unmagnetized state these have a random distribution so that their fields cancel each other and no external indication of magnetism exists.

When a demagnetized recording medium, such as a length of tape, is moved past a recording electromagnet, the small magnetic particles of which the tape consists are all rotated so as to align themselves with the instantaneous direction of the magnetic field which excites them. Then the magnitude and polarity of their own magnetic field will depend upon the magnitude and polarity of the current through the coil of the recording head. Thus after the tape has passed over the recording-head gap, it will consist of a number of very short sections, each of them in effect a narrow line of magnets all pointing in the same direction. Since these successive lines of magnetization, when taken as a whole, will have various opposing directions in accordance with the signal, the tape will exhibit little or no external magnetic effect, except when these lines are scanned one at a time by a playback head.

Lateral magnetization, as shown in Fig. 929-b, has the head moved 90°, so that the gap is parallel with the length of the tape. High-frequency response is very poor with this method and it is normally used as a secondary type in conjunction with longitudinal magnetization. It is useful in this case for applying cueing or synchronizing signals, which are not detected by the longitudinal playback head, but which are picked up by a separate lateral playback head and transmitted to a control device of some sort. Quite frequently employed in motion-picture photography, especially for television, this method helps maintain perfect synchronism between recorder or playback and camera or projector.

In transverse magnetization the poles of the recording head are on opposite sides of the tape, so that the magnetization is in the direction of the tape thickness and perpendicular to its surface. It has been claimed that a wider frequency range is possible at lower speeds with this method, but its critical operating adjustment has prevented its wide use. Furthermore, to be practical with tape the oxide would have to impregnate the medium, but all commercial recording stock manufactured in the United States has the oxide coated on one side of the base. And finally, even if impregnated tape were used, recording it in this manner would remove all magnetic isolation between adjacent layers on a reel, and there would probably be a great amount of layer-to-layer transfer of magnetic fields. This phenomenon, which sometimes occurs even with longitudinally magnetized coated tape, is often referred to as *print-through* and results in "ghost" and "echo" effects.

The recording medium

The ideal recording tape would be practically the equivalent of a homogeneous magnetic fluid, uniformly coated on or impreg-

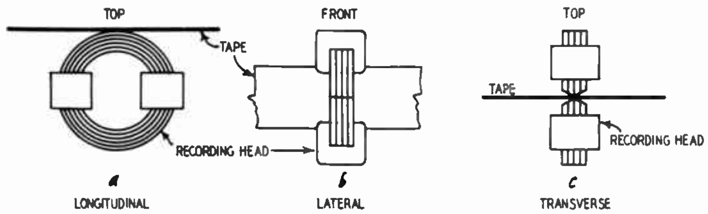


Fig. 929-a, -b, -c. *Longitudinal, lateral, and transverse magnetization.*

nated in a base carrier of high tensile strength, perfectly smooth and with absolutely no shrinkage or expansion. The magnetic material itself would be extremely sensitive and would exhibit no frequency discrimination or other distortion effects anywhere within the audible range.

Several types of coated and impregnated tape were developed in Germany after World War II. All of them used finely powdered red iron oxide for the magnetic material, a chemical compound of iron and oxygen having the formula Fe_2O_3. This is still the basis of most tape manufactured in the United States, although some work has been done with black iron oxide (Fe_3O_4). More recently certain additives have been introduced with a considerable degree of success to provide greater sensitivity. But the precise nature of these additional materials is regarded as a trade secret by the manufacturers who have developed them.

All of the widely used tapes in the United States are of the coated type in which the powdered oxide and a suitable binder are applied to one side of a smooth ribbon of paper or plastic. The character of this base is of extreme importance, for unless it is perfectly smooth the coating will likewise be rough and cause noise in the output. This is why the smoother plastic base provides a better signal-to-noise ratio than paper tape with the same coating. Paper does have the advantage of somewhat better dimensional stability, but the plastic type is generally of such better quality that paper recording tape is not very popular.

The tape base is $\frac{1}{4}$-inch wide and on the average is 1.5 mils thick, coated with about 0.5 mil of magnetic material, most of which is binder. There is known to be a strong correlation between frequency response and tape coating thickness, a thinner coating providing a better high-frequency response. Consequently a great deal of effort has been expended on developing thinner and more sensitive coatings, more effective binders with less bulk and thinner bases with adequate strength and stability.

Not very much can be done in the way of making the base thinner which doesn't involve mechanical troubles as well as increased layer-to-layer transfer, but a high-remanance coating has been recently devised which has made it possible to reduce its thickness to only 0.3 mil while still retaining full output and sensitivity. Improvements in the base have made it possible for its thickness to be reduced to 1.0 mil. By reducing both thickness dimensions while still retaining the same performance, an increased playing time of 50% is gained for that much more tape can be accommodated on any given reel.

Layer-to-layer signal transfer

When a length of recorded tape is wound on a reel, each layer of tape will fall under the magnetic influence of the adjacent layers. The amount of magnetization induced between layers is ordinarily undetectable in reproduction, but in some instances may be nearly as loud as the signal itself. This effect is nonlinear, like recording without bias. For each decrease of 1 db in recording level, the interlayer transfer level will drop about 2 db. Thus print-through is a much more serious problem in the case of very heavy modulation and overloading and much trouble can be avoided simply by keeping the level within bounds.

The effect becomes more noticeable after the tape has been stored for a period of time. It is also accelerated by high temperatures, increasing by about 1 db for every temperature increase of 10°F. Thus, it is wise to store recorded reels in a cool place.

Nearby magnetic fields will not only induce noises, but they will sometimes increase the transfer level by as much as 30 to 40 db. It is therefore extremely important to keep the recorded reels well away from any sources of stray fields, such as magnets, motors, generators and high-tension power lines. When the precautions against overmodulation, excessive temperature and stray fields are all carefully observed, the print-through problem is seldom of any real consequence.

Tape frequency characteristics

Since a voltage will be induced in a conductor whenever it cuts magnetic lines of force, the playback process simply involves pulling the recorded tape past the surface of a specially designed electromagnet known as the reproduce head. The induced voltage follows the usual laws of magnetism and is directly proportional to the rate of flux cutting, the higher frequencies with their more rapid magnetic reversals inducing higher voltages in the reproduce head. Since doubling the frequency simultaneously doubles the voltage, it is characteristic of tape that its frequency response has a rise of 6 db per octave. This is true only up to a point, however, and at the higher frequencies beyond this, other factors which enter into the playback process tend to counteract this characteristic.

Whenever a magnetic material leaves the influence of its exciting fields, its actual residual magnetism will fall slightly below the maximum strength of the excitation. This is due in part to some interaction between the opposite poles of the infinitely small magnets in the material and is known as *self-demagnetization.* Its effects in recording are more pronounced at the higher frequencies, thus reducing the treble response. And there are the usual eddy currents and other iron losses typical in magnetic phenomena, all of which tend to pull down the high-frequency response even farther.

Frequency response in reproduction is also largely dependent upon the dimensions of the playback-head gap. At the higher frequencies the gap will approach or even exceed the dimension of one wavelength, and beyond this point the frequency response is seriously impaired. The recording-head gap is not as critical. Machines which have separate record and playback heads may have a record-head gap of around 1 mil and the reproduce-head gap in the order of 0.25 to 0.5 mil. Ideally then, the gap should be as small as possible, but there are practical limits. A very small gap dimension approaches the state of a magnetic short, with the signal applied to the tape being very small or even nonexistent. There are also the problems of manufacturing which make such a head very expensive, if not impossible, to build.

The contact between tape and reproduce head must be very close for optimum response. Likewise the alignment of the head gap with respect to the tape is very important. Since the method of magnetization is longitudinal, the head gap should be at precisely 90° relative to the direction of tape travel. This angle would not be so critical, were it not for purposes of standardization. If this procedure is not followed, a tape recorded on one machine and then played back on another having some angular displacement of the gap would reproduce rather badly, particularly at the high end. These relationships are illustrated in Fig. 930. We would obtain the same difficulty, of course, in a machine having separate

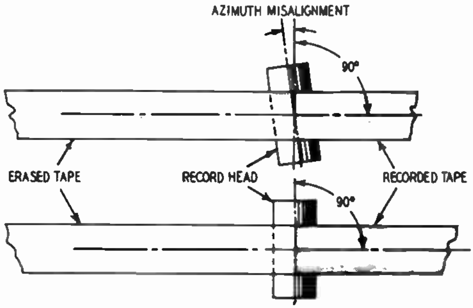


Fig. 930. *Angular displacement of the record head will produce distortion.*

record and reproduce heads not perfectly aligned with each other.

The azimuth (angular) adjustment may be checked by reproducing a special azimuth tape available from manufacturers of most professional tape machines. The reproduce head is simply adjusted for maximum output at all frequencies. The record head may then be adjusted to conform to the playback-head azimuth by recording a series of high-frequency tones and noting the effect on the reproduced output as the record-head adjustment is varied.

Recording speed

The speed of tape travel has a considerable effect in determining the high-frequency response. A higher speed will increase ratio of the shortest audio wavelength to the dimension of the reproduce-head gap and it will also reduce the self-demagnetization. The frequency of maximum response (see Fig. 927) varies directly with the speed so that doubling the speed will raise this peak exactly an octave. From the standpoint of frequency response, it is obviously desirable to use as high a speed as possible. A higher speed, however, requires greater tape consumption for the same amount of program time, and so the speed will depend largely upon the requirements of the service to which the equipment is to be put.

The original German equipment operated at a speed very close to 30 inches per second, and this was the speed adopted for the first professional machines introduced into the United States. This speed is seldom used any longer, except for the most exacting professional applications; 15 i.p.s. is now at least as good as 30 i.p.s. was when first introduced. All other tape speeds presently in

TAPE LENGTH, FEET	RECORDING TIME, MINUTES FOR TAPE SPEEDS IN INCHES PER SEC.				
	1 7/8	3 3/4	7 1/2	15	30
150	15	7 1/2	3 3/4	—	—
300	30	15	7 1/2	3 3/4	—
600	60	30	15	7 1/2	3 3/4
1200	120	60	30	15	7 1/2
2400	—	—	60	30	15
4800	—	—	—	60	30

Fig. 931. *Relationship between tape speed, length, and playing time.*

use are submultiples of 30 i.p.s., and the relationship between these various speeds, tape lengths and playing times is shown in the table of Fig. 931. The speed now used for most professional applications is 15 i.p.s. while 7½ i.p.s. is employed in the majority of home machines. The slower speeds, however, are quite adequate for certain speech uses where maximum fidelity and intelligibility are not required.

Tape equalization

It is obvious that the 6-db-per-octave characteristic of unequalized recording tape, such as is shown in the typical curves of Fig. 927 is not satisfactory for reproducing purposes. To achieve anything approaching a flat characteristic in reproduction, a considerable amount of playback equalization would be necessary. At 6 db per octave, the difference between the recorded levels at 1,000 and 50 c.p.s., for example, is 24 db. But a bass boost which would flatten this characteristic would also bring up with it a large amount of hum and rumble as well.

A much more practical method is a system of pre-equalization which is ideally based upon the overload vs. frequency characteristics of the tape. The intent of this is simply to put as much signal as possible on the tape at all frequencies .This will involve pre-emphasis at both the high and low frequencies and, since the peak response varies by an octave for each doubling of the tape speed, a different equalization characteristic must be employed for every speed. Unfortunately, there has been nearly as much disparity in equalization characteristics between tape machines of different manufacturers as there once was in disc recording characteristics. For this reason, a recording made on a given machine and played on another of a different type may often give disappointing results.

Magnetic recording is widely used in radio and television broadcasting, phonograph records and pre-recorded tape and literally thousands of applications in commerce, education and the home. It has been employed successfully for the recording of television picture signals in full color and new uses are being discovered for it every day with the end nowhere in sight. It is without question the most outstanding and important development in the audio art for several decades.

index

Other GERNSBACK LIBRARY Books

47—RADIO & TV HINTS. Hundreds of short-cuts. The hints and kinks in this book are practical, and can be put to work by you. 112 pages. $1.00

48—HIGH FIDELITY. Audio design, construction, measurement, techniques. Complete section on building amplifiers. 128 pages. $1.50

49—RADIO AND TV TEST INSTRUMENTS. Build CRT circuit analyzer, CRT tester, 3 inch scope, portable marker generator, etc. 128 pages. $1.50

50—TV REPAIR TECHNIQUES. Service TV in the home; high-voltage troubles, CRT circuit difficulties, TVI, intercarrier buzz, etc. 128 pages. $1.50

51—TRANSISTORS - THEORY & PRACTICE. Semi-conductor theory, characteristics, equivalent circuits. Practical transistor circuits, tests and measurements. 144 pages. $2.00

52—THE OSCILLOSCOPE. Waveforms, sweep systems, typical scopes, alignment techniques, tests, measurements, experiments. 192 pages. $2.25

53—RADIO-CONTROL HANDBOOK. Radio control for planes, boats, trucks, tractors, etc. Practical construction ideas. 192 pages. $2.25

54—PROBES. Crystal-demodulator, voltage-doubler, low capacitance, high-voltage, isolation, transistorized and direct probes. 224 pages. $2.50

55—SWEEP & MARKER GENERATORS for TV and Radio. Complete theory and use of sweep and marker generators. 224 pages. $2.50

56—HIGH FIDELITY CIRCUIT DESIGN. Feedback, drivers, inverters, attenuators, equalizers, speaker systems. Hard cover. 304 pages. $5.95

57—The V.T.V.M. Supplies full information on the theory and practical operation of the vacuum-tube voltmeter. 224 pages. $2.50

58—MAINTAINING HI-FI EQUIPMENT. Instruments; circuits; diagnosis; distortion; pickups; turntables; equalizers; tuners. 224 pages. $2.90

59—SERVICING RECORD CHANGERS. Detailed drawings show how to make repairs. Full servicing information. 224 pages. $2.90

60—RAPID TV REPAIR. Alphabetical listing of hundreds of TV troubles — symptoms and specific information on repair. 224 pages. $2.90

61—TRANSISTOR TECHNIQUES. How to work with and test transistors; measurements; practical building projects. 96 pages. $1.50

62—TV—IT'S A CINCH! Snappy conversational style with hundreds and hundreds of drawings helps you breeze through TV. 224 pages. $2.90

63—TRANSISTOR CIRCUITS. Over 150 practical circuits, using transistors only. A guide for the experimenter. 160 pages. $2.75

64—UNDERSTANDING HI-FI CIRCUITS. Explains how modern audio circuits work and what you can expect from them. 224 pages. $2.90

65—SERVICING COLOR TV. How to repair color receivers. Illustrated with photos, drawings and troubleshooting charts. 224 pages. $2.90

67—ELEMENTS OF TAPE RECORDER CIRCUITS. Complete coverage of the electronic portion of tape recorders. 224 pages. $2.90

68—TV AND RADIO TUBE TROUBLES. Service TV and radio receivers by understanding tube troubles. Recognize symptoms quickly. 224 pages. $2.90

69—ELECTRONIC HOBBYISTS' HANDBOOK. Scores of practical circuits for the hobbyist. Parts lists on construction projects. 160 pages. $2.50

70—ELECTRONIC PUZZLES AND GAMES. Build and design them. Dozens of projects which need no special parts or tools. 128 pages. $1.95

71—AUDIO DESIGN HANDBOOK. Each chapter covers a separate section of an audio system, from preamp to speaker. 224 pages. $2.90

72—OSCILLOSCOPE TECHNIQUES. Derivation and interpretation of waveforms. Hundreds of photographs of actual traces. 224 pages. $2.90

73—AUDIO MEASUREMENTS. Test equipment and all phases of audio measurement from basic amplifier to microphone. 224 pages. $2.90

74—MODEL RADIO-CONTROL. Coders, transmitters, receivers, power control, servos, transistors. Theory and construction. 192 pages. $2.65

75—TRANSISTORS—THEORY and PRACTICE. Theo.y, equivalent circuits, amplifiers, oscillators, tests and measurements. 160 pages. $2.95

76—SERVICING TRANSISTOR RADIOS. Fundamentals, types of construction, testing, stage-by-stage servicing procedures. 224 pages. $2.90

77—GUIDE TO MOBILE RADIO. Industrial and railroad radio, selective and remote systems, maintenance and licensing. 160 pages. $2.85

78—RAPID RADIO REPAIR. Fix radios quickly and easily. Separate sections deal with receiver types, servicing techniques, troubles. 224 pages. $2.90

79—DESIGNING & BUILDING HI-FI FURNITURE. Fundamentals of design, woods, tools, professional finishing, polishing and retouching, furniture styles, placement. 224 pages. $2.90

80—STEREO . . . HOW IT WORKS. Complete coverage of stereo, discs, tapes, multi-channel, installation, pseudo-stereo. 224 pages $2.90

81—PRINTED CIRCUITS. Designing and making printed circuits, repairs, subminiaturization, applications. 224 pages. $2.90

85—HOW TO GET THE MOST OUT OF YOUR VOM. Basic meter, kits, accessories, measurements, servicing. 224 pages. $2.90

86—INSTALLING HI-FI SYSTEMS. Includes every aspect of home installations: interiors, decor, furniture and acoustics. 224 pages. $3.20

90—HI-FI MADE EASY. Audio explained for the layman in easy-to-understand form. Profusely illustrated. 224 pages. $2.90

102—PRACTICAL TV TROUBLESHOOTING. Useful repair techniques based on actual servicing experience. An easy guide for tough repair jobs. 128 pages. $2.35