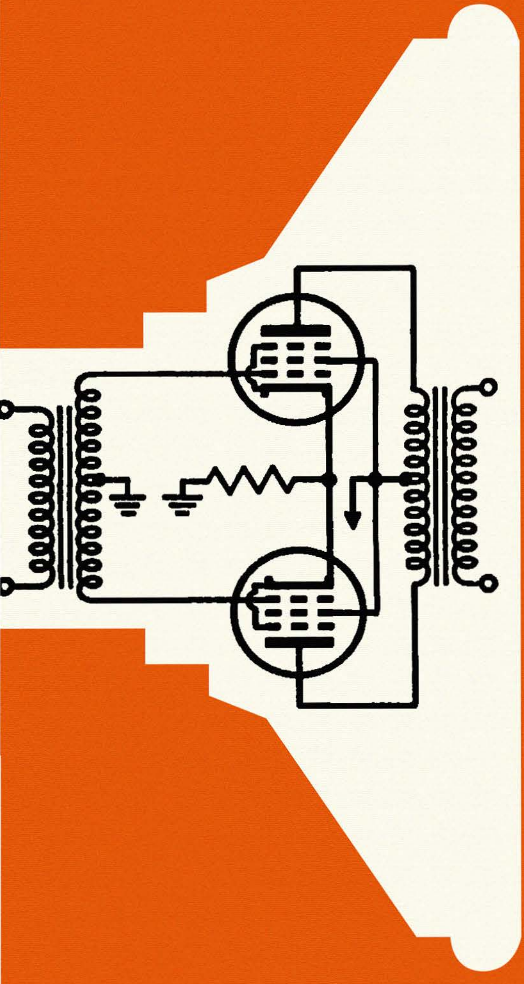
No. 201

basic audio

by NORMAN H. CROWHURST



VOLUMES

1 to 3

a RIDER publication

basic audio

by NORMAN H. CROWHURST

VOL. 1

JOHN F. RIDER PUBLISHER, INC.
116 West 14th Street • New York 11, N. Y.

Library of Congress Catalog Card Number 58-13425

Printed in the United States of America

PREFACE

Audio is like Topsy: it wasn't born, it just growed. Whatever Topsy may have been like, Audio has grown like a gawky child—not always in proportion! Originally associated with radio and later with high fidelity, audio now finds application in many other places—to name a few: computors, automation, ballistics and guidance for missiles, sonar detection for navigation, ultra- and infra-sonics for medicine, both diagnostic and therapeutic, as well as geophysical and other work. In fact Audio is now one of the largest and most basic divisions of electronics.

Courses in audio were nonexistent not too many years ago. Since then, textbooks and courses have appeared. But their approach follows the principle of many professors: "I learned it the hard way—you'll have to!" It's like learning watchmaking from a bridge-building man.

My wide experience in various aspects of audio has shown the need for a better way. In industry, in academic education, and particularly in working with graduates from college, this need is evident. My extensive technical writing for magazines and consultant work in the industry have also shown me audio's educational needs.

Many competent "practical men" find themselves hindered by lack of academic background in the subject. They can do their job in their own established "groove." But they do not have—and find it impossible to acquire—the background to enable them to expand outside this groove. These people need help in closing the gap between "theory" and "practice."

Engineers are conversant with the accepted "technical language," but they read the literature with only an "intuitive comprehension" (or should it be apprehension?). Their education dragged them past many "awkward spots" about which they have never felt really "comfortable." Like the King of Siam in "The King and I," they find many facts of which they wish they were more certain they are sure.

Very important are the new students, technicians, and audiophiles. They will need a basic education in audio to enable them to add their contribution to progress (and to earn themselves a living!). Why make it difficult? They'll do much better if they can get a good start.

All-in-all, it is time that certain roundabout approaches to this key subject were eliminated. We need a direct, meaningful way to take the place of the difficult detours. Then each of our three groups can not only "learn audio," but also understand it! This three-volume book results from the author's extensive education research. The finished arrangement achieves a completely new directness.

Let me give just one example: how many understand the behavior of a coupling capacitor, particularly its contribution to amplifier transient performance, and what sometimes happens to feedback? This has always been based on the concept of capacitive reactance, which does not adequately *explain* all the effects. We have adopted a practical "what happens?" approach.

As a result, someone who learned this the old way may miss the familiar landmark of the reactance concept—when he expects it; a closer examination will reveal the reason for postponing it: the whole presentation has been arranged to avoid the "dead spots" left by the traditional approach.

Inevitably such a change of approach will mean a change of stress. I make no apology for this. I know from practice that it is far more successful in getting Basic Audio "across."

It would be impossible to acknowledge the very many who have, knowingly or unknowingly, contributed to my experience, making this book possible. But I would like to express my thanks to the John F. Rider staff for their cooperation in "packaging" it in a form that interprets my intentions so well.

NORMAN H. CROWHURST

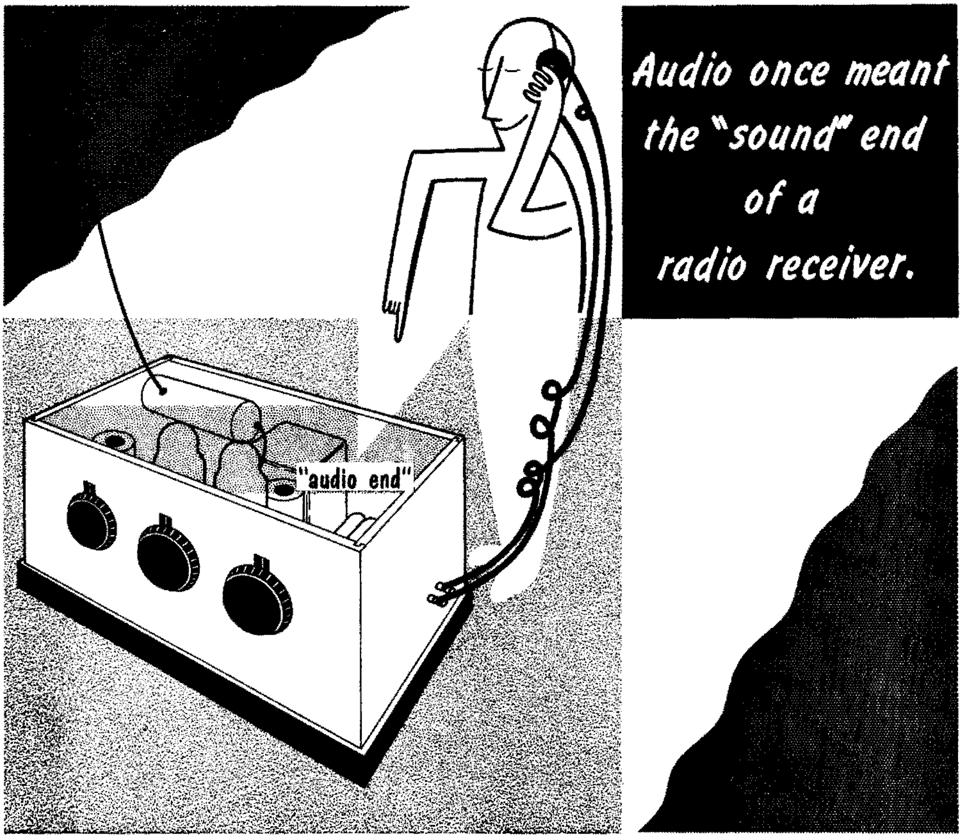
New York, N.Y.
August 1959

CONTENTS

VOL. I—BASIC AUDIO

What Audio Is

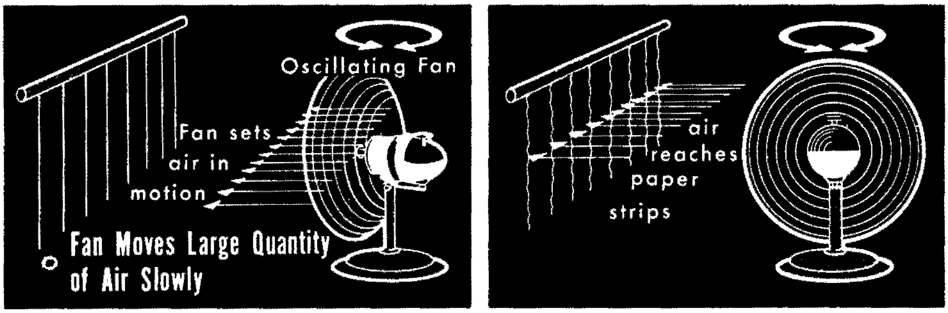
In the early days of radio, the word "audio" was used to describe the part of a radio receiver (or transmitter) that amplifies the audio, or sound frequencies. After detecting (later called *demodulating)* the radio waves or carrier, the signal was rather feeble, hardly sufficient to be audible in the headphones, so an *audio* amplifier was used to make the feeble signal audible in comfort. A similar amplifier was needed at the transmitter to magnify the feeble microphone currents for modulation of the radio carrier.

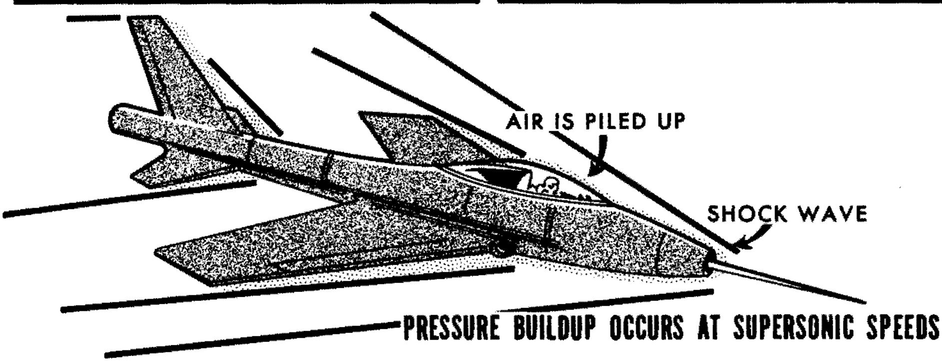


From those early beginnings, audio amplifiers have found more and more uses, and have improved in performance. Quite early, audio amplification was applied to the making and reproducing of phonograph records, giving us the first "electrical" reproduction. Since then, the methods developed have been used for many things: all kinds of control mechanisms for industry; sonar-type detection devices; vibration measuring equipment for aircraft and other types of development or research; and many of the "brain cells" used in electronic computers, to name just a few.

What Sound Is

Air can move in different ways. One way can be illustrated by an oscillating electric fan. It blows the air at a comparatively slow speed, but moves quite a quantity of it. The slowness with which the air is moving is apparent because the draft of air reaches you a little while after the fan has stopped blowing in your direction. Here a large quantity of air is moved at a speed of only a few feet per second (much slower than sound travels). The air movement is large, and there is hardly any compression.

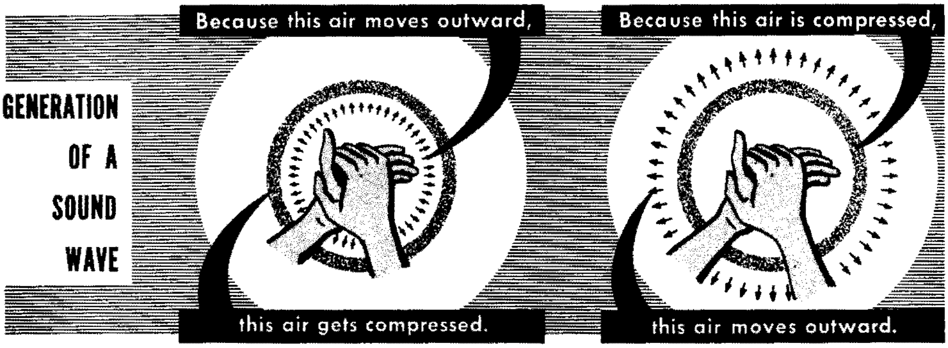




When a jet aircraft breaks through the sound barrier, however, it encounters different (and, at first, strange) kinds of air movement. The aircraft is now traveling faster than sound, and the air in front of it is no longer moving freely, but is "piled up" at high pressure on the front surfaces of the plane. As the aircraft goes through space, more air piles up in front of it, while some escapes at the side, producing the well-known "shock wave." The air in front of the plane hardly moves (relative to the aircraft), but is in a state of high compression—just the opposite condition from the air moved by the fan. However, the aircraft must move faster than sound (about 750 miles an hour) to cause this effect.

Both these forms of air movement do not "carry" very far: air movement from a fan soon gets "lost," and the pressure buildup in front of the aircraft in supersonic flight is not very deep—at most a few feet.

Characteristics of Sound Waves

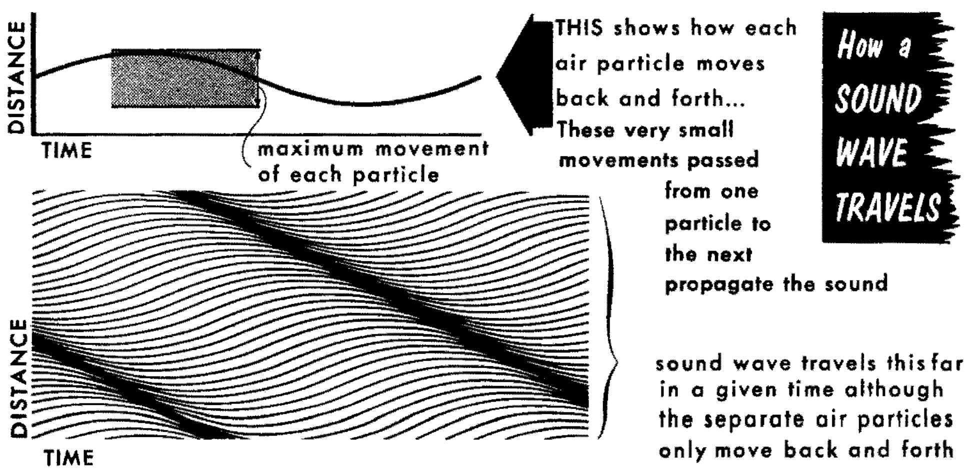


Sound waves combine these two forms of movement so as to cover great distances with only small air movements. Take the sound of a hand clap—when your hands come together, a small amount of air is forced out quite suddenly at the last instant.

The air pushed out by the hands clapping has a momentary movement *and* some pressure, although both are quite small compared to the movement produced by the fan or the pressure buildup of the supersonic jet. Two things now happen at the same time:

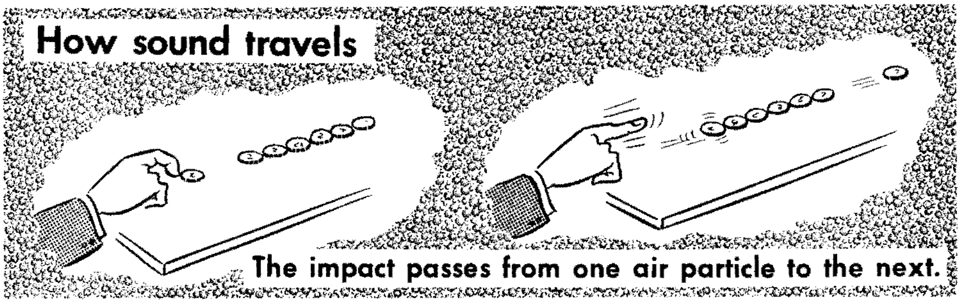
1. Because the pressure of the air close to the hands is momentarily greater than that further away, air moves outward.
2. Because the air close to the hands is moving outward, the air immediately beyond it also gets compressed.

The two actions combine to keep the pressure and movement wave traveling at a natural speed—the propagation velocity of sound in air. This is *not* to be confused with the speed at which individual air particles move due to the wave.



Characteristics of Sound Waves (contd.)

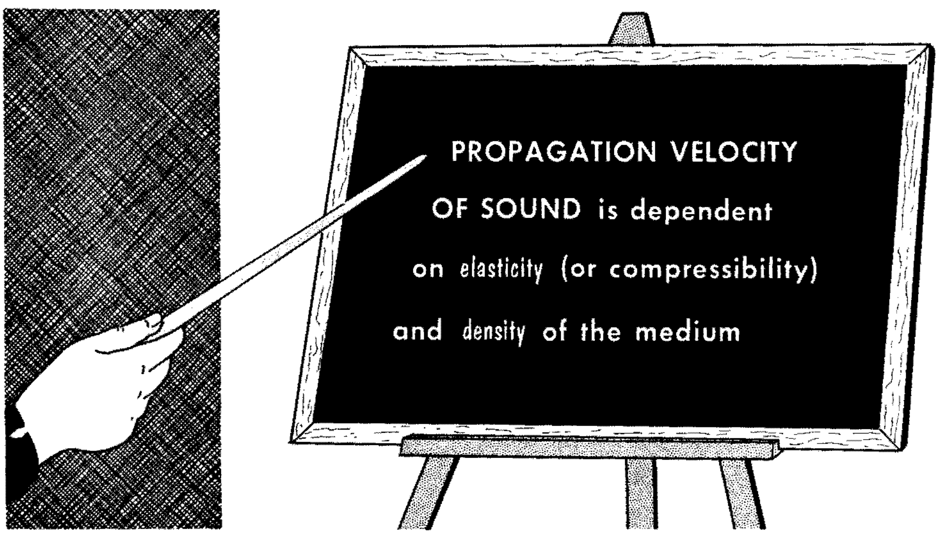
Note that, after the wave has passed, the air stops moving. A body of air does *not* move with the wave, but the energy in the wave is passed on from particle to particle, in much the same way that (in the illustration) the impact of one penny is transmitted along a line of pennies.



The natural speed, or propagation velocity, of sound is controlled by two properties of the air through which it travels:

1. How much squeezing more of it into a given space will increase its pressure—its compressibility (known technically as elasticity). This feature takes control of the air piled up in front of the supersonic jet.
2. How much force is needed to get a given quantity of it moving, or to stop its motion—its *density* (or mass per unit volume). This feature is responsible for carrying the draft created by the fan.

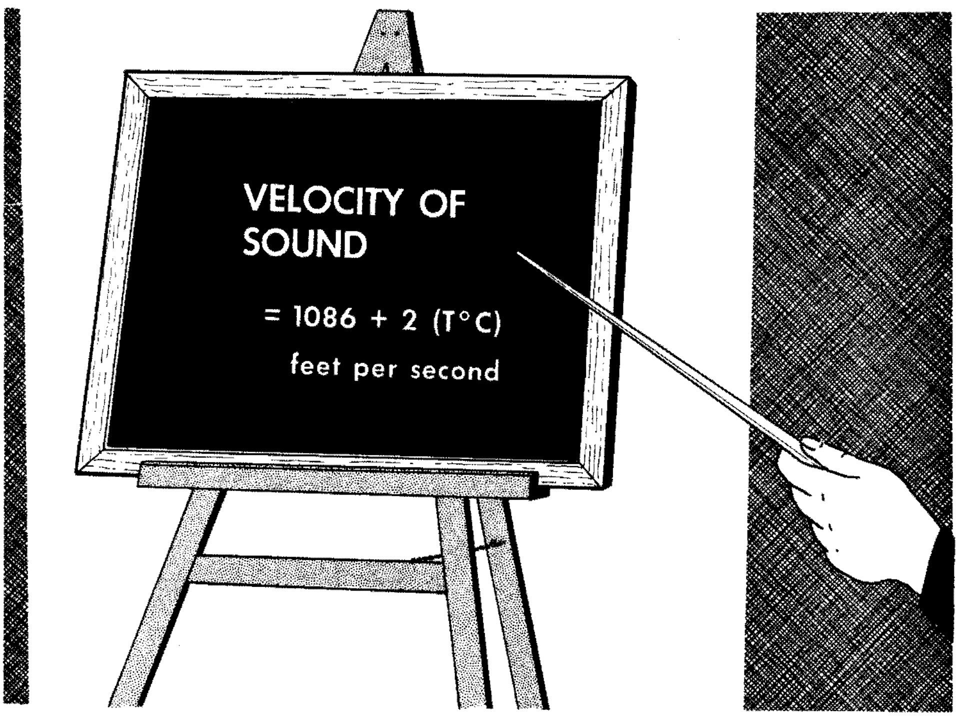
The natural speed, or propagation velocity, of sound in air is roughly 1100 feet per second, or 750 miles per hour.



How Fast Sound Travels

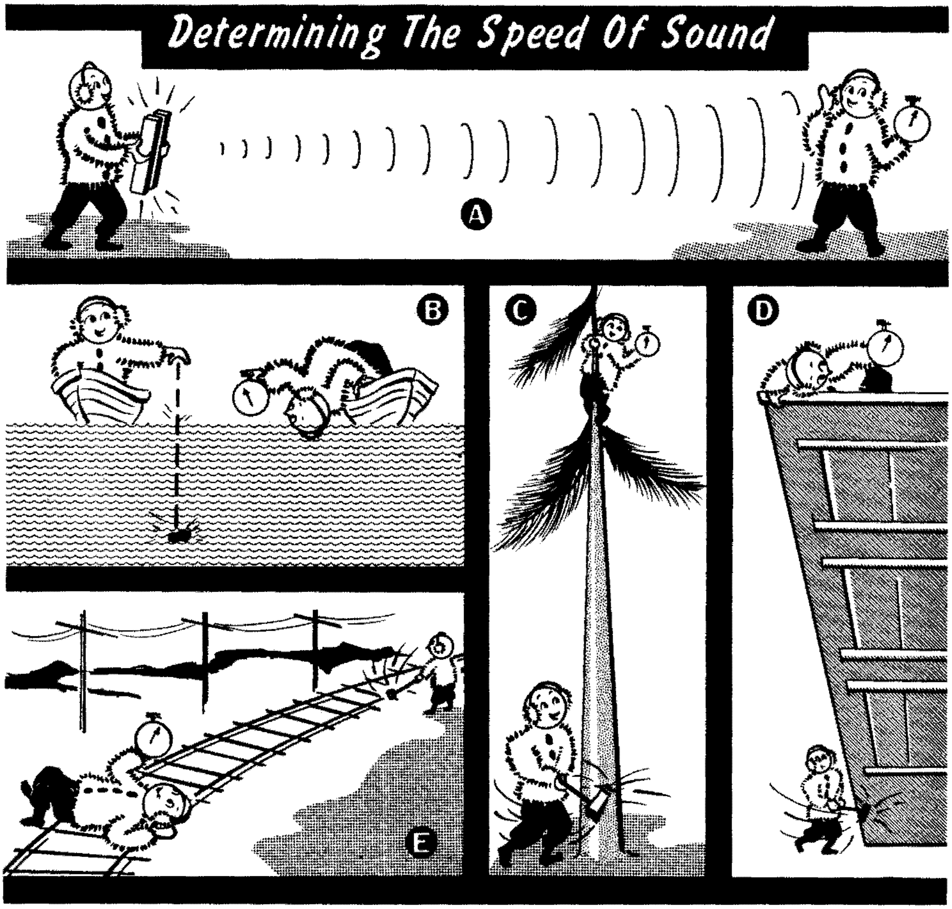
The speed, or propagation velocity, of sound depends on the elasticity of air and on its density. Although we do not normally measure these properties, we can easily find the temperature and barometric pressure of the air, which affect both its density and elasticity. What we would like to know is: "How does the speed of sound change with atmospheric pressure and temperature?"

Experiments and advanced theory provide the following answers:
1. Change in barometric pressure alters the density of air (how much of it there is in a given space) and its compressibility (the more compressed it is, the more it resists further compression). Both change together and neutralize their effect on velocity of sound, so it is not appreciably dependent on barometric pressure.
2. Change in temperature also affects both the density and compressibility of air, but not so as to neutralize as pressure does. A useful formula says that the velocity of sound in air is approximately 1086 feet per second at 0° Centigrade (the freezing point of water), and rises 2 feet per second for each Centigrade degree rise in temperature.



At 20° C (normal room temperature), sound travels at $1086 + (2 \times 20) = 1126$ feet per second.

How Fast Sound Travels (contd.)

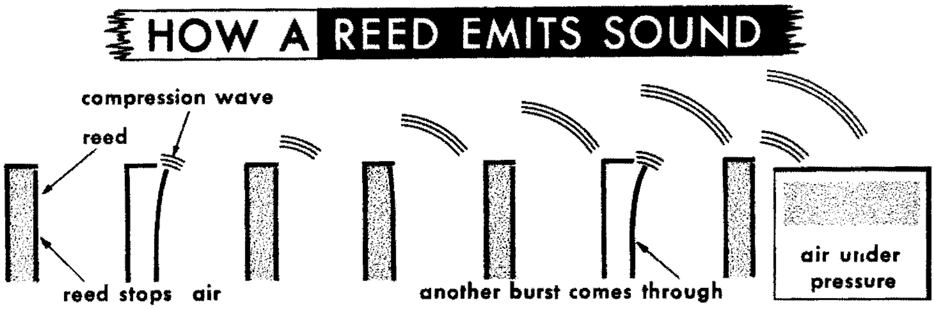


Sound also travels in almost any medium besides air, because everything (except a vacuum) possesses the two necessary properties, density and compressibility. (Even the most "incompressible" substances "give" a little.) For example, water is about 80 times more dense than air, but it also offers an even greater opposition to compression. Owing to its greater density, we should expect sound to travel more slowly in water than it does in air. Water, however, is less compressible, which tends to make sound travel faster. The combined effect makes the speed of sound in water about 4.7 times faster than in air. The following table gives the speed of sound in different substances measured at 0°C:

Air	1086 feet per second	Brick	11,980 feet per second
Water	4700 feet per second	Iron	16,000 feet per second
Pinewood	10,900 feet per second	Steel	16,360 feet per second

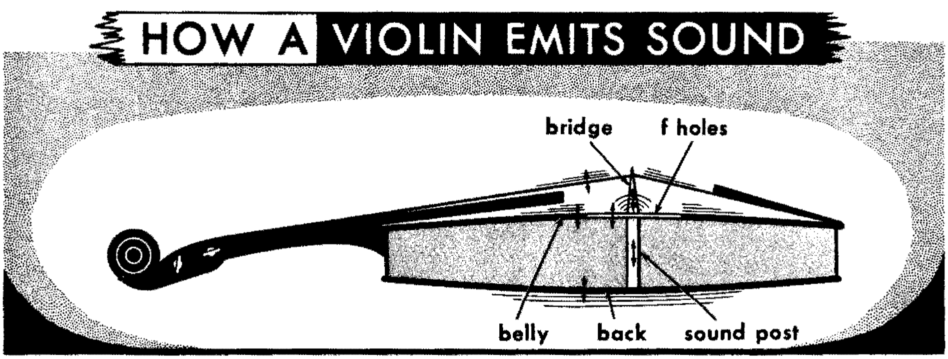
Sources of Sound

Most sounds are more than a single "pulse" of sound, such as a hand-clap. For example, the reed of a harmonica or accordion vibrates at its resonant frequency and allows the air to be emitted in "bursts." Each burst of air pressure and movement is radiated in the same way as a single hand-clap.

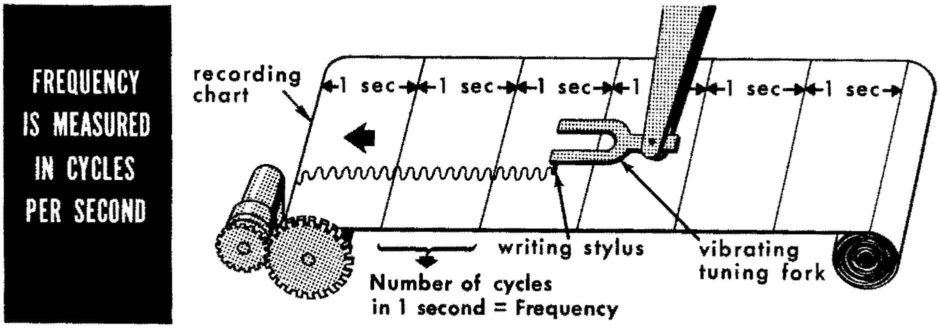


There are also other forms of vibration that produce sound: a vibrating piano string causes the sound board of the piano to vibrate. This moves the air in contact with it, producing alternate waves of compression and expansion *(rarefaction)*. All stringed instruments use the same principle; in the violin, the string vibrations are transmitted to the body of the instrument, which moves the air in contact with it. In addition, there are wind instruments in which a column of air inside a tube vibrates in a manner controlled by the internal dimensions of the instrument. The contact of the air column with outside air at one end of the column or through an opening allows sound to be radiated.

All these sounds are produced by a *periodic* vibration at regular intervals (definite *frequency*). Other sounds are not rhythmic, and do not give musical tones because they are due to vibrations that are not regular or periodic—clapping, rattling, and scraping sounds, and all kinds of noises, like those that come from a factory, street, or kitchen, or even voices, except when singing.



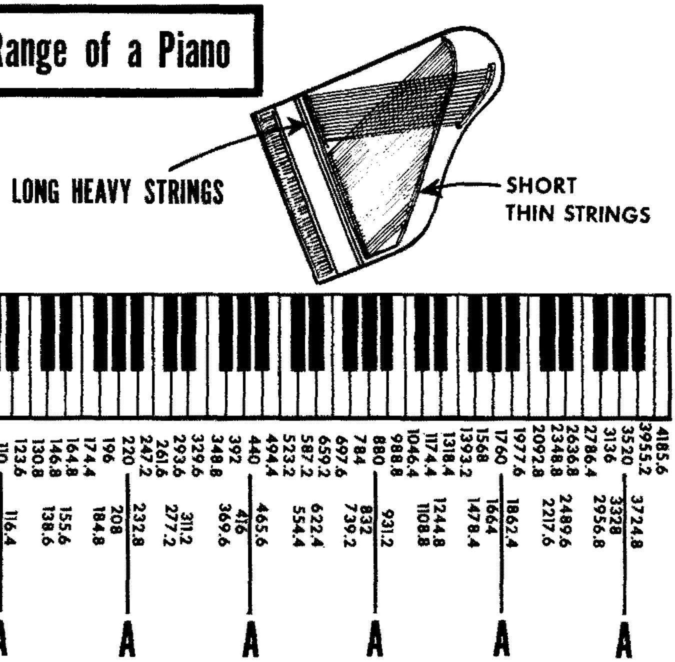
Frequency and Pitch



The frequency of a sound is a measure of how many vibrations occur in a given time; it is usually measured in vibrations per second. When the sound vibrations are converted into electrical waves for amplification, the frequency is referred to as *cycles per second,* often called *cycles* for short.

In music, difference of frequency is recognized as a change in *pitch.* In a piano, for example, the long, heavy strings vibrate slowly (you can actually see the individual vibrations), whereas the short light ones vibrate very rapidly. The slow (low-frequency) vibrations are recognized musically as low in pitch; the rapid (high-frequency) vibrations are recognized musically as high pitched.

The Frequency Range of a Piano



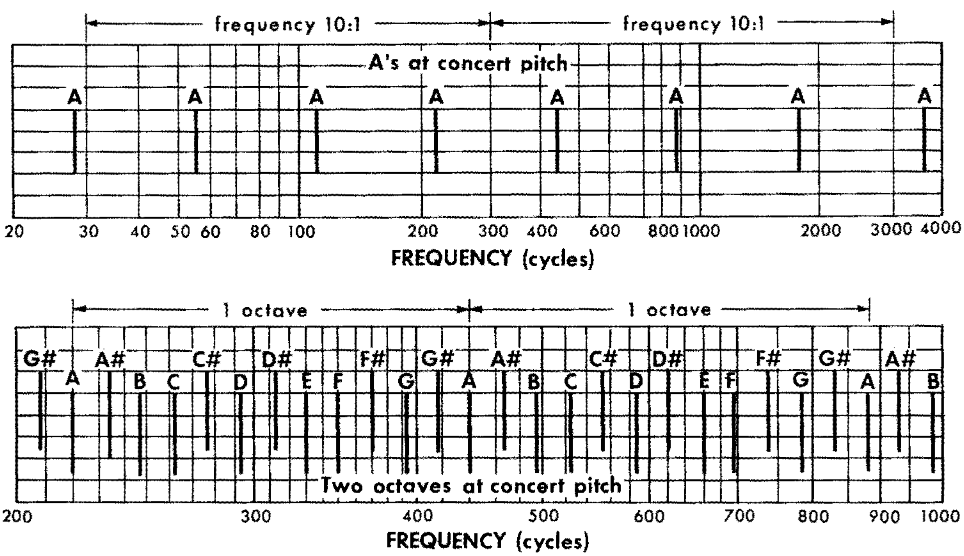
Frequency and Pitch (contd.)

Every time frequency is *doubled,* pitch changes one *octave.* A two-octave rise in pitch quadruples the frequency; a three-octave rise in pitch multiplies it eight times. Any two adjacent notes on a piano keyboard have a constant ratio between their frequencies; since there are twelve notes per octave, this ratio is a number that, multiplied by itself twelve times, equals 2. The difference in frequency between two successive notes on the keyboard is not constant, but the *ratio* between their frequencies is constant; the frequency of the upper note is 1.059 times that of the lower one.

$$1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059\times1.059 = 2 \text{ (very nearly)}$$

In ordinary arithmetic, the difference between successive numbers is always the same—one. A system of numbers or quantities where successive items are separated by a *constant fraction,* rather than by the same amount, is called a *logarithmic system.* The relationship between frequency and pitch is logarithmic, as may be seen from the fact that to raise the pitch each octave, the frequency has to be *multiplied* by a power of two, and to raise the pitch by any desired musical interval, the frequency has to be *multiplied* by a figure corresponding to that pitch interval.

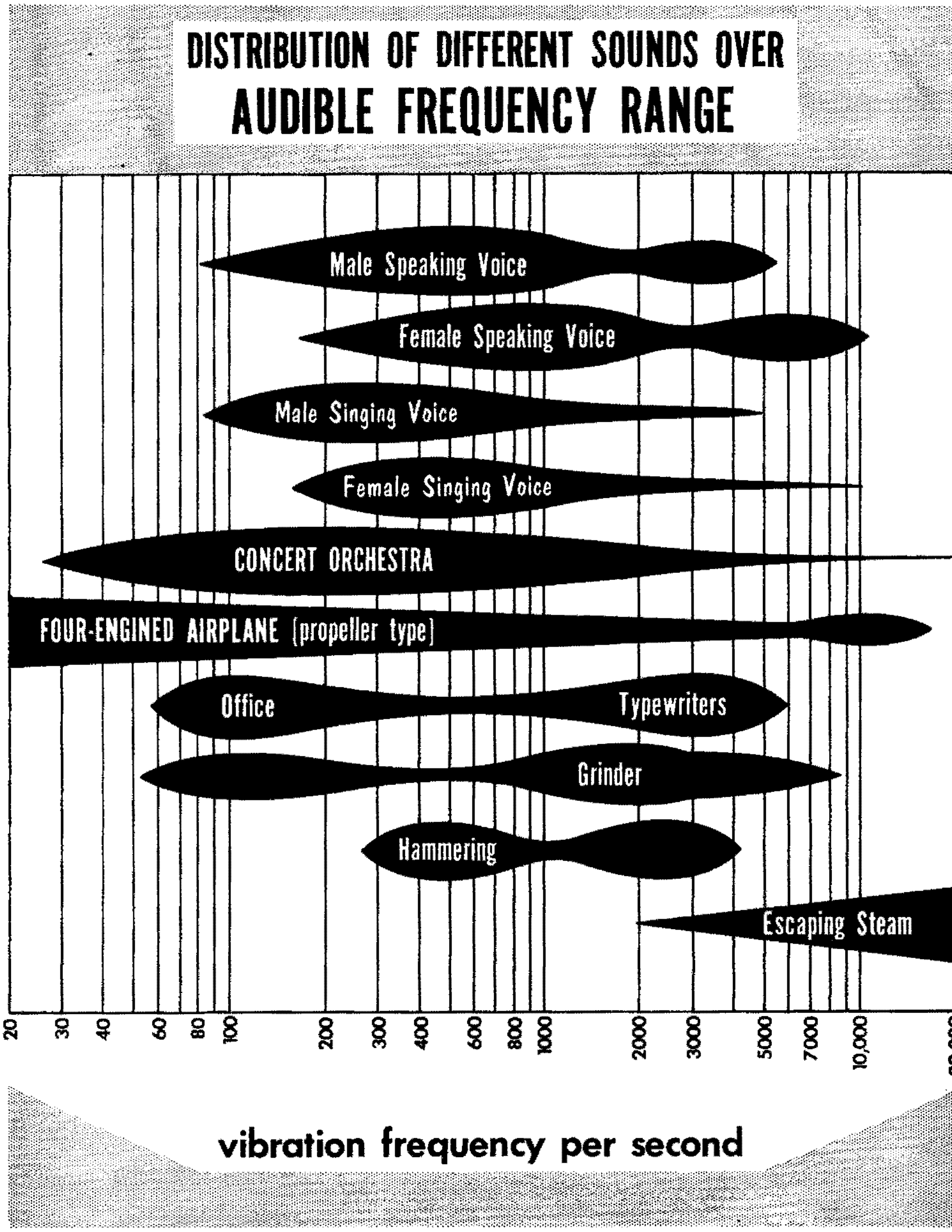
This can be shown visually by using "logarithmic" paper, where the pattern of lines is repeated for every numerical increase of ten times, and the distance covered by each change of ten times is the same. Similarly, each ratio of two is the same distance on the scale, as is any other ratio.



PLOTTING FREQUENCIES ON A LOGARITHMIC SCALE

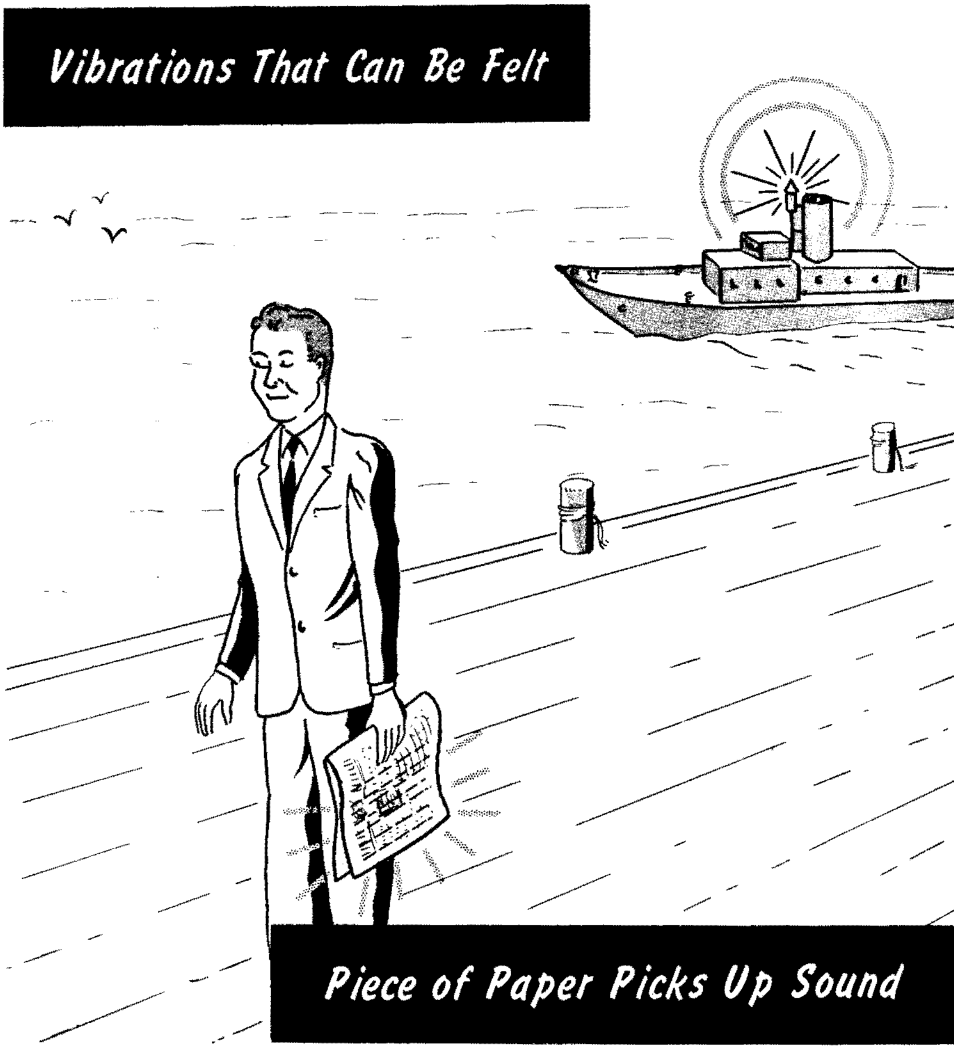
Frequency and Pitch (contd.)

Different kinds of sound, both musical and otherwise, are characterized by different ranges of frequencies:



Intensity and Loudness

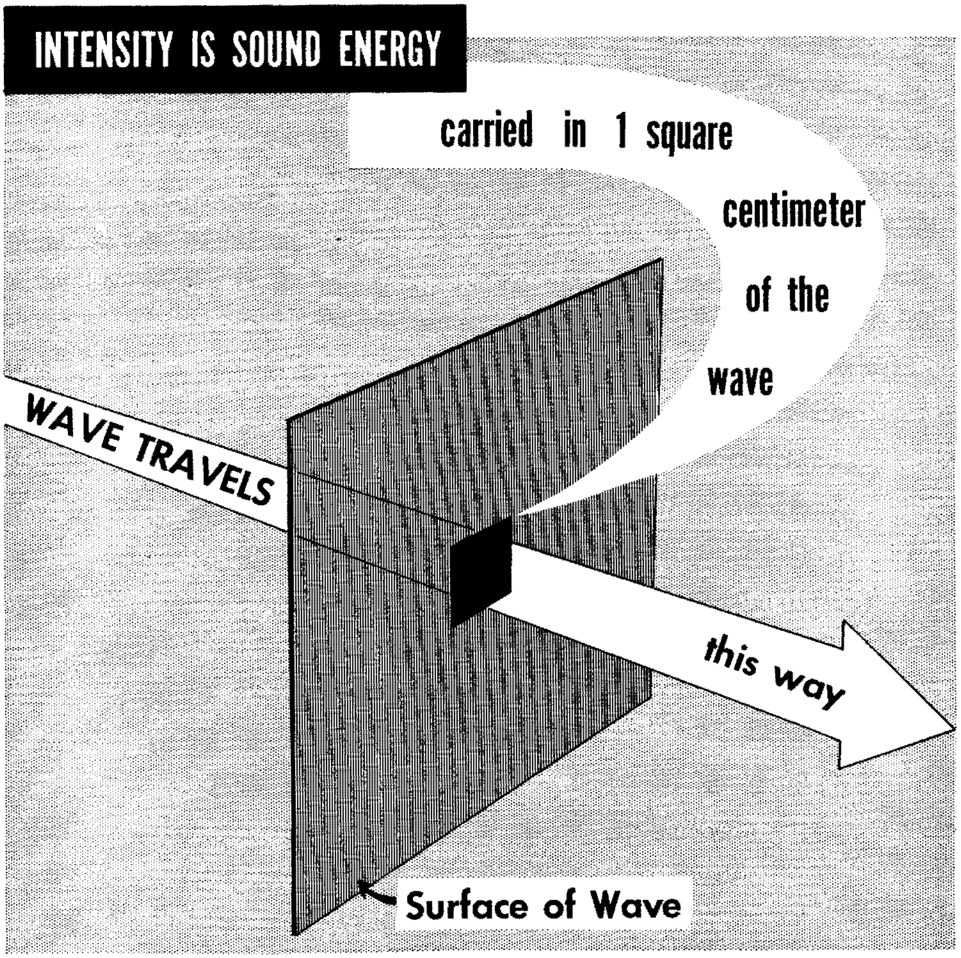
As well as having different frequencies, sounds also differ in loudness. This means that the vibrations we hear are of greater or smaller intensity. The horn of an ocean liner produces such an intense vibration that you can feel it as well as hear it. A piece of paper held near the source of such a sound will vibrate hard enough to numb your fingers.



The ticking of a watch, however, is of very low intensity. Unless the surroundings are fairly "quiet" you may be unable to hear it at all. Certainly you would never hear it near the horn of an ocean liner.

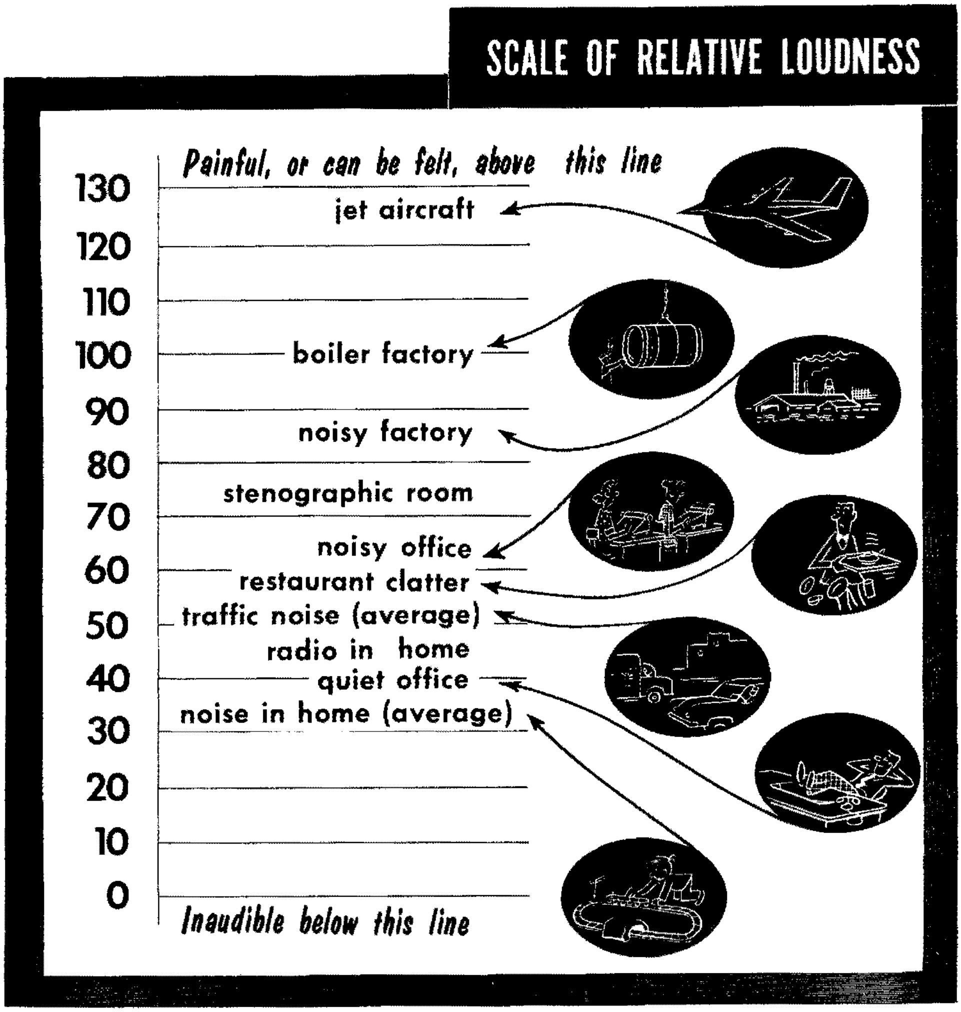
Intensity and Loudness (contd.)

Sound intensity is a measure of the *acoustical power* transmitted by the sound wave. Intensity is measured in terms of a certain section of the wave, specified as one square centimeter in scientific measurements. (There are 6.45 square centimeters in a square inch.)



Any kind of power can be measured in watts. This is true of sound waves. One-tenth of one quadrillionth of a watt (.000,000,000,000,0001 watt) of sound power passing through an area of one square centimeter is not quite audible, using a vibration frequency of 1000 cycles per second. One-quadrillionth of a watt (.000,000,000,000,001 watt), however, is easily audible. A sound that is loud enough to be almost painful represents an intensity of less than one-thousandth of a watt per square centimeter.

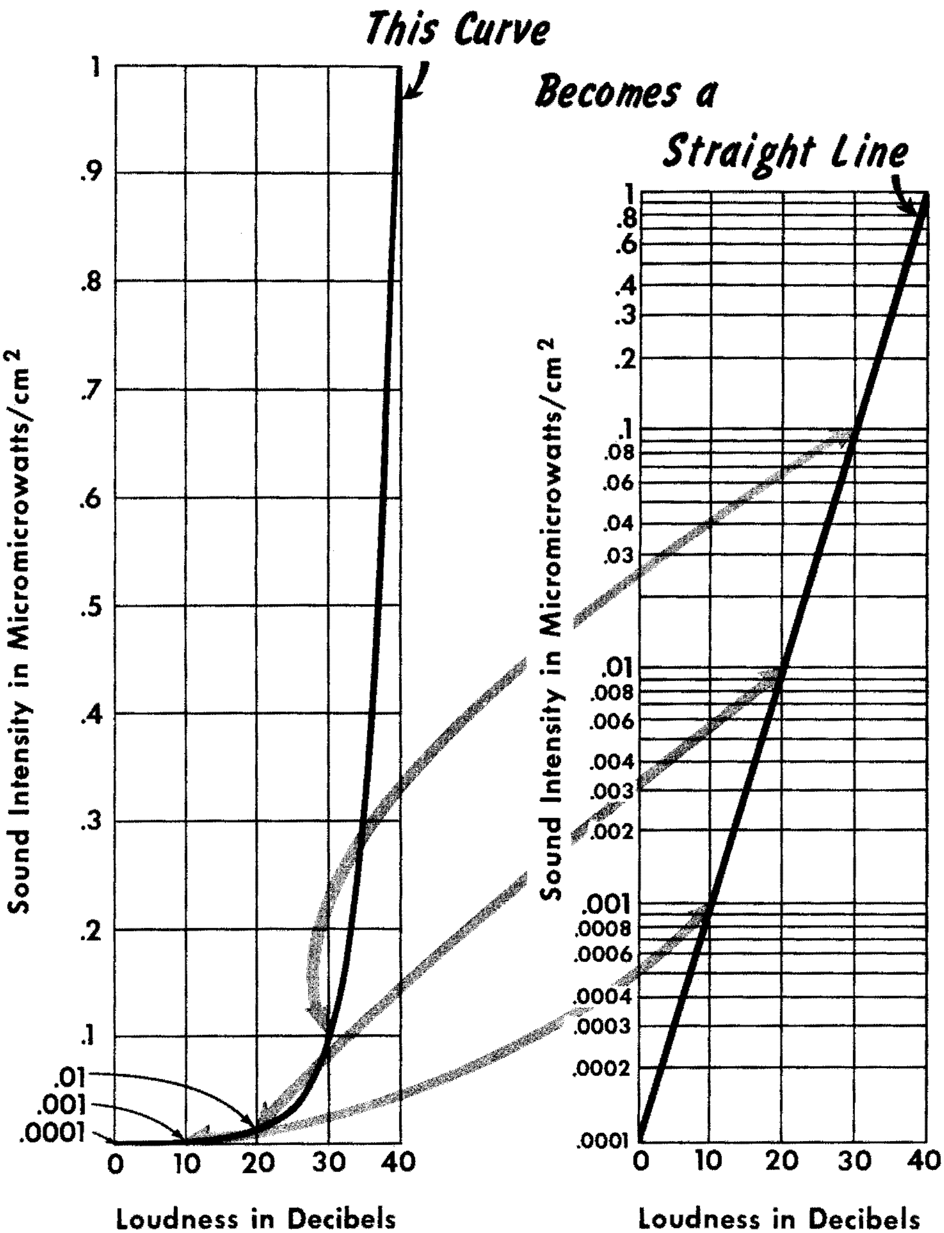
Intensity and Loudness (contd.)



Every time that the intensity (or power) of a sound wave is *multiplied* by ten, it sounds louder by about the same amount. (It sounds as if a similar "quantity" of sound has been *added.)* A change in intensity of ten times does not represent as great a change in loudness as one might expect. In fact, a change in intensity of 26% is just barely detectible. The range between the intensity at which a 1000-cycle sound is first heard (the *threshold of hearing)* and a point at which an increase in power ceases to give the impression of further increase in loudness is a trillion times. Thus each multiplication of ten in intensity is equivalent to about one-twelfth of the range from audibility to saturation.

Intensity and Loudness (contd.)

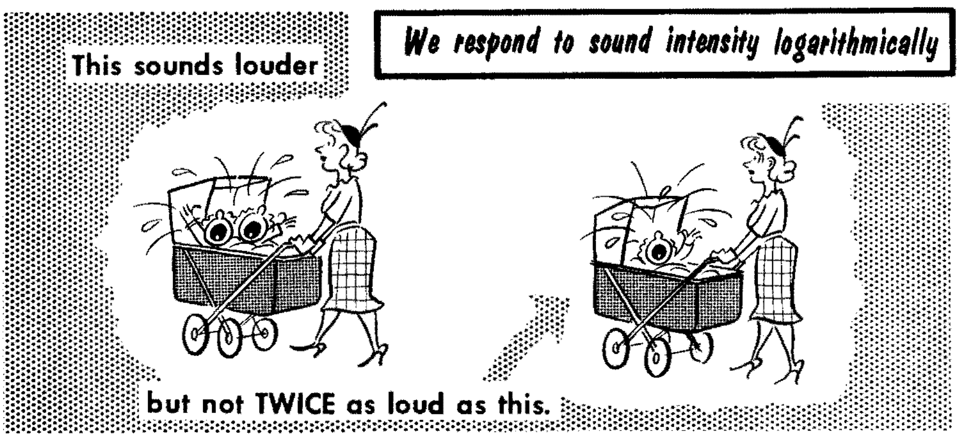
WHEN INTENSITY IS PLOTTED ON A LOG SCALE,



Intensity and Loudness (contd.)

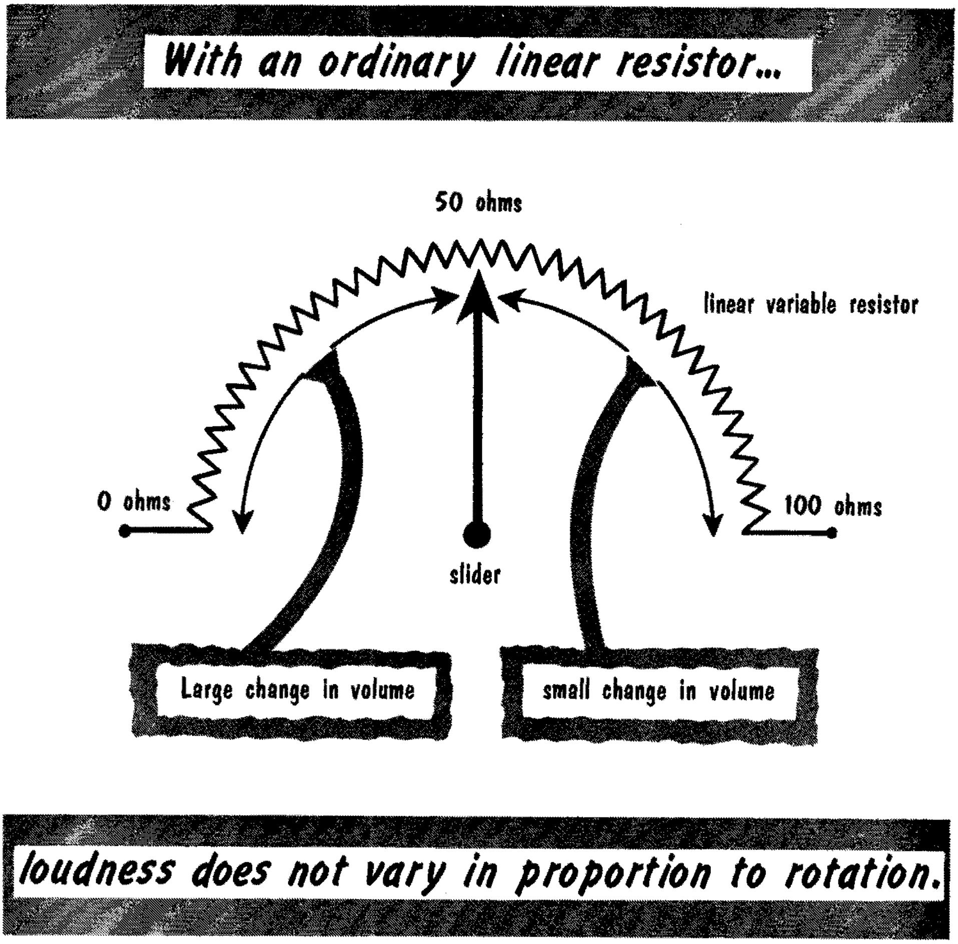
Loudness increases by equal amounts not with equal additions of sound intensity, but rather with equal multiplications of intensity. In this respect, therefore, our response to a change in intensity is similar to our response to a change in frequency—it is logarithmic. (Note the similarity of the graph of intensity and loudness to the graph of frequency and pitch; both are plotted on a logarithmic scale.) The fact that this logarithmic relationship exists is generalized in a principle known as Fechner's Law, which states that "For a sensation to increase in arithmetic proportion, the stimulus must increase in geometric progression." Fechner's Law may be applied to the other senses as well as to hearing.

The basis of the loudness scale is a multiplication factor of 10. This unit (called a *bel)* is inconveniently large, since there are only 12 bels in the entire useful range of audibility at 1000 cycles. For this reason, a smaller unit, the *decibel* (one-tenth of a bel) is more commonly used. (Thus the range of useful audibility at 1000 cycles is 120 decibels.) The decibel (abbreviated *db)* is also a convenient unit because it represents the 26% intensity change that is the smallest possible change an average person can hear in the range of loudness at which the ear is most sensitive to change. Over most of the range a 2-db change is difficult to detect, and at higher levels an even greater change is necessary.



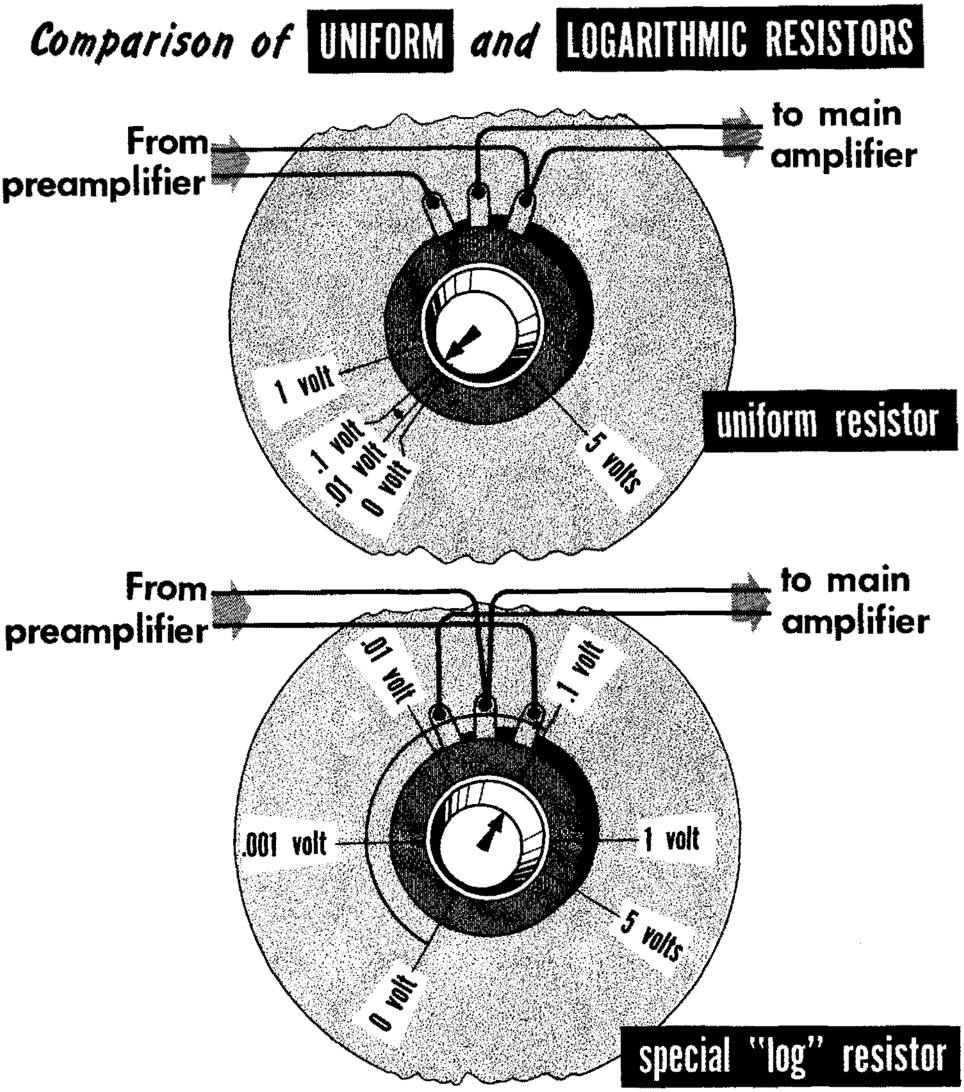
Because our response to sound intensity is logarithmic, our impression of loudness can be quite deceiving. If, for example, we are listening to the radio at low volume, we may not realize just how low the volume is, until an airplane passes over. The radio seems to become even less loud as the sound of the airplane drowns it out and our ears become less sensitive to the quieter sound. (This effect, known as *masking*, will be discussed in greater detail later.) Our instinctive reaction when this occurs is to turn up the volume control.

Demonstration of Relative Loudness



This fact of the logarithmic sensation of loudness can be very easily demonstrated with a simple test, using an amplifier with a separate volume control and a loudspeaker. (A volume control is a resistor with a sliding contact; if an audio program is fed through the resistor and only part of it is picked off by means of the slider, the volume can be changed by moving the slider.) If the volume control is an ordinary variable resistor (with the resistance uniformly distributed), the intensity will be proportional to the angle between the slider position and the "zero" position. Using this kind of resistor for a volume control, the loudness does not sound as if it varies in direct proportion at all. In the first few degrees of rotation there is a big change in loudness, but further rotation makes hardly any additional change.

Demonstration of Relative Loudness (contd.)

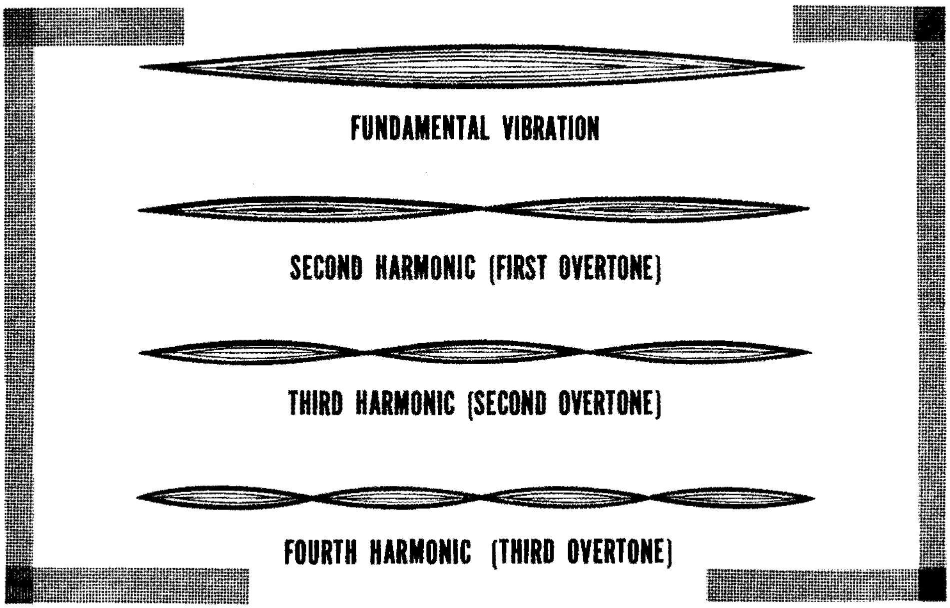


For this reason, variable resistors intended for use as volume controls are made differently. The resistance is not uniformly distributed, but rather is proportional to the *logarithm* of the angle of rotation from the zero end. Thus 50° of movement of the slider allows the passage of one-thousandth of a volt of audio signal; 100° may produce one-hundredth of a volt; 150°, one-tenth of a volt; 200°, one volt, and so on. When this kind of control is used, the volume, or loudness, seems to be proportional to the amount of rotation of the control.

Harmonics or Overtones

We have described the two main properties of any single sound: frequency (which we recognize as pitch) and intensity (which we recognize as loudness). But there are other differences by which we can tell one sound from another, even if both are of the same frequency and intensity (pitch and loudness). For example, a violin and a flute do not sound the same, even when they play the same note equally loudly. The difference in sound *quality* or *timbre* arises from the fact that every note on any musical instrument consists of not just one frequency, but a combination of several frequencies.

A STRING VIBRATING IN DIFFERENT WAYS PRODUCES DIFFERENT OVERTONES.



These additional frequencies are multiples of the lowest frequency (called the *fundamental)*, which is usually the one that determines the pitch of the note. If the fundamental is 440 cycles per second (A above middle C), the same instrument will produce a range of frequencies that are multiples of 440 cycles at the same time: 2×440, or 880 cycles, 3×440, or 1320 cycles, and so on up the scale. What makes the different instruments easily distinguishable is the fact that these *overtones* (or *harmonics)* may be present in different relative intensities; some may be absent entirely. Each instrument has its own characteristic "selection" of overtones in characteristic relative intensities.

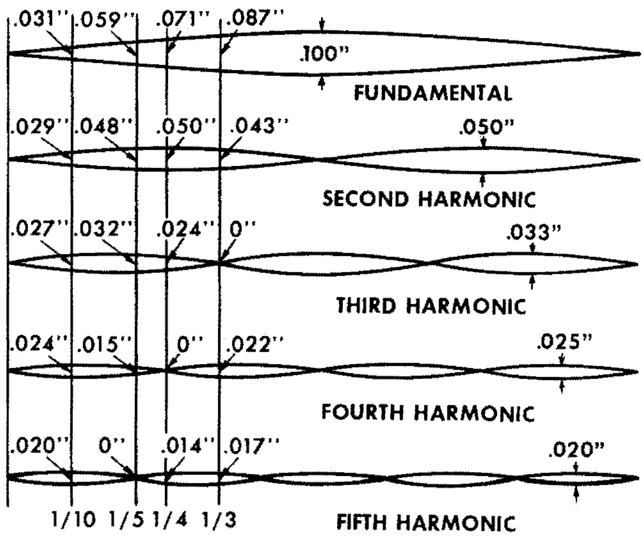
Harmonics in Strings

That it is differences in the overtone structure that account for the different "sound" of various instruments can be verified with an electric guitar. With this instrument, the strings themselves produce relatively little sound. An electrical pickup placed close to the strings is used to pick up their vibration, which is amplified electronically, the sound being heard from a loudspeaker. If you take one of these guitars and try moving the pickup to different parts of the string, you will find that the *timbre* of the note changes.

Assume that after the string has been set vibrating, the amplitude of vibration of the string is 1/10 inch at the lowest frequency, 1/20 inch at the next above this, which is twice the frequency (the second harmonic or first overtone), 1/30 inch at the third harmonic or second overtone, 1/40 inch at the fourth harmonic or third overtone, and so on. Now suppose that we move the pickup to positions 1/10, 1/5, 1/4, and 1/3 of the length of the string from one end. The differences in timbre will be clearly audible. From the drawings we can find the magnitude of vibration at each point along the string for the fundamental and each harmonic. These can be made into a table:

Tone	*Maximum vibration in inches*	*Vibration at pickup points in inches* one tenth	one fifth	one fourth	one third
Fundamental	.100	.031	.059	.071	.087
2nd harmonic	.050	.029	.048	.050	.043
3rd harmonic	.033	.027	.032	.024	.000
4th harmonic	.025	.024	.015	.000	.022
5th harmonic	.020	.020	.000	.014	.017

EXTENT OF VIBRATION *at various pickup points* FOR THE FIRST FIVE HARMONICS

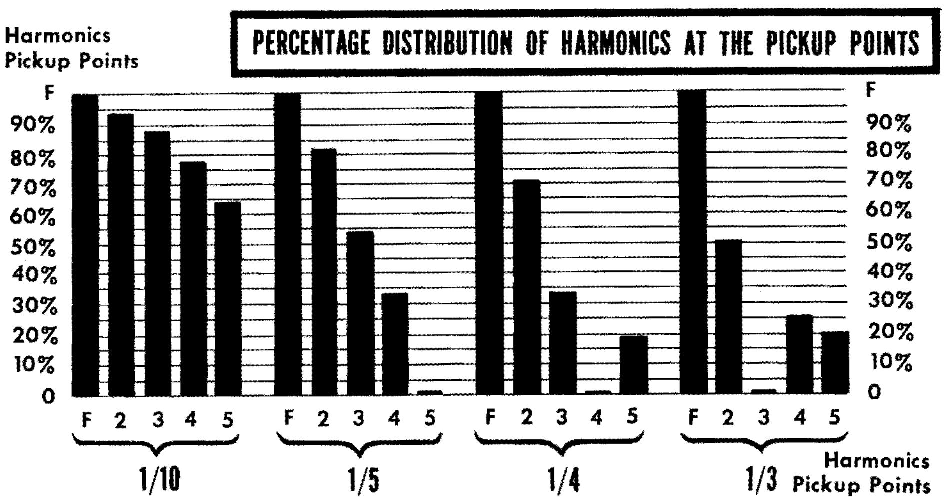


Harmonics in Strings (contd.)

Of course there is less output from the positions nearer the end, but this can be compensated for by extra amplification, so the tonal quality, as judged by the ear, will depend on how much of each overtone there is, *compared to the fundamental*. Let us adjust the table to give this information, by expressing the strength of each harmonic as a percentage of fundamental. We do this by dividing each vibration figure by that for the fundamental at the same point, and multiplying by 100%. For example, when the fundamental vibrates 0.031 inch and the harmonic is 0.029 inch, the harmonic is $(0.029 \times 0.031) \times 100\% = 93.7\%$. The table, completed in this way, then looks like this:

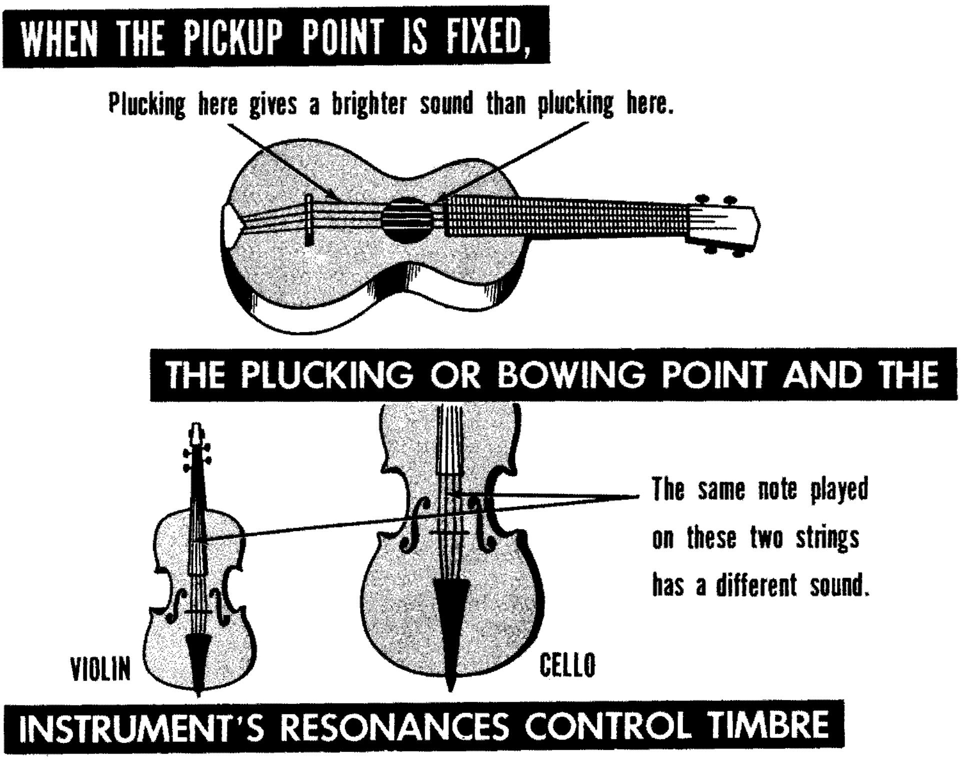
		Vibration percentages			
			Harmonics		
Pickup point	*Fundamental*	*2nd*	*3rd*	*4th*	*5th*
one tenth	100	93.7	87.1	77.5	64.5
one fifth	100	81.4	54.3	25.4	0
one fourth	100	70.5	33.8	0	19.8
one third	100	50.0	0	25.0	20.0

Now notice the difference in composition: the first position, 1/10 from the end, gives a large proportion of all harmonics; even the fifth harmonic is more than half as strong as the fundamental. The second position, 1/5 from the end, eliminates fifth harmonic, and leaves the others in different strengths. Each position gives different proportions of harmonics. But the most important difference is that the further we go from the end, the weaker *all* the harmonics get, compared with the fundamental. At 1/10 from the end, all the harmonics have more than half the amplitude of the fundamental. At 1/3 from the end, only the second harmonic is even half as strong as the fundamental; the others are much weaker.



Harmonics in Strings (contd.)

From this we can see that the tone quality, or timbre, of a stringed instrument will vary considerably if we change the point along the string where the sound is picked off. In most stringed instruments, the sound is taken from the bridge, which transmits it to the body of the instrument. This means that the pickup point is fixed.



In this case, the precise point at which the string is plucked or bowed influences the harmonics or overtones of the string that are set in vibration together with the fundamental, thus changing the tone quality. In addition, the natural vibration properties of the body of the instrument (called resonances) influence the relative strength of the various frequency components as they are radiated into the air as sound. This accounts for the characteristic differences between different instruments using the same kind of strings.

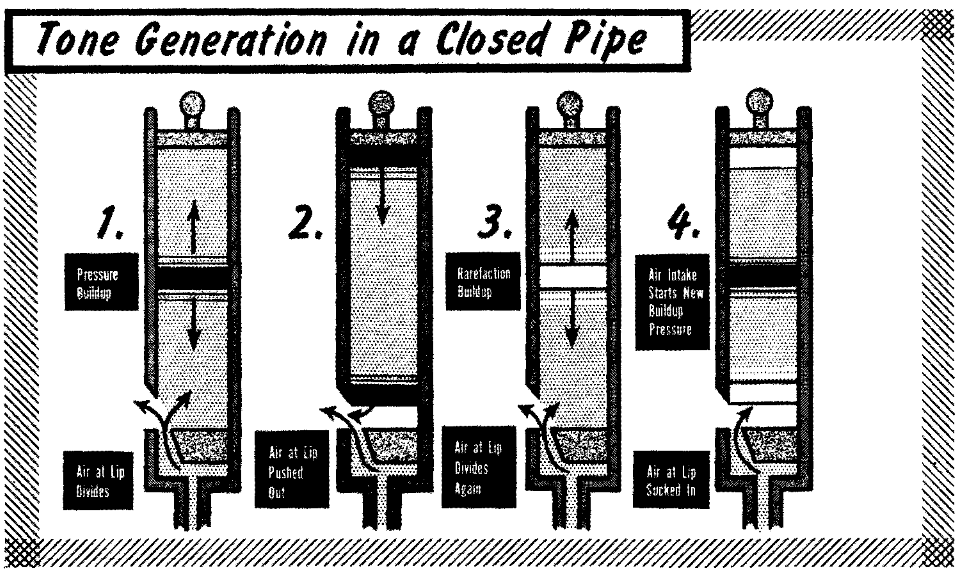
We can extend the same general idea of overtones, and the way they are excited, to other kinds of instruments, using pipes, reeds, vibrating bars or rods, or other basic sound generators—even triangles, bells or drums. Variations in the complex pattern of harmonics give each instrument its own character.

Harmonics in Organ Pipes

Variation in overtone structure makes different stops on an organ give different tonal qualities. In a pipe organ, the wide pipes produce tones in which the fundamental predominates. Their sounds are deep and smooth. The thinner pipes of the same length produce many more harmonics than the wider pipes, and give full-bodied sounds. Still thinner pipes suppress the fundamental and give a thin or reedy tone. In an electronic organ, the harmonics put in (or left out) are controlled electronically, giving a similar variety of "tone color."

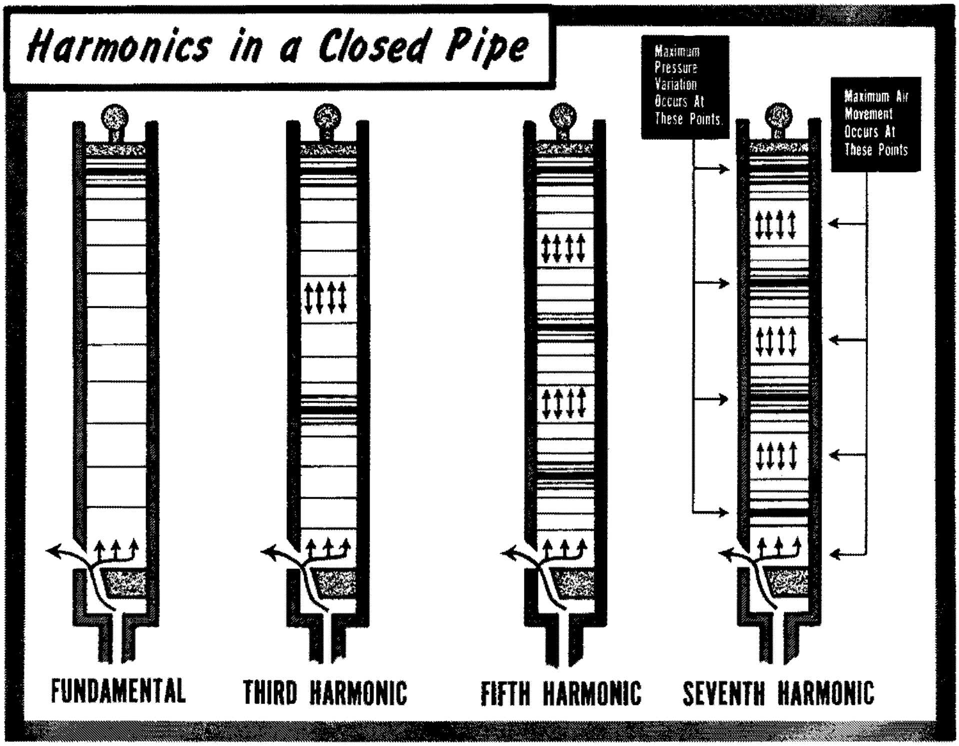
The pattern of overtones produced by a pipe also depends on whether the far end of the pipe is open or closed. Acoustic vibration is set up in the air column by a sheet of air directed toward the upper lip of the pipe. When this air passes inside the lip, it starts to compress the air inside the pipe. This inside pressure soon forces the air out again. As a result, the blown air alternately goes inside and outside the pipe, with a considerable back and forth movement of air particles occurring at the pipe mouth.

The frequency of oscillation at the mouth is determined by the pipe. The pressure wave travels up the pipe at the speed of sound. When it gets to the far end, if the end is closed, the pressure increases because it cannot be passed on. The pressure wave is therefore reflected back toward the mouth of the pipe. When it reaches the mouth, the air directed against the lip is forced to go outside the pipe. Since this represents half of a complete cycle and the pressure wave has traveled twice the length of the pipe, it is clear that a closed organ pipe is one quarter-wavelength long for its lowest frequency.



Harmonics in Organ Pipes (contd.)

The air at the mouth of a closed pipe can move freely, whereas the air at the far end cannot move at all. Thus the lowest (fundamental) frequency of such pipes is determined by the length of a wave that travels up to the top, produces increased pressure, and travels back down the pipe to produce outward motion of air at the mouth.

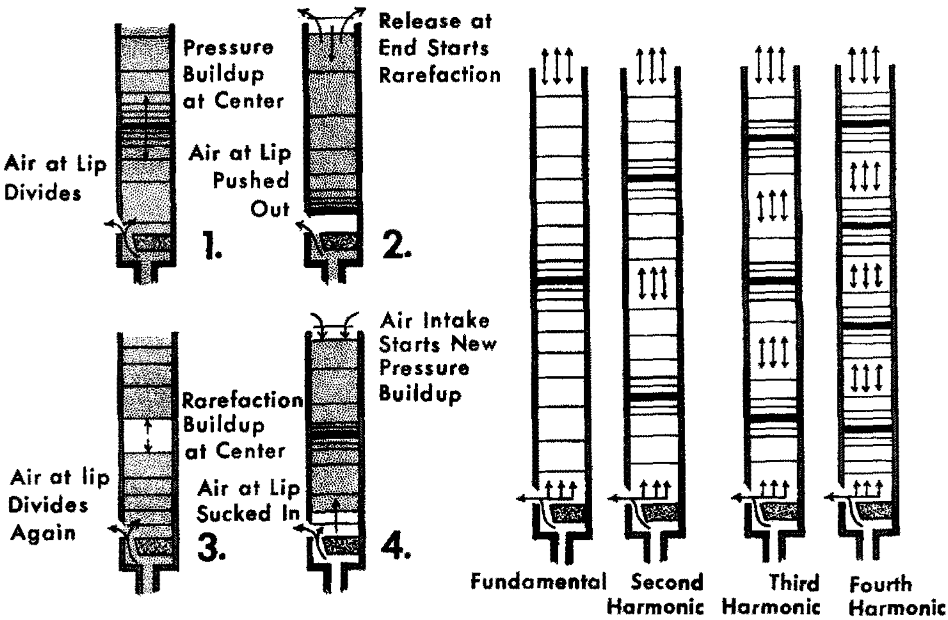


Any wave whose length is such that the return journey gets it to the mouth when the blown air is moving in the opposite way to the initial pulse—where the pipe is ¼ wavelength long, ¾ wavelength long, 1¼ wavelengths long, and so on—will tend to be present in the pipe. These wavelengths correspond to the *odd* harmonics of the fundamental. The even harmonics (whose wavelengths are such that the pipe is 1½ wavelength long, 1 wavelength long, etc.) are absent. Notice that, at the mouth and for waves other than the fundamental, at half-wave distances along the pipe, the movement of air in the going wave adds to that in the return wave, to produce a point where air *movement* is a maximum. At the closed end, and for waves other than fundamental, at half-wave distances from the closed end, the air pressure in the going and returning waves adds, producing a point where air *pressure* fluctuation is a maximum.

Harmonics in Organ Pipes (contd.)

If the far end of the pipe is open, the pressure will drop suddenly because the wave is no longer confined in the pipe. In this case, the pressure wave is reflected as a *rarefaction.* When the reflected wave reaches the mouth, the air blown against the lip is directed into the pipe again and thus completes a full cycle while the wave and its reflection have traveled twice the length of the pipe. Thus the fundamental frequency of an open pipe is such that the pipe is half a wavelength long. Any wave whose length is such that the length of the pipe is an exact multiple of a half-wavelength—1 wavelength, 1½ wavelengths, etc.—can be sustained in an open pipe. These wavelengths correspond to all the harmonics of the fundamental frequency, and an open pipe produces a complete harmonic series.

TONE GENERATION AND HARMONICS IN AN OPEN PIPE



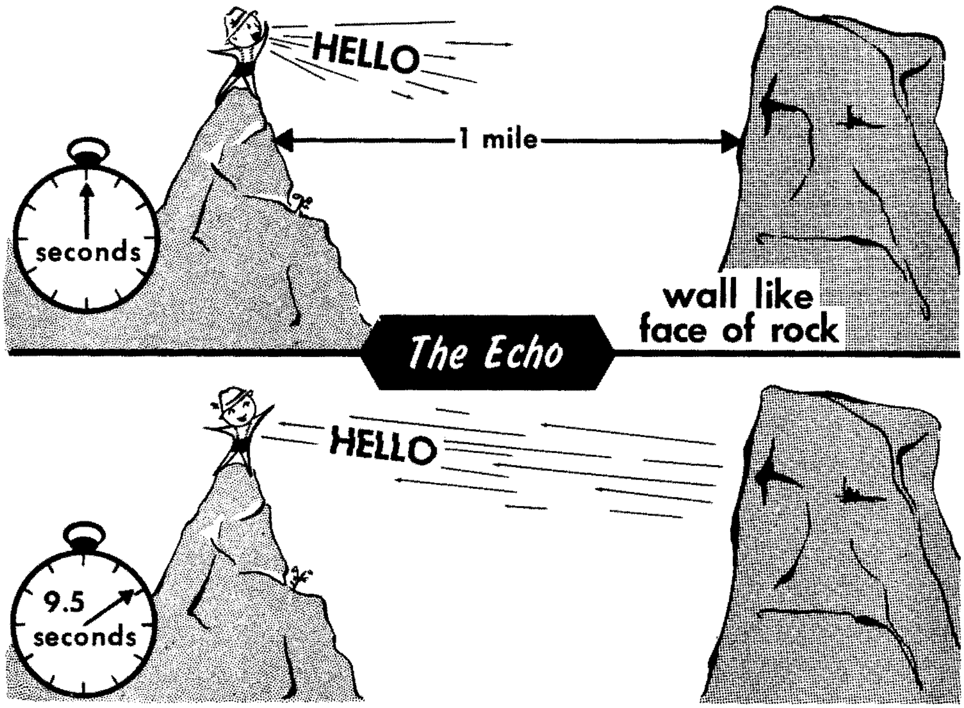
Just as in the closed pipe, there will be points where the pressures add up, and points where movements add up. Because of this action inside pipes, the waves do not seem to move. At the points where pressure has a maximum fluctuation, as at the end of a closed pipe, the air does not move, but changes in pressure according to the sound wave. At the points where movement is a maximum, as at the open end of an open pipe, or the mouth of either, there is little pressure fluctuation and a maximum of movement. For this reason, waves of this kind are called *standing waves.*

QUESTIONS AND PROBLEMS

1. What two kinds of air movement combine to produce sound waves? Illustrate with a wave started by bursting an inflated paper bag.
2. What two properties of air (or any other medium through which sound travels) control propagation velocity of sound waves?
3. How does the speed of sound change with (a) barometric pressure, (b) atmospheric temperature? State any approximate rule that can be used.
4. State why you think sound waves travel faster in steel than in iron or brick.
5. What part of (a) a piano, (b) a violin, is responsible for radiating sound waves into the air?
6. Explain the relationship (a) between frequency and pitch, (b) between intensity and loudness. What frequency interval corresponds with a pitch interval of one semitone in music?
7. What is the ratio between intensities corresponding to (a) the full range of average human hearing, (b) the smallest change in loudness that can be detected with very careful listening?
8. How would Question 7 be answered in decibel units?
9. How does Fechner's Law explain the usefulness of the decibel scale?
10. Why do potentiometers for use as volume controls have a logarithmic gradation?
11. What are the three basic properties of a musical tone? On what properties of sound does each of these properties depend?
12. Why can a string vibrate at more than one natural frequency? Explain the relationship between the different frequencies at which it can vibrate.
13. Why do you think strings for the lowest notes on the piano are "loaded" by having a spiral of wire wound on over the central stretched one?
14. Suggest why the same violin, played by an accomplished musician, gives a much sweeter tone than when played by a novice.
15. What is the basic difference in the overtone structure of organ pipes with the "far" end open or closed?
16. If the same pipe is provided with a removable end plug, how will its pitch change from open to closed?
17. Why is it that a narrow organ pipe sounds thin or reedy, while a wide pipe gives a deep, smooth tone?
18. What are standing waves? Explain how they build up inside (a) a closed and (b) an open organ pipe.

What Rooms Do To Sound

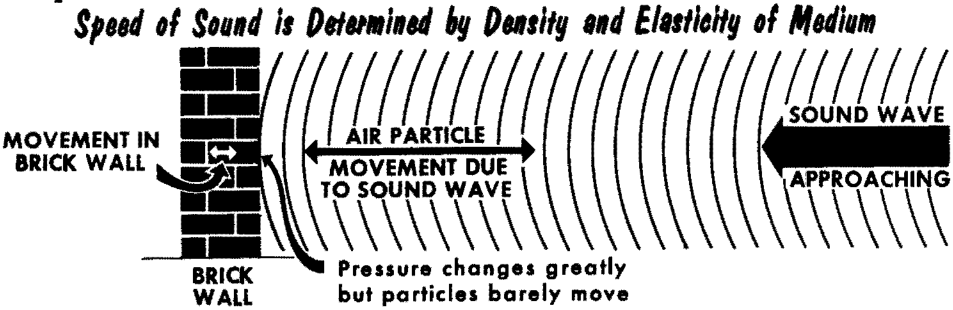
Most of the sounds that we are concerned with in audio exist in rooms. But to understand what happens to sound in rooms, remember that echo that can be heard in the mountains. Every sound made comes back a few seconds later, like a perfect mimic. A wall-like face of rock reflects the sound waves that are generated, and sends them back one or more times.



Because of the large open spaces in mountain areas and the time it takes sound to travel (about 4¾ seconds for each mile of travel), the reflected sound is heard so long after the original sound that it sounds quite separate from it. However, all surfaces reflect sound in the same way, even the walls in your living room. The difference is that the sound does not take so long in going to the reflecting surface and coming back, so the reflected sound does not get completely separated from the original sound.

You must have noticed at some time the difference in a room when all the furniture and rugs are removed (before moving into a new apartment or in preparation for the painters). Without the furniture and carpeting, the room sounds "hollow." When the furniture is in it, the room becomes pleasant to talk in. The hollow effect is due to the echo in the room from the wall surfaces. When the room is empty, the echo goes on bouncing from wall to wall a great many times; when the furniture is in, the echo is deadened.

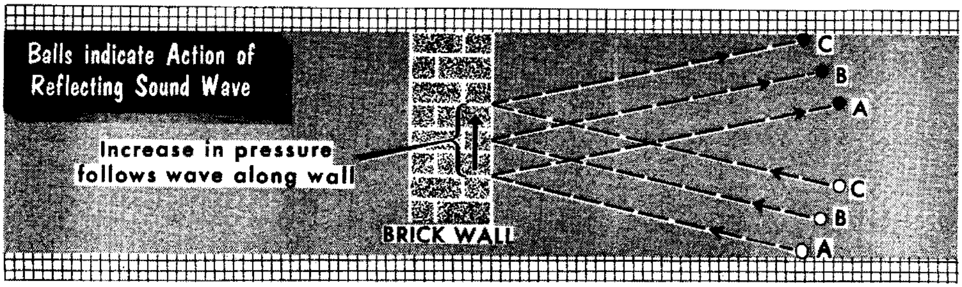
Why Sound Waves Reflect



Sound will travel through anything except a vacuum, but the speed at which it travels is set by the density and elasticity of the material through which it is traveling. When sound, traveling in air, comes to the end of the air, it will start to penetrate whatever it strikes. The wave consists of both pressure and movement of the particles of the material through which it goes. The power in the given area of a sound wave is found by multiplying the pressure by the velocity, or rate of particle movement. Transmission in air uses a combination of large movement with small pressure, compared to transmission in, say, a brick wall.

Since the brick wall is much heavier, or denser, than air, the sound wave will not move the brick as much as it does the air. The pressure transmitted to the bricks will be the same as that built up in the air where it strikes them, but because the movement in the brick is very small, something different happens in the air where it touches the wall—it does not follow the same pressure and movement combination as air elsewhere, because the air hardly moves at all. This means that the pressure of the sound wave is almost doubled at this point. This "surplus" pressure starts another sound wave, directed away from the wall.

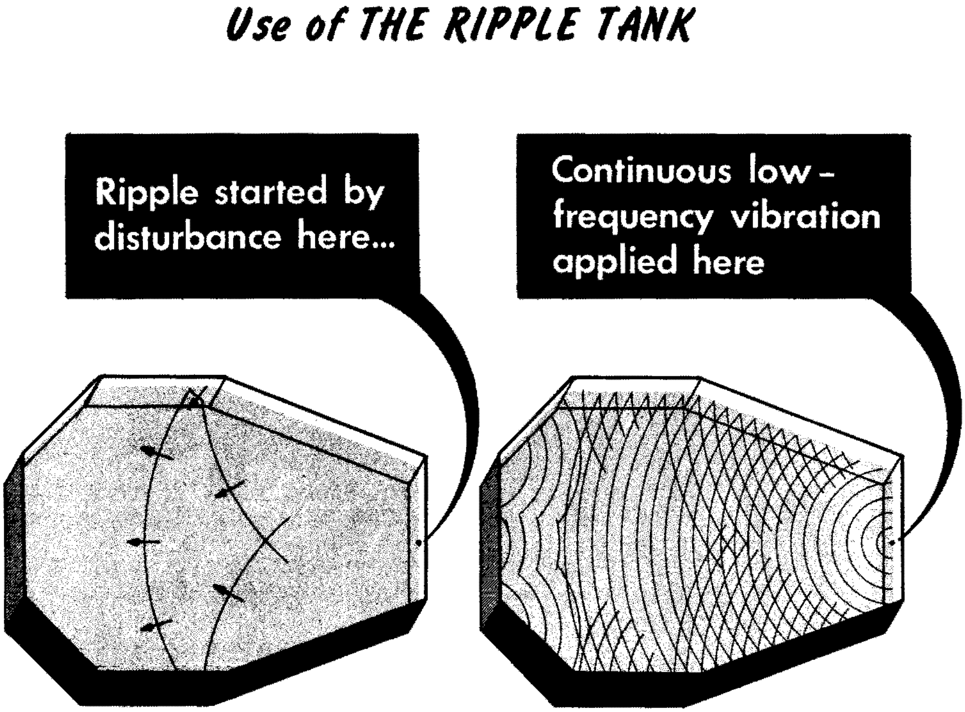
If the original wave strikes the wall at an angle, instead of "head on," the increase in pressure will follow the wave along the wall, as different parts of the original wavefront reach the wall. This results in the wave leaving the wall, just the same way that light gets reflected from a mirror.



The Ripple Tank

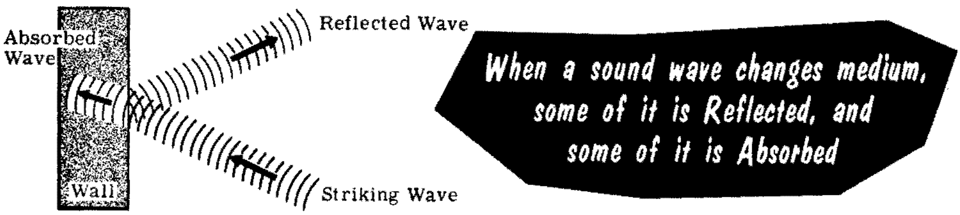
Waves of water in a tank are reflected in the same way. You can watch this at high tide, where the incoming waves of the ocean strike a breakwater at an angle. This gives a kind of slow-motion picture of what happens to sound waves.

Acoustic architects use ripple tanks for examination of the effects of various shapes in building structure on the way waves get reflected round an auditorium. It proves a useful way of working out a good shape for a building without having to make a full-scale model and then try all over again if the acoustics are improper.



When a tone or sound is continued, reflection effects build up. A single clap, or other pulse of sound, is reflected in a sequence that can easily be traced in a ripple tank. To some extent, we can consciously discriminate between direct and reflected sounds in normal listening. But a continuous tone causes a sound *pattern* to be set up, called a *standing wave pattern*. This is similar to the standing wave set up inside an organ pipe, except that the one in the organ pipe is deliberately controlled and has a pattern in only one direction —along the pipe. Patterns in rooms or buildings are not so organized, and "stand" in several (at least two as a rule) directions across the room.

Absorption



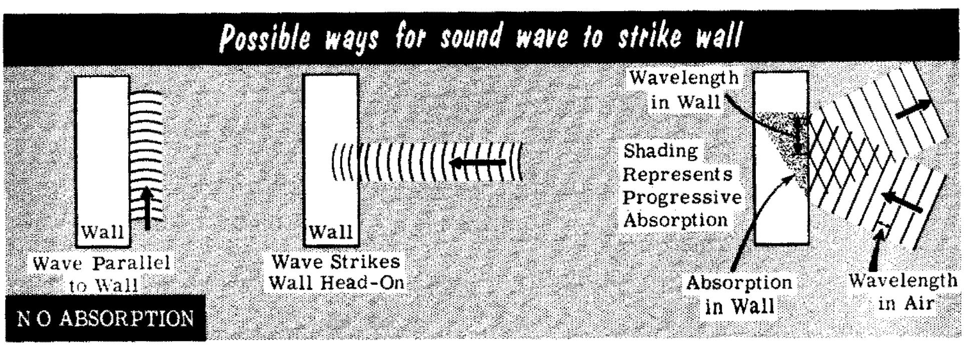
Sound is never completely reflected. Some of it is also absorbed. If a sound wave hits the wall of a room, most of it will be reflected back into the room. Some of it, however, will go on into the wall. Some of it may even go out into the next room. This is the way sounds in one room can be heard in the next—through the wall.

Each time a wave encounters a change in medium, this happens. Some of it goes into the next medium; some of it gets reflected. The proportions depend on the differences between the two media (in the example, the air of the room and the substance of the wall), the wavelength of the sound waves, and the angle at which they strike.

If a sound wave is traveling parallel to a wall, the wavelength along the wall will be the same as in the wave traveling in air. Very little absorption will occur because the wavelength of the same frequency in the wall is much longer (since sound travels faster in the material of the wall).

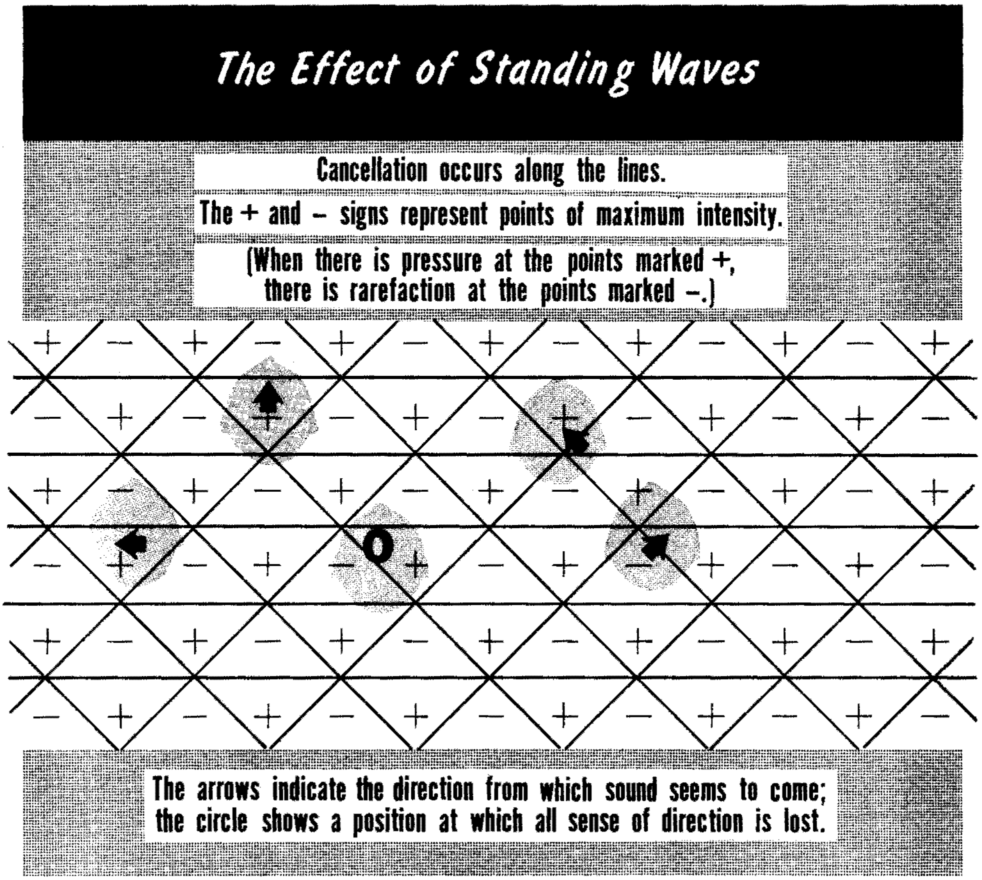
If the wave hits the wall "head on," the reflection and absorption will divide according to the density and elasticity of the wall material compared to air. But if the wave hits at a particular angle, the wavelength along the wall due to the striking wave may be the same as the natural wavelength for this wave in the wall.

Because this tends to make the wall take up more of the sound wave, there is a critical angle at which a sound wave will strike a surface, at which it will absorb much more than either a head-on strike or traveling parallel.



Standing Waves

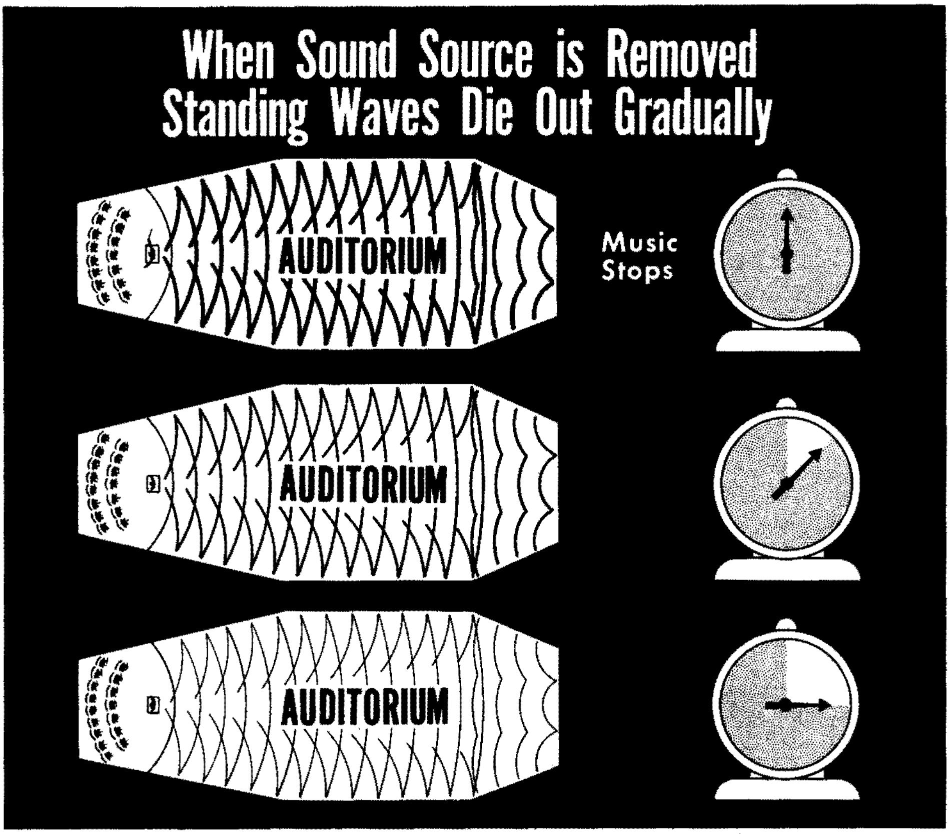
As a result of this standing wave pattern, the normally definite sense of direction that enables us to tell where sounds come from is lost. This can be verified with an oscillator, amplifier, and loudspeaker, or by just getting an instrumentalist to play a single long note. The sound seems to fill the room. If you have an impression that the sound comes from one direction (with your eyes shut) you are probably wrong, and if you move your head slightly, it will seem to come from a different place.



If your head is placed where a maximum intensity occurs by your right ear, and the left ear is at a point of less intensity, the sound will seem to come from your right side. If both ears are on a line of maximum intensity, or are at equal intensity, the sound will seem to be in front of you. And if both ears receive pressure at the same intensity, but one receives a pressure wave when the other receives a rarefaction wave, and vice versa, the impression of direction is confused.

Standing Waves (contd.)

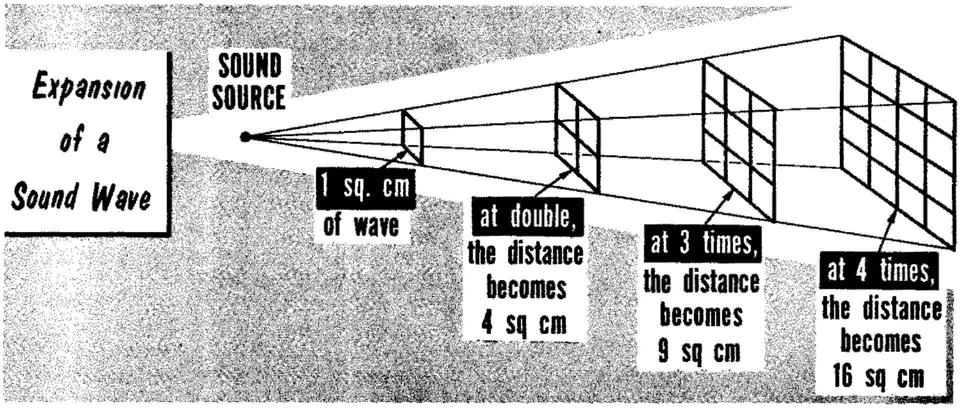
Standing wave patterns take time to build up and, what is more important in acoustics, also take time to die away *(decay)*. When the vibration causing the standing waves in the water ceases, or when the tone in the room stops, the vibrations all over the area die out gradually, rather in the way that a swinging pendulum comes to rest when nothing continues to drive it.



This dying away occurs in a fraction of a second in an ordinary living room. Most of us are so used to the brief presence of these patterns that we do not normally notice it. In a large hall or auditorium, however, or in a stadium with walls all around it, the time required for sound to die away can be quite noticeable.

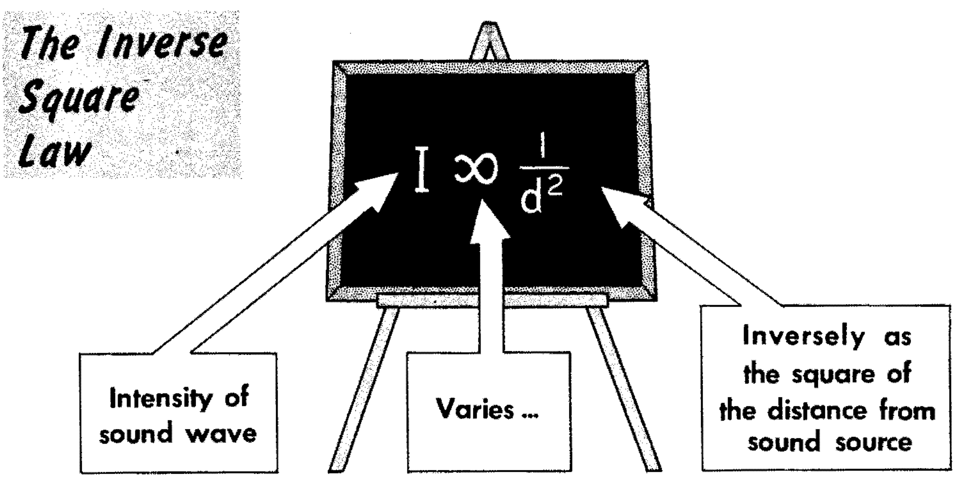
The time required for sound to die away after the originating tone ceases is called the *reverberation time* of the room or building. (Reverberation is the name given to "echo" when it is not sufficiently separated from the original sound to be noticed as a separate repetition.)

The Inverse-Square Law

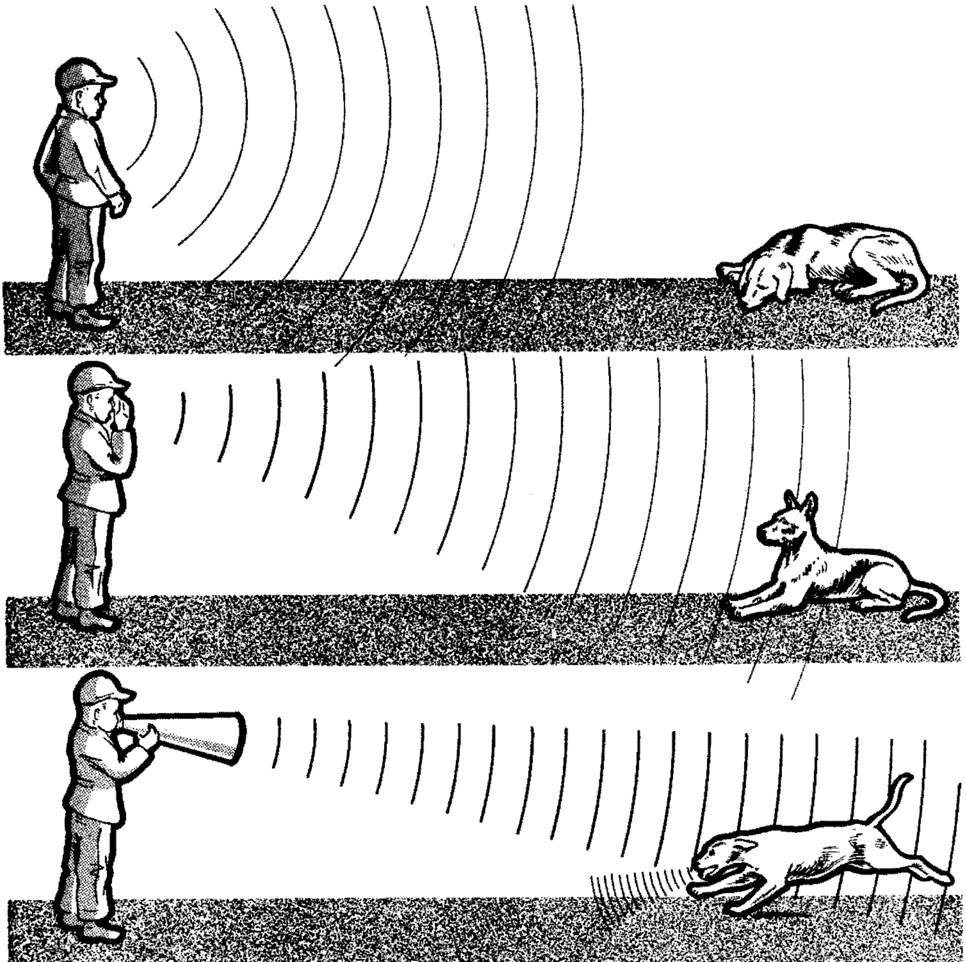


In the absence of any reverberation, the sound goes on outwards in an ever-expanding wave. The farther the wave goes from its starting point the larger its area becomes. The energy in the wave does not increase, because the wave can only pass on the original amount of energy put into it. This means the intensity in a *square centimeter* of the wave (which is how intensity is measured) must decrease as we go farther from the source. It is like spreading a fixed amount of butter on two slices of bread. If one slice is larger than the other, the butter on it will be thinner.

Doubling the distance means the area is quadrupled, so the intensity must be divided by four. Multiplying the distance by any number means that the area is increased by the square of that number, which, in turn, means that the intensity must be divided by the square of the number. This fact is known as the *inverse-square law.*



The Inverse-Square Law (contd.)



Concentrating Sound into narrower angle permits sound to carry further

The inverse-square law gives one reason why sound does not carry very far in the open air, unless there is something to make an echo. Cupping your hands or using a megaphone concentrates more of the original sound within a narrower angle, so that the power does not get scattered quite so widely. For this reason, the sound carries further in a particular direction. Because of reflection and reverberation effects, sound does not get "lost" so readily indoors.

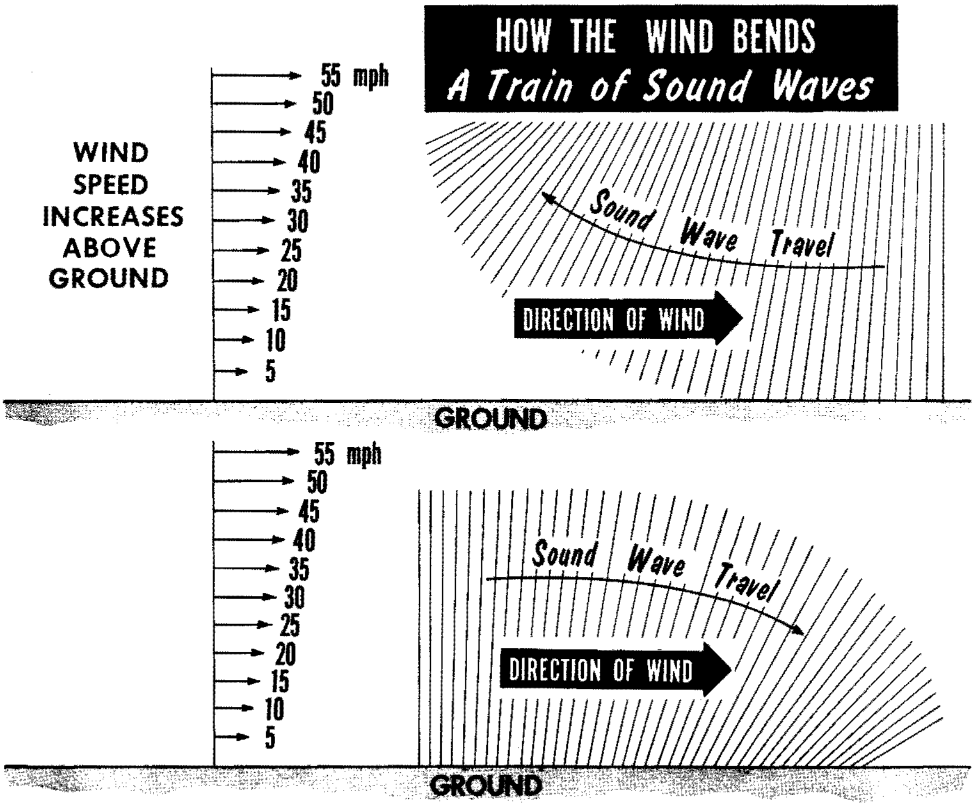
Masking

Although there is a big power difference from the quietest audible sound to the loudest that the ear can negotiate—about 1,000,000,000,000 times—we cannot listen effectively to both a very soft and a very loud sound at the same time. If sound A is too much louder than sound B, A drowns B out. The scientific name for this effect is *masking*.

While it is sometimes a matter of getting the sound we *want* to hear loud enough to be audible against a noisy background, at other times—when we want to listen to a high fidelity reproducer, for instance—our object is to get the sound loud enough so that we do *not* hear other, *unwanted* sounds. Or, more specifically, we wish to get the unwanted sounds so quiet that they cannot be heard while we listen to the music.

Outdoor Acoustics

Have you ever wondered why sound travels much better "downwind" than against the wind? With sound traveling at about 750 miles an hour, a wind of only a few miles an hour will obviously not be able to *stop* the sound, although the speed of sound relative to the ground does vary with the wind speed.



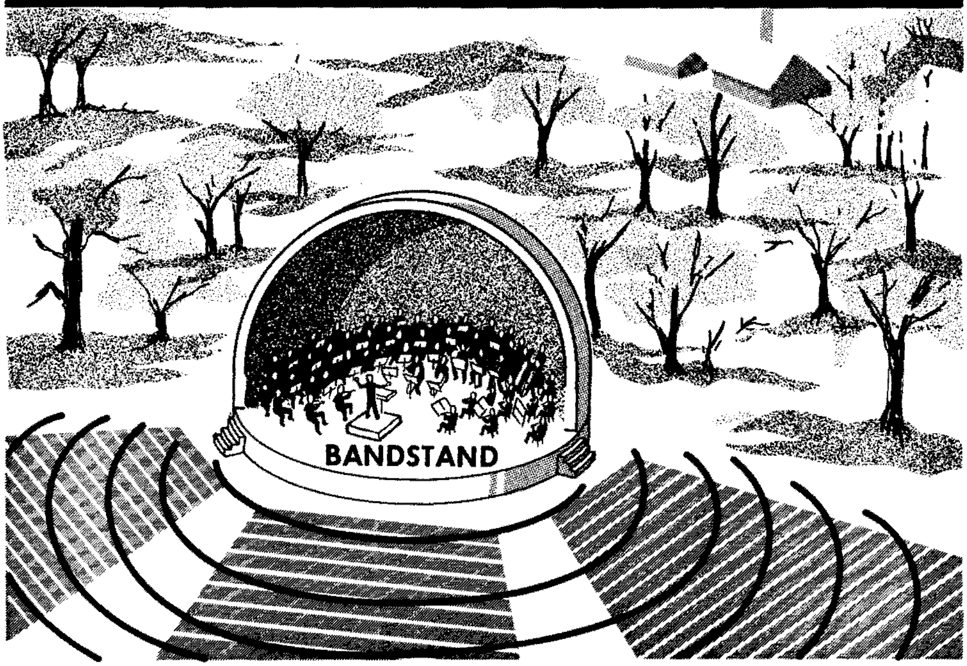
The wind travels at different speeds at various heights above ground. Air in contact with the ground hardly moves at all, but the higher you go, the stronger the wind. For this reason, the higher part of the sound wave traveling *with* the wind moves faster than the part near the ground. (The wind speed is added to the normal speed of sound.) This makes the wave lean forward, and bear down on the ground, and the sound heard here is more intense than without the wind.

Traveling *against* the wind, sound at higher levels moves a little slower than on the ground, the wave leans back. It therefore "takes off" and does not carry to a listener on the ground.

Listening Acoustics

It might be thought that to enjoy listening best, we should go outdoors where there is no reverberation. But in most outdoor locations, there is so much other noise going on, and the sound we want to listen to gets lost so quickly in all directions that we will not be satisfied. However, some quiet country locations are ideal for listening to music, which accounts for the increase in popularity of summer outdoor music festivals.

Quiet Country Locations are ideal for Music Listening

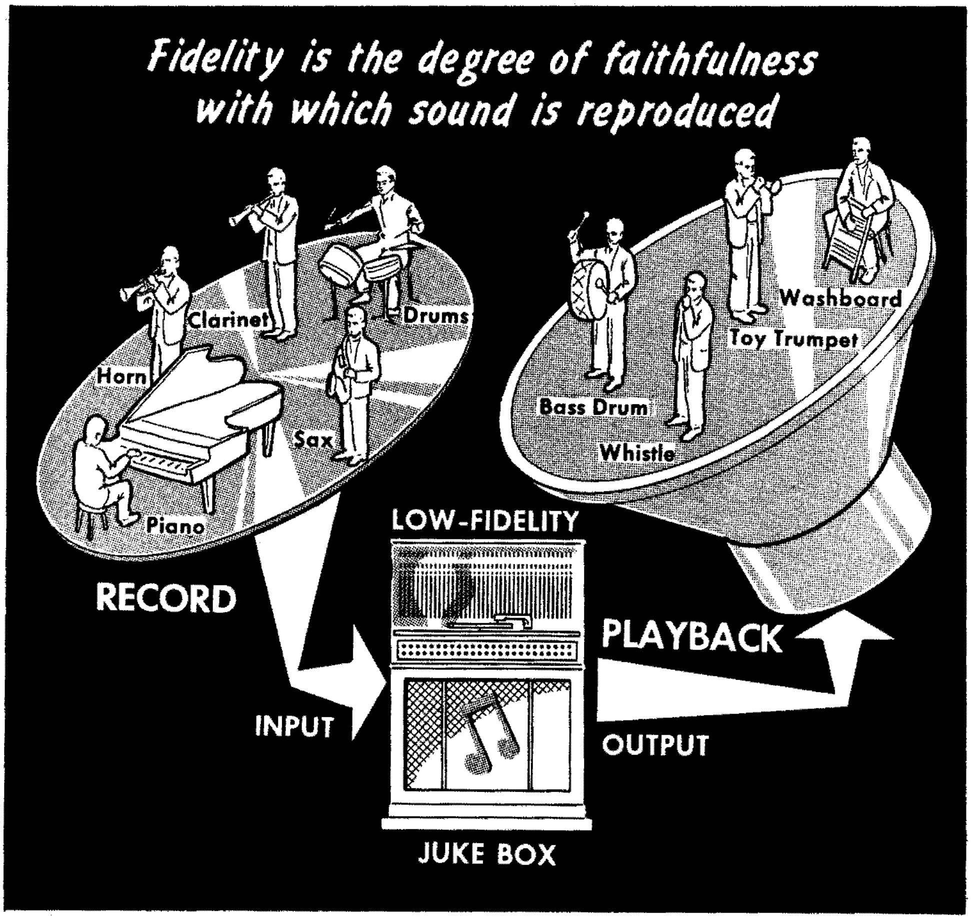


Using a room to listen in, whether it be our living room or an auditorium, helps to keep in the sound we want to hear, and to keep out what we don't want to hear. Reverberation helps to build up the intensity of the sound we want to hear, but there is a limit to the useful amount of reverberation. If we have too much, the sound seems to go on and on and on. Each new sound is blurred by the reverberation of the sounds that preceded it.

This phenomenon is closely linked with masking: reverberation should build up the sound, but not so much that it masks the original sound or succeeding ones. Every size of room or auditorium has an optimum amount of reverberation and length of reverberation time for music or speech that will give the most enjoyment from listening.

Sound Composition

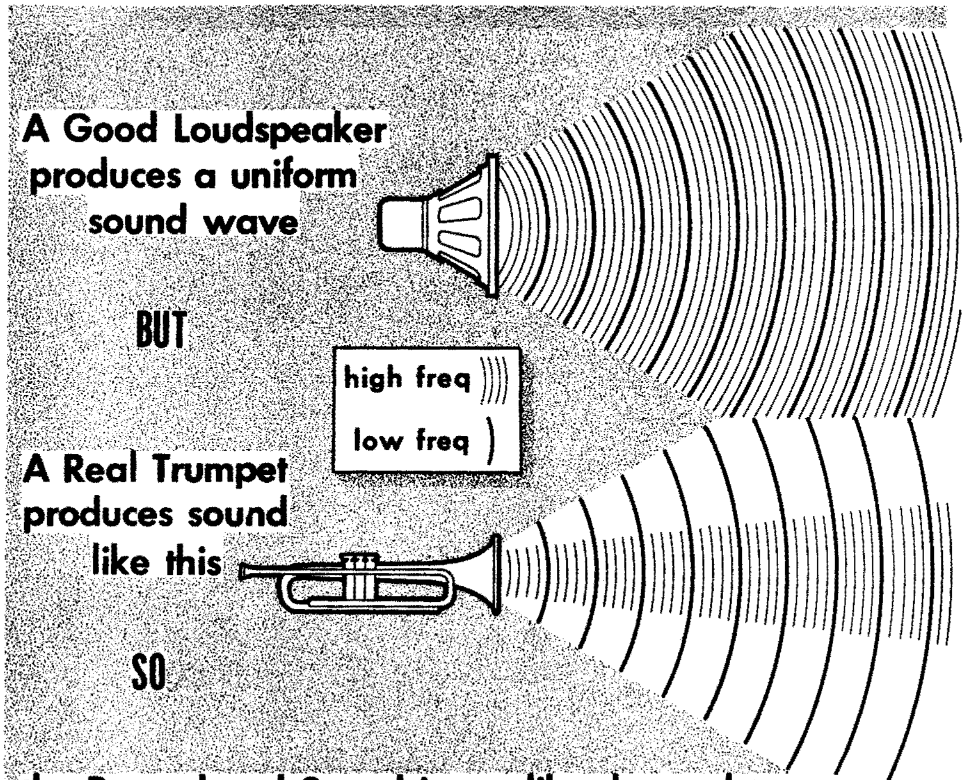
Every sound has its own individual composition: different frequencies at varying intensities, and recurring at intervals. Musical instruments have a definite pitch range, from the lowest note to the highest note that each of them can play. Above each note there is a range of overtones which must also be heard if we are to distinguish one instrument from others playing the same note.



The faithfulness with which the sound reproduced through a loudspeaker copies the original is called the *fidelity* of the system. If all the original frequencies are present in their original proportions, the system has "high fidelity." In the average juke box—especially the older ones—the fidelity is very low. The tune, of course, can be recognized, but you would have difficulty in recognizing some of the instruments playing it.

Directional Effects

Hearing all sounds in correct proportion is not only a matter of making sure that they come out of the loudspeaker that way. Different frequencies do not travel in rooms in quite the same way, so we may notice some peculiar effects because all the frequencies do not reach us properly.

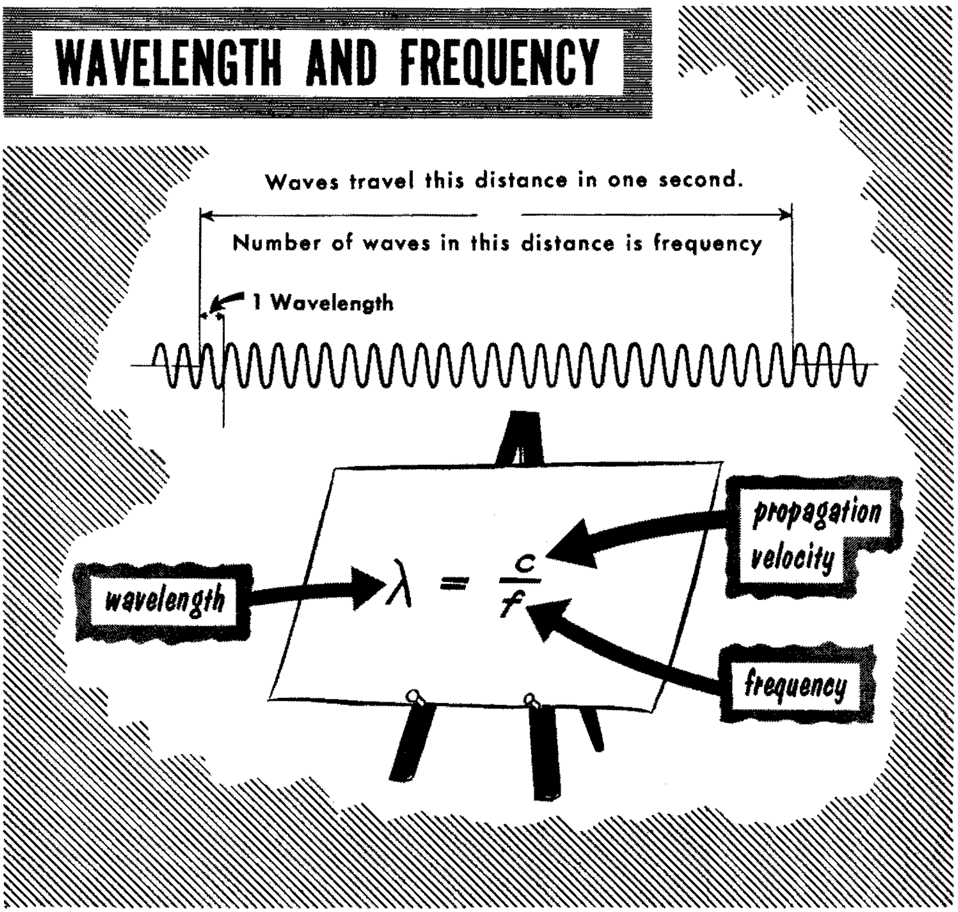


For example, a trumpet gives out sound that is very rich in overtones. The lower (fundamental) notes of the trumpet go out in all directions, whereas the high-frequency sound is almost squirted in the direction in which the trumpet is pointed. This kind of sound production and the way in which the sound bounces around before reaching our ears have become part of our experience in listening to trumpets.

If a loudspeaker radiates all the correct frequencies to make up a trumpet sound, but distributes the high frequencies in all directions (in the same way as it does the lower frequencies), reproduction of the trumpet will not be realistic.

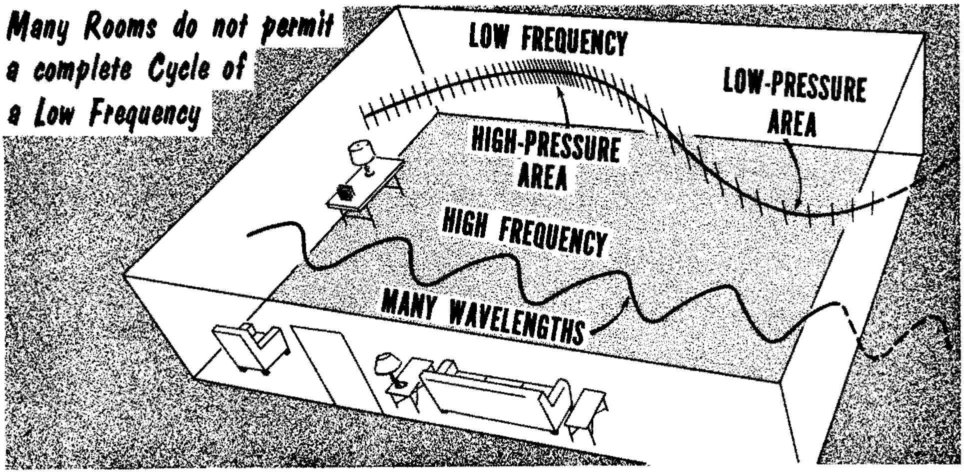
Frequency and Wavelength

All sounds travel at the same speed, regardless of frequency. Because the waves are traveling, the length of a wave (measured along the direction it travels) from the peak of one wave to the peak of the next will be different, according to the frequency.



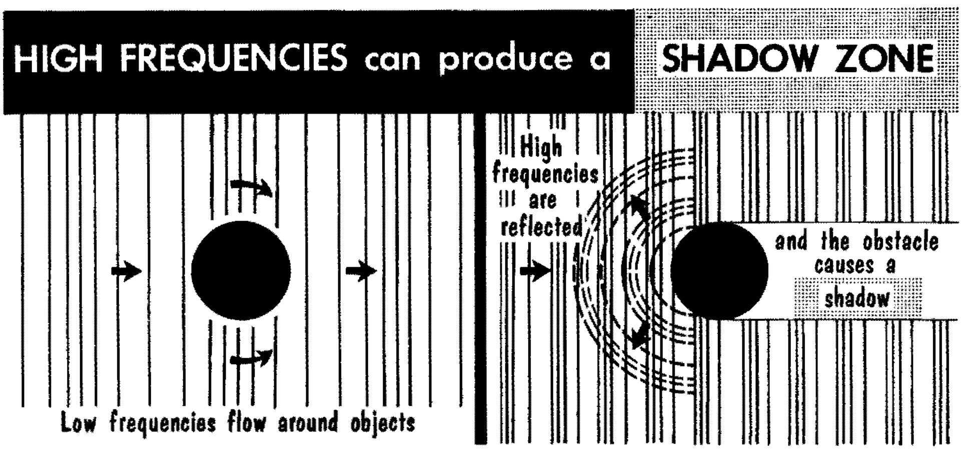
Sound moves at about 1100 feet per second. If the frequency of a sound is 440 cycles per second (A above middle C), there must be 440 cycles in a 1100-foot piece of the sound wave. Each wave is thus about 2½ feet long in this case. If the frequency is 8800 cycles (an octave beyond the end of a piano keyboard) the same space (1100 feet) will be occupied by 8800 cycles, or waves, and each must be only one eighth of a foot (about 1½ inches) long. Low frequencies from organ pipes may be down in the region of 32 cycles, with a wavelength of 34 feet.

Long and Short Sound Waves



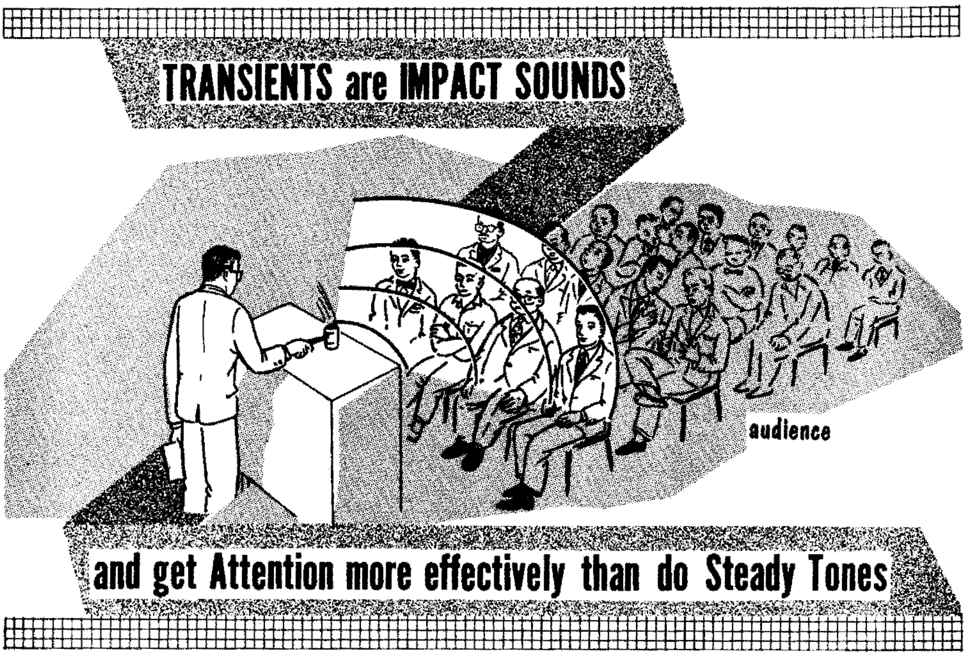
As a result of the tremendous differences in wavelength, the low frequencies may not even have one complete wave in an average-sized room at the same instant, whereas a high-frequency sound will have many complete waves in the room traveling in a number of directions at the same time.

Because of its size, a low-frequency wave can hardly be recognized as such in a comparatively small room—it is more like a fluctuating pressure throughout the whole room at the same time. For this reason, it will fill the room, regardless of the room's shape. The higher frequencies, however, since they are smaller than most objects, such as walls and furniture, will be reflected whenever they strike a surface. For this reason, you may miss some of the high frequencies, not because they are not present, but because you happen to be sitting in a *shadow* zone.



Transients

The word "transient" means something that is passing or changing. Within the strict meaning of the word, any change, from silence to sound, for example, or from sound back to silence, would be classified as a transient, while the steady, unchanging tone is not a transient.

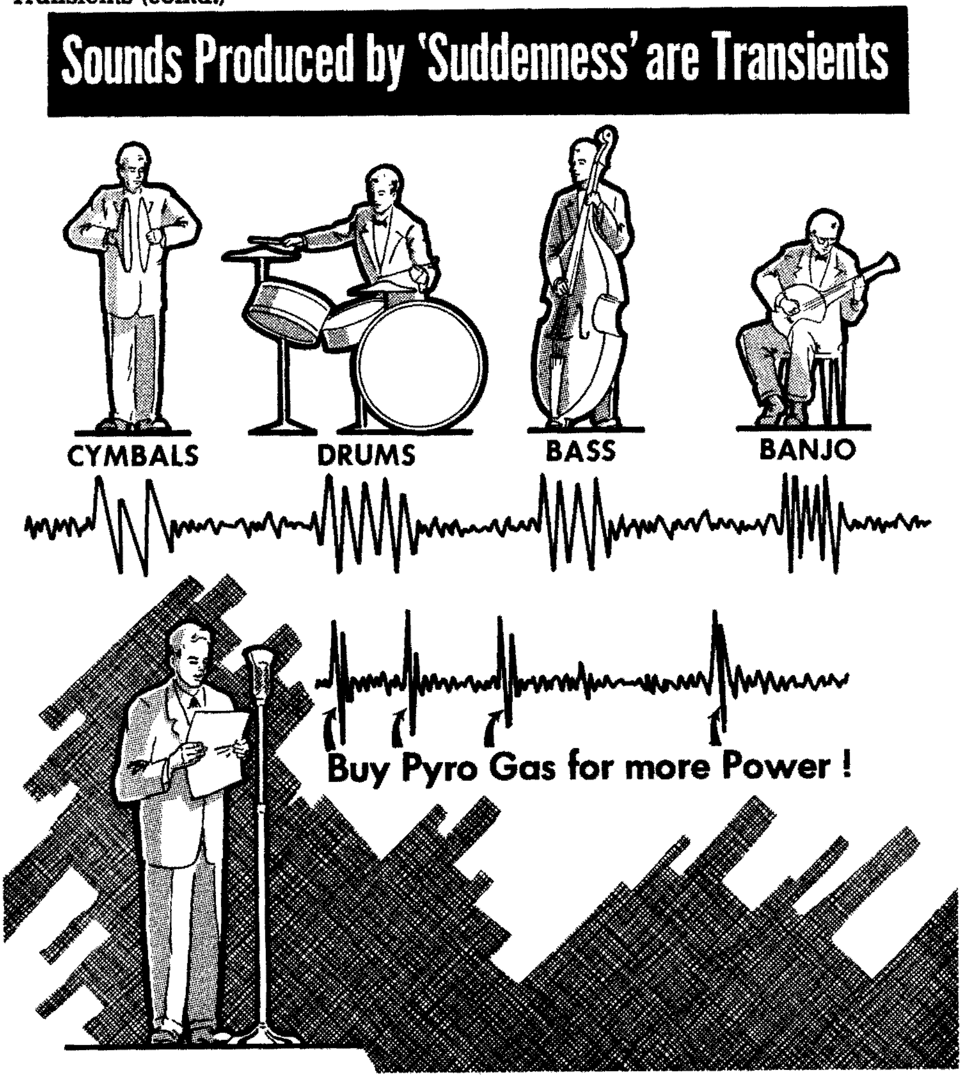


In audio, for two reasons, the word is used with a somewhat more restricted meaning. The first is connected with everyday listening experience, the second with the performance of audio equipment—microphones, amplifiers, loudspeakers, etc.

The start of any sound, whether sudden or gradual, will reach a listener before the reinforcement of the same sound by reverberation. When the sound finishes, the reverberation goes on. Because this happens all the time around us, our hearing faculty has formed the subconscious habit of paying more critical attention to the beginning of sounds than to the endings.

If the sound builds up relatively slowly, as in the deep notes of a pipe organ, the reverberation builds up almost as rapidly as the direct sound from the pipe. On the other hand, a sound that starts suddenly, like a hand clap, a drum beat, or any sound that has what we may call "impact," reaches the listener well ahead of its reverberation, and gives him a good chance to tell where it came from. Thus musical sounds that have impact, like a hand clap, can be regarded as attention-getting sounds.

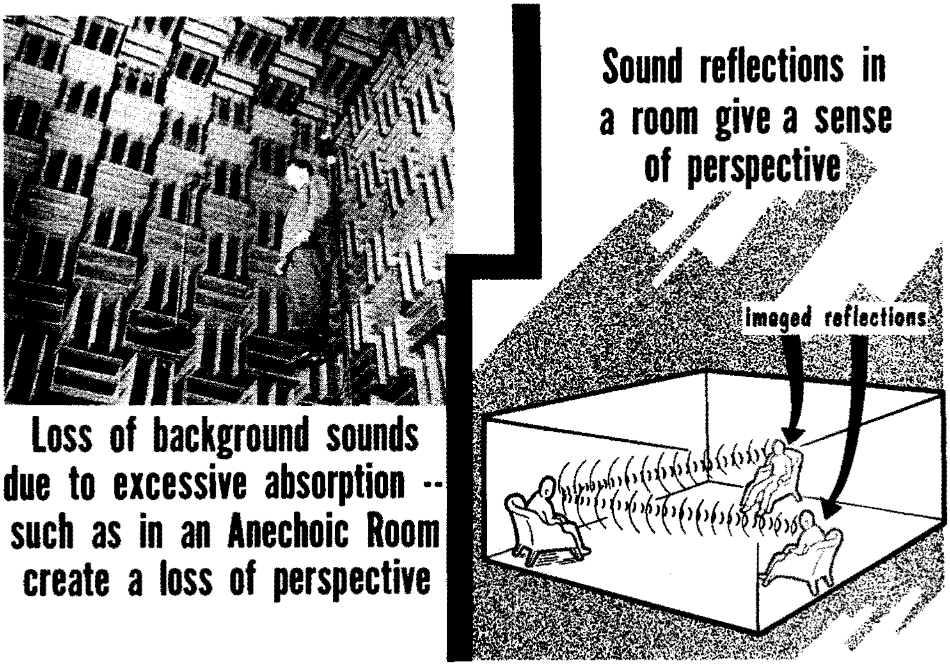
Transients (contd.)



Any sound distinguished by suddenness is thus a transient: all of the percussion instruments, drums, cymbal, etc., as well as plucked or struck strings; and in speech, the sounds made in pronouncing the letters, b, d, g, k, p, t, are always transients. (Other consonents sometimes are too.)

The second reason for paying attention to this more restricted kind of transient is that audio equipment has particular problems in handling these more "sudden" types of sound, as compared with steady tones of unchanging, or relatively slowly changing, frequency or intensity.

Background



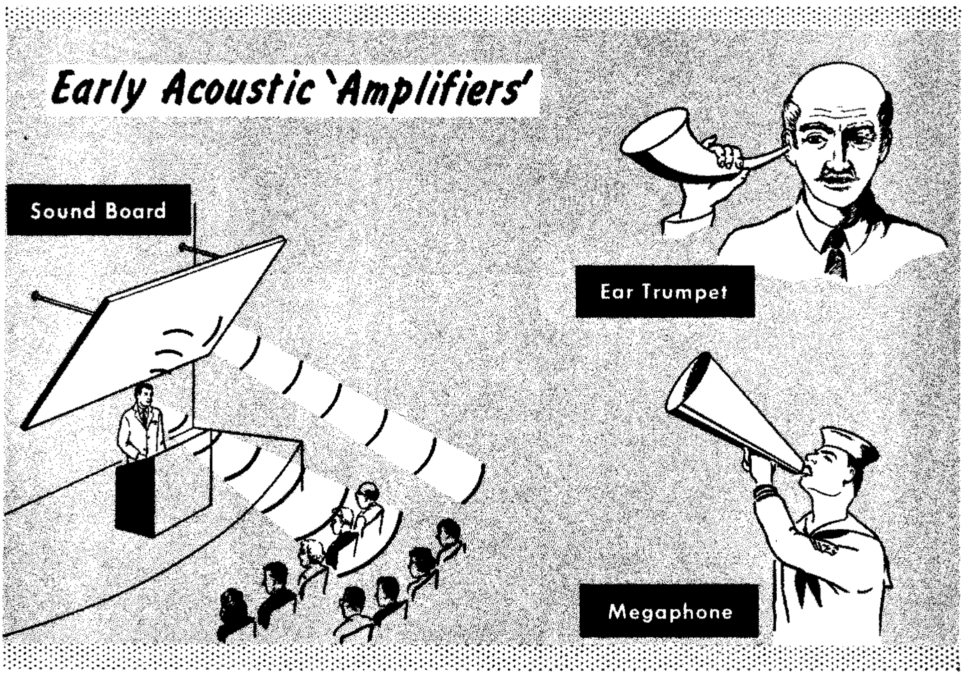
Although reverberation makes the endings of sounds less important than their beginnings, this does not mean we can ignore the endings. Human hearing is very conscious of the reverberation even though it may not listen to it so critically.

You notice nothing unnatural about talking in an open field. There is no echo to your voice because the sound of it can keep on going without being reflected. You also notice nothing unnatural about talking in a room where there is a very definite echo, or reverberation, to your voice. But try talking in an anechoic room (a room used for acoustic testing in which walls, floors, and ceiling are made completely absorbant of sound); try one of those padded rooms they use for violent cases in mental institutions (if you ever have the opportunity). Either of these places will give you a quite unnatural sensation of "soundlessness," rather than of silence.

When you are outdoors, it may be quiet but there are little sounds going on that give you a subconscious perspective of where you are: birds singing in summertime or other incidental sounds that are usually "in the background" outdoors. When you are indoors the reverberation of your own voice subconsciously tells you what kind of room you are in. But in the padded cell or anechoic room, because of the excessive absorption that removes all background sound, either from outside or from your own voice, you feel your voice is "lost." There is no background to give you perspective.

Amplification

Normal sound vibrations are very small—so small that the movement of a loudspeaker diaphragm is not visible, except at the lower frequencies. The only reason that sound can be transmitted so efficiently with such small movements is that it uses the natural transmission speed of the air. Nonetheless, the magnitude of movement of the particles diminishes with distance traveled (recall the inverse-square law). Although hearing covers a very wide range of intensity variation, sound will only carry a certain distance before it becomes inaudible.



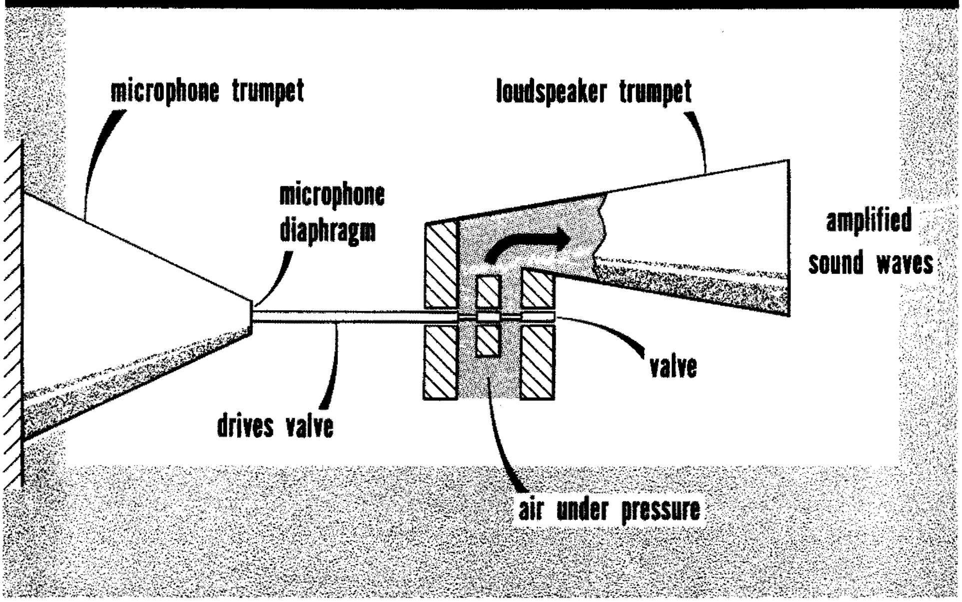
From earliest times man has had this problem to overcome. Long before electronic amplifiers became possible, some kinds of acoustic "amplifiers" were used. Actually these were not amplifiers in the true sense of the word—they did not increase the sound power but merely conserved what power was available. The ear trumpet, for example, collects a larger area of the sound wave and thus increases the intensity at the earpiece. The megaphone concentrates sound at the sending end, to restrict it within a narrow angle. The sound in front of the megaphone is louder, but you can hear less than normal in all other directions. The speaking tube is more efficient; it virtually prevents any sound escaping at all, so that all the power is conveyed along the tube. In this way, sound can be transmitted for considerable distances. The reflector board over the speaker's platform serves a purpose similar to the megaphone by making use of sound that otherwise would escape upwards.

Amplification (contd.)

An early attempt at real amplification, (the *pneumatic amplifier)* was entirely acoustic. Sound vibrations striking a diaphragm were used to operate a valve, somewhat like a reed, through which air under pressure was driven. The amount of air passed by the reed was controlled, or "modulated" by the sound vibrations reaching the diaphragm, and produced a more intense replica of the original sound.

The amplification given was rather crude. The apparatus was far less convenient than modern electrical, or electronic amplification. The "microphone" and "loudspeaker" had to be mechanically coupled, and thus close together. It could not be used for as many purposes, and its quality was rather rough, to say the least.

PNEUMATIC AMPLIFICATION



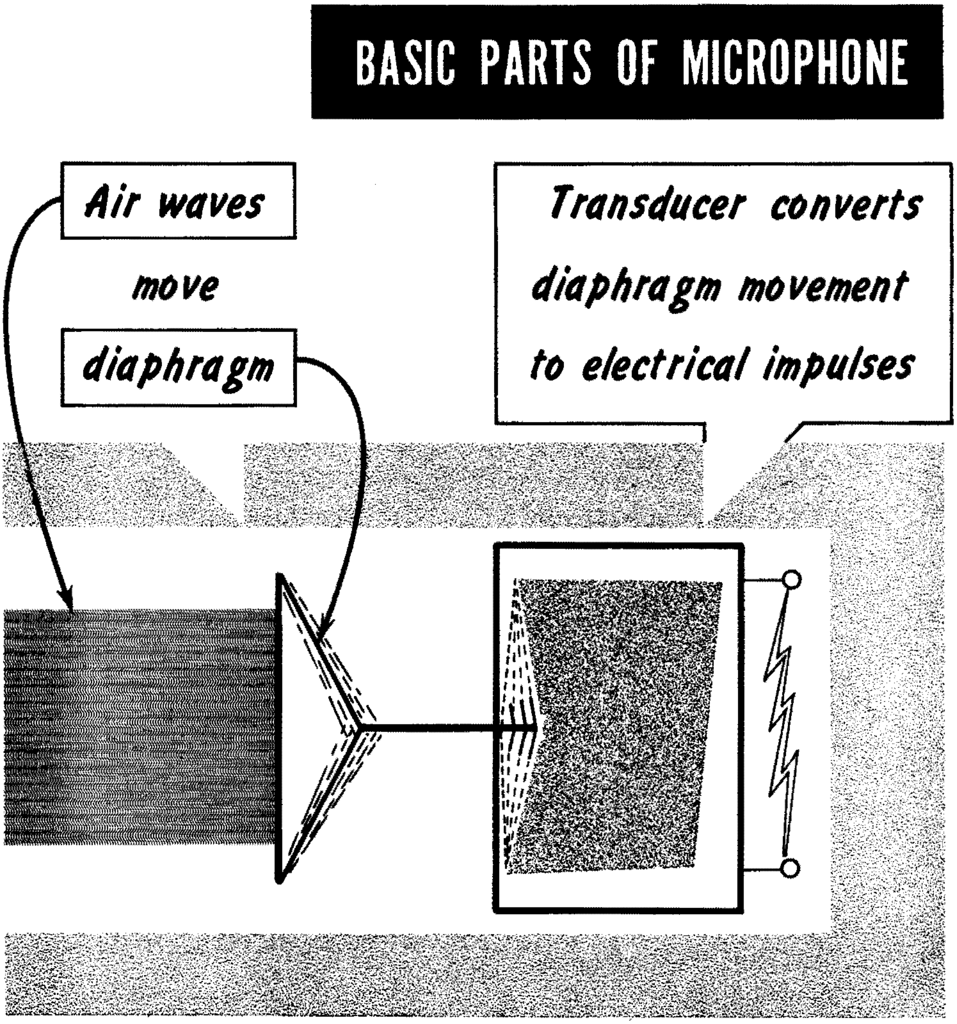
Nonetheless, this acoustic device had one thing in common with any real amplifier. Extra power had to come from somewhere. In that case it was air under pressure supplied from a suitable pump. In electrical amplification, extra electrical power is added and later converted to sound. Very small electrical impulses, or waves, put into an amplifier, control a larger amount of power taken from a battery, power line, or some suitable source and give a large amount of audio power to drive a loudspeaker or other transducer.

QUESTIONS AND PROBLEMS

1. What are the differences between echo and reverberation?
2. How does the standing wave pattern in a room or auditorium differ from that in an organ pipe?
3. Explain what causes sound waves to reflect when they encounter the surface of a different material.
4. When sound is reflected, is all of the wave reflected, or does some of it pass into the reflecting material?
5. What is the reverberation time of a building or auditorium?
6. What is the inverse-square law? Explain when it applies to sound waves.
7. What makes listening to anything outdoors different from listening to the same thing indoors?
8. Explain how the principle of masking applies (a) to trying to hear a conversation at a noisy airport, and (b) in getting the full benefit of listening to high fidelity.
9. Does wind stop sound waves? If not, why is it more difficult to hear any distance up-wind than down-wind?
10. Explain why you think a reproduction of violin playing through a loudspeaker with a metal horn would not sound very realistic.
11. What is the connection or relationship between frequency and wavelength in sound waves? What would be the wavelength corresponding to a note whose frequency is 250 cycles, if the speed of sound is 1100 feet per second?
12. Listening to a particular high fidelity setup, the following facts are noticed: (a) an organ note of 41 cycles is equally audible anywhere in the room; (b) a power-line hum (that proves to have a frequency of 180 cycles) seems quite strong in places and almost inaudible at others; (c) the sound of a snare drum played with a wire brush on a certain recording is audible in front of the loudspeaker, but not toward the sides of the speaker enclosure. Explain these differences.
13. What does the word transient mean, but to what is its meaning restricted (a) musically, (b) in audio?
14. Why do instruments or other sources of sound with "impact" transients (a) attract attention, and (b) give a good indication of their direction, which smooth-starting tones do not?
15. Why does a sound of particular intensity only seem to carry a certain distance? How can a megaphone help sound to carry further?
16. Do the following devices amplify sound: a megaphone; a speaking tube; the reflecting board over a speakers platform? If so, why? If not, why not?

The Purpose of Microphones

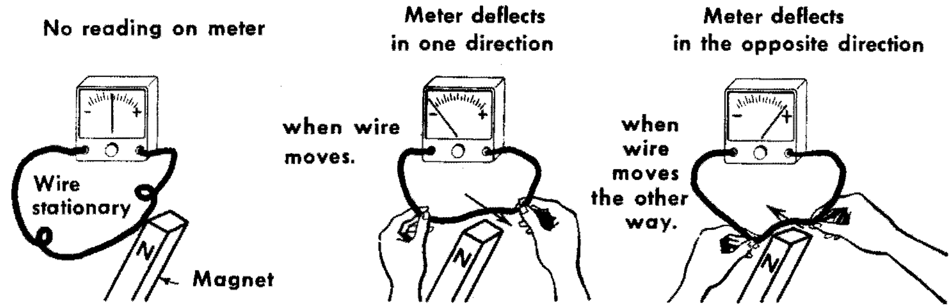
Sound waves have to be converted into electrical impulses to be amplified. A microphone is needed to convert the tiny acoustic vibrations into electrical waves.



Every microphone has two basic actions: first to convert the acoustic vibrations of the air into mechanical vibrations by having the air move a diaphragm—a light stiff surface; second, it must act as a *transducer*, and convert this movement of the diaphragm into electrical currents or voltages. One important difference between microphones is the arrangement used for making this conversion.

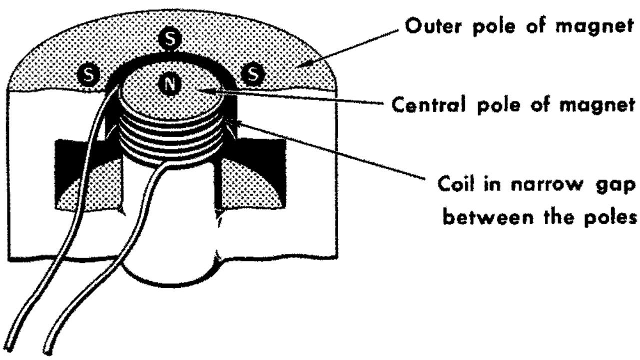
The Dynamic (Moving-Coil) Microphone

THE PRINCIPLE OF DYNAMIC-MICROPHONE OPERATION



If a wire connected to a meter that will indicate when current flows is moved about near to a magnet, the meter will show current fluctuations. When the wire is *moved,* the meter deflects. Holding the wire still produces no current. The direction of current indicated on the meter depends on the direction in which the wire is moved. This is an ideal basis for converting movement into electrical current. Because *movement* is the essential feature for conversion, a microphone using this principle is called a *dynamic* microphone. The problem in making a microphone of this type is that a large movement is needed to produce even a small current, while the movement of the air particles due to sound waves is small.

INCREASING THE EFFECTIVENESS OF THE MAGNET

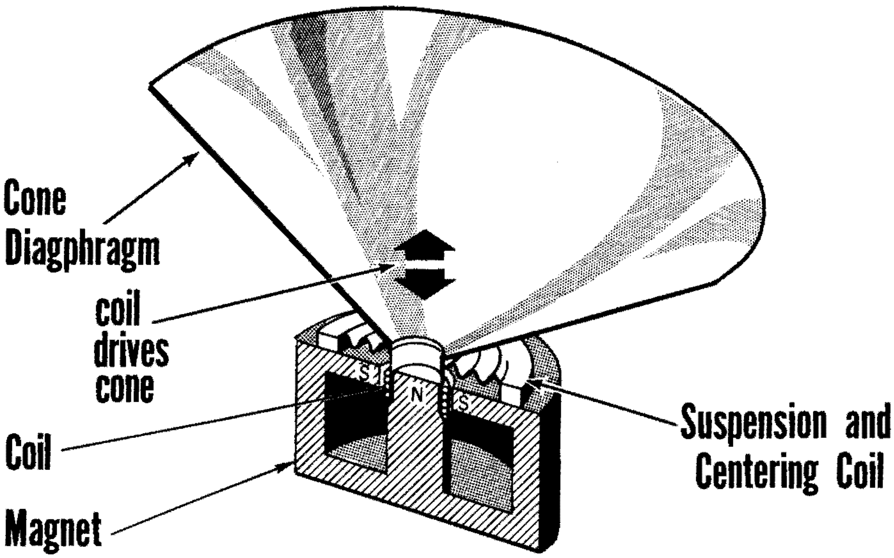


This problem is overcome, to some extent, by increasing the intensity of the magnetic field. A North and South pole are brought close together, and the wire moves in the narrow space between them. To increase its effectiveness, many turns of wire move in the same gap. It is convenient to make the gap circular, because this simplifies construction of the coil, gets the poles close together, and gives the coil free space in which to move.

The Dynamic (Moving-Coil) Microphone (contd.)

Converting movement into current is only part of the job. We must first move the coil by means of the sound waves, which requires a diaphragm. To move freely, the diaphragm must be light—as little heavier than air as possible. Because the coil is also attached to the diaphragm, it, also, must be as light as possible, or it would load the diaphragm down. Hence, a small coil must be used.

CROSS SECTION OF A SIMPLE MOVING-COIL MICROPHONE

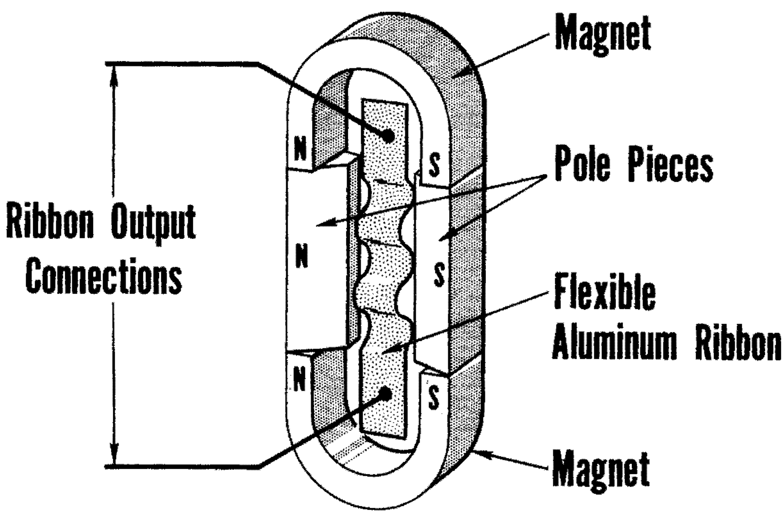


The use of a small coil requires a very intense magnetic field to get the best results. To accomplish this, the gap is made very small. To prevent the coil rubbing against the magnet poles, a centering "spider" or suspension is used, which allows free movement in the direction of vibration, while preventing the coil from moving against the pole faces.

The Velocity (Ribbon) Microphone

Another microphone that uses the same basic idea is the ribbon type. Instead of having a coil of wire, however, a single flat flexible ribbon of aluminum, or aluminum alloy, is used. It moves with the air vibrations, and the magnet poles on either side of it cause it to generate currents. The two basic microphone actions are thus served by the ribbon alone—it acts as both diaphragm and transducer.

BASIC CONSTRUCTION OF A RIBBON MICROPHONE

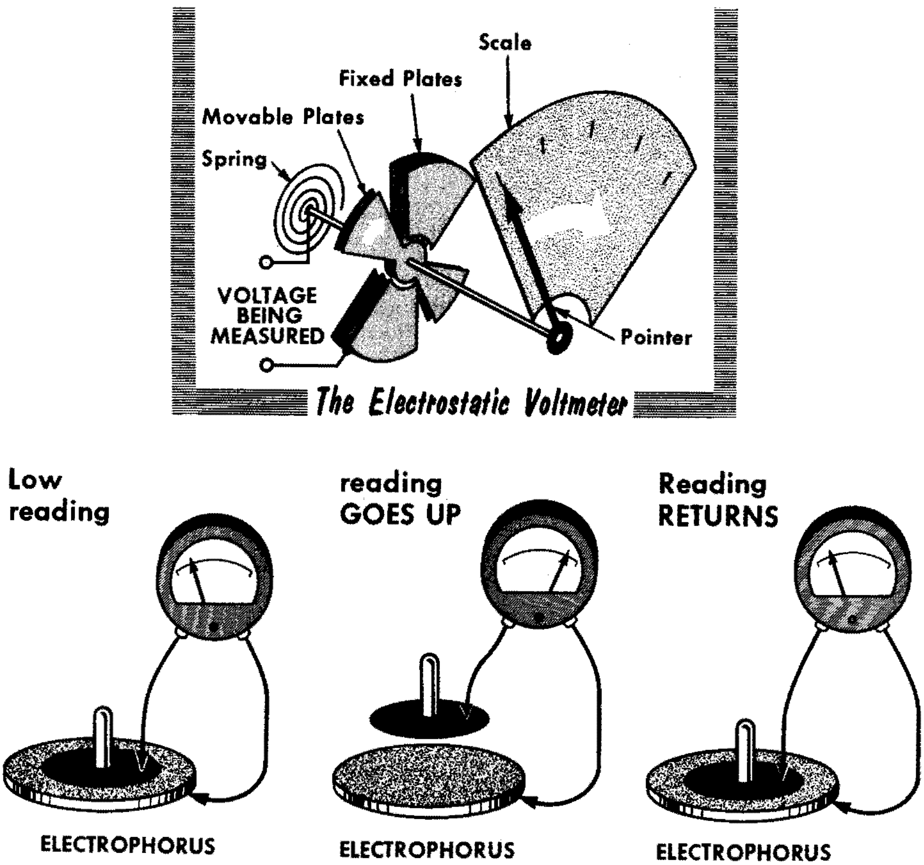


The ribbon microphone is also called the *velocity* type, because its response is proportional to the velocity of motion of the air particles in the sound wave rather than to pressure fluctuations. These microphones are also called *pressure-gradient* microphones because the movement of the ribbon is due to the pressure gradient—the difference in a pressure caused by the sound wave—between its back and its front. It is this constantly changing difference in pressure, of course, that controls the velocity at which the air particles move around the ribbon. Hence both "velocity" and "pressure gradient" are equally descriptive of the action of the ribbon microphone.

The Electrostatic (Condenser) Microphone

If an electrophorus (a simple instrument used in demonstrating the properties of electrical charges) is charged up and connected to an electrostatic voltmeter, a low reading will be obtained with the plates of the electrophorus in contact with each other. When the moveable plate of the electrophorus is lifted from its base, the reading will rise—probably shoot off the scale. When the plate is replaced, the reading will return to the earlier value.

The Principle of **CONDENSER MICROPHONE OPERATION**

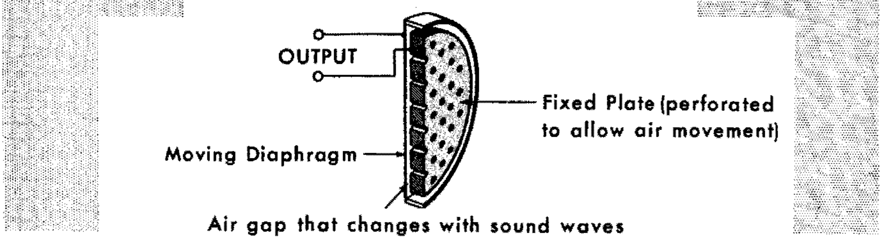


This shows that when the distance between the charged plates changes, the voltage due to the charge changes. This principle is used in the electrostatic or condenser microphone. (Nowadays, the use of the term "condenser" is discouraged for most purposes—it should be *capacitor*, but most people still use the older term for the electrostatic microphone.)

The Electrostatic (Condenser) Microphone (contd.)

In a condenser microphone, one plate is flexible, whereas the other has holes in it that permit air to flow into the space between them. (This permits the flexible plate to move freely due to sound waves from either direction.) The motion of the flexible diaphragm changes the spacing between it and the fixed plate, and produces voltage fluctuations.

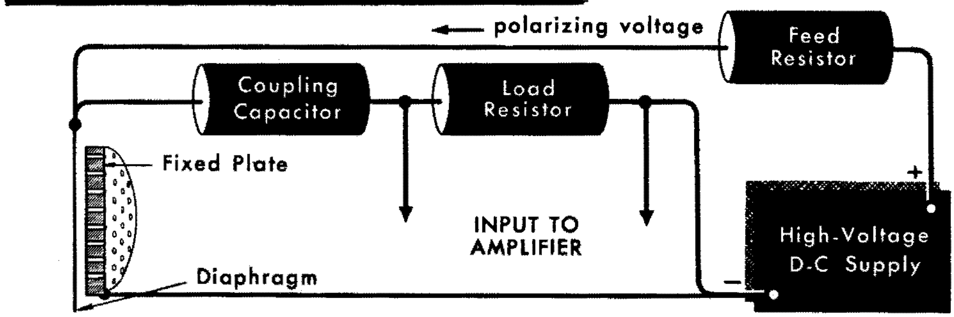
CUTAWAY VIEW OF A CONDENSER MICROPHONE



To make a condenser microphone work, it must have a steady electric charge upon it (obtained from a source of high voltage) that is isolated from the microphone so that a change in the spacing between the plates due to incident sound waves causes the voltage between them to go up and down. This is done by connecting a high voltage across the plates through a large resistance. The fluctuating voltage due to sound waves is fed through a coupling capacitor to an output resistor.

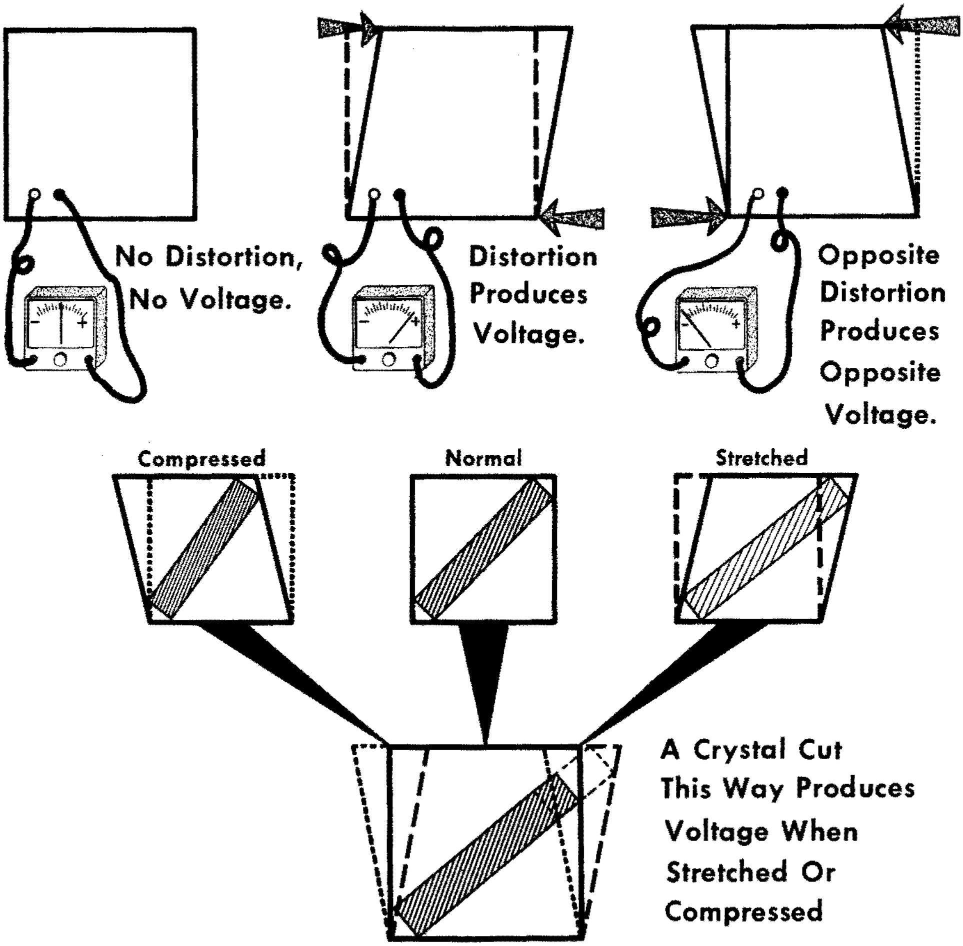
The capacitance of the coupling capacitor is greater than that between the microphone plates, and both the resistors are so large that the charges on capacitors do not have time to change during the slowest fluctuations due to sound waves. Because the charge on the coupling capacitor does not change, the voltage across it also must be constant. Therefore, the voltage that appears across the output resistor has the same fluctuations put out by the microphone, without, however, the polarizing voltage applied to microphone plates.

CONDENSER MICROPHONE CIRCUIT



The Crystal Microphone

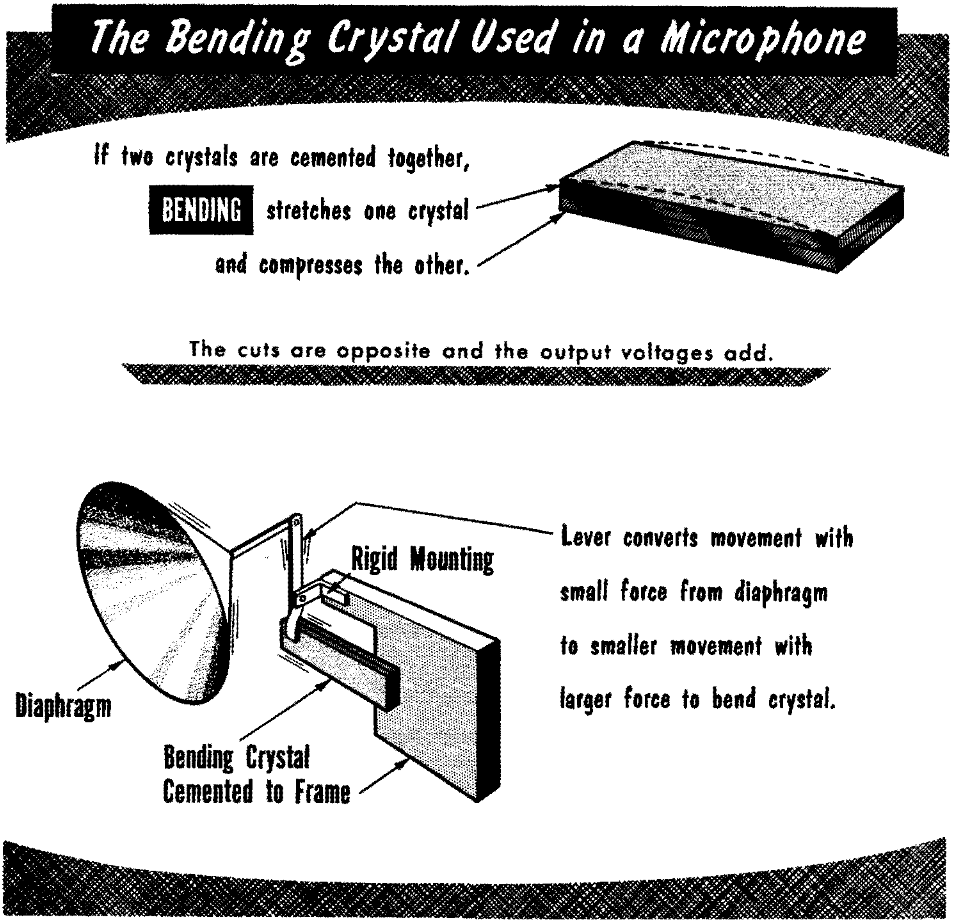
HOW A PIEZOELECTRIC CRYSTAL WORKS



A *piezoelectric* crystal is another kind of transducer. When such a crystal is put under mechanical strain, a voltage is set up inside the crystal structure. (Similarly, when a voltage is applied to the crystal, mechanical bending occurs—this makes the crystal useful for loudspeaker use.) The best or most efficient method of getting a voltage from a crystal is by changing its shape. By cutting pieces of it in a particular manner, the crystal can be made to produce its best voltage by stretching or compressing.

The Crystal Microphone (contd.)

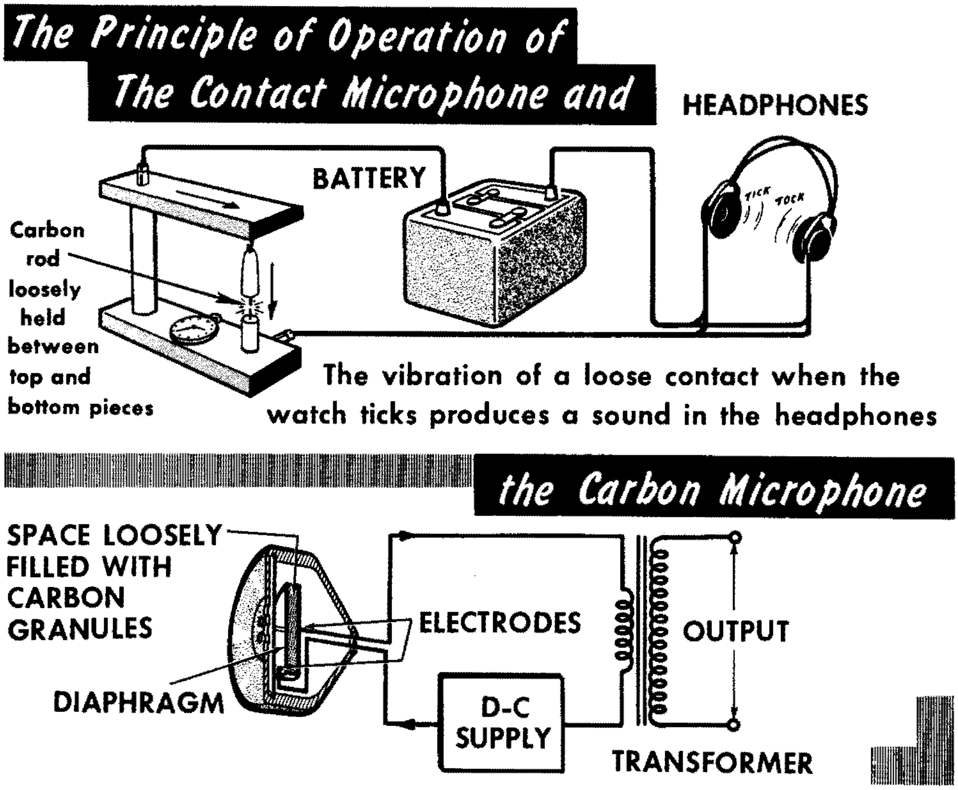
Even this type of cutting does not make an efficient transducer, because the amount of force needed to produce any voltage is great and the force imparted by sound waves is small. The action can be improved by cementing two crystals together, so that when one is compressed and the other is stretched, the voltage between surfaces adds up in the same direction. Bending this combination gives increased output for a smaller applied force.



This double crystal is still much too stiff for use as a microphone. For this reason, the diaphragm is coupled to the crystal by a lever action, which transforms a larger movement with smaller force to a smaller movement with a larger force at the crystal. In this way, the tiny movements of air particles in contact with the diaphragm are efficiently coupled to the crystal element to produce as sensitive a microphone as any other.

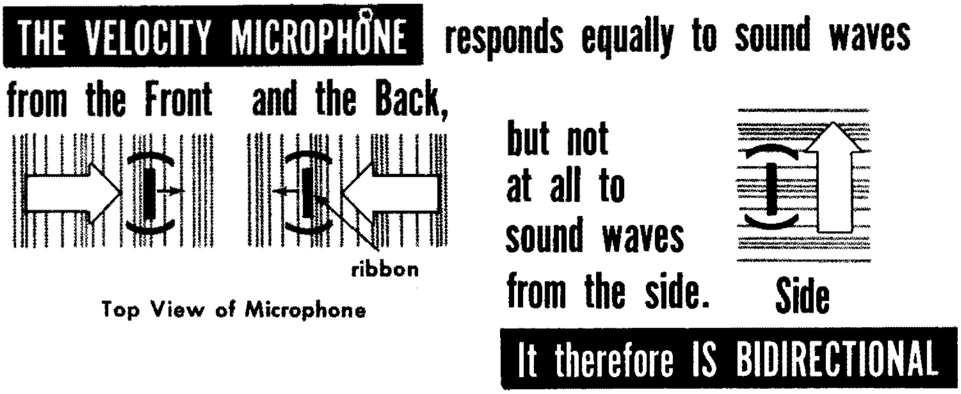
The Carbon Microphone

There is still another kind of microphone in common use; in fact, the carbon microphone forms the basis of millions of telephone instruments in use today. Any loose contact can be susceptible to vibrations around it that will alter its effectiveness in sympathy with the vibrations. A single loose contact, however, makes a poor microphone—all it can do is make noises that keep time with the speech or music. (It is, for example, suitable for relaying the ticking of a watch.)



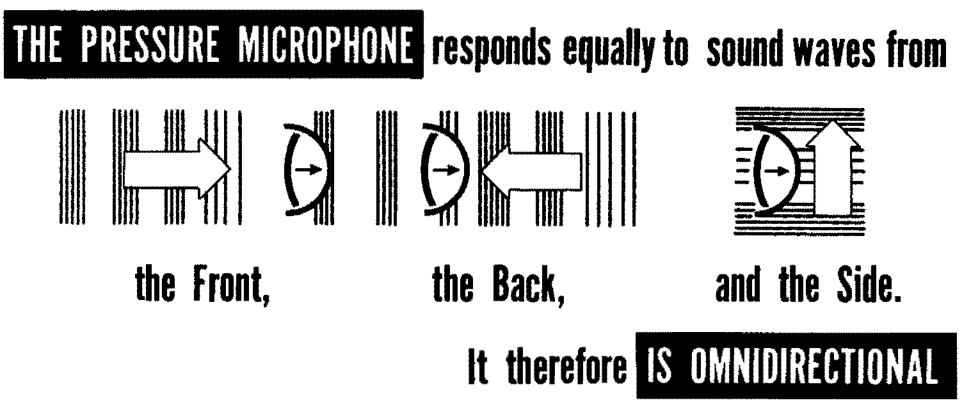
A carbon microphone extends this principle by using thousands of very small loose contacts. The space behind the diaphragm is loosely filled with tiny carbon granules. When the diaphragm vibrates due to sound waves reaching it, the granules are agitated. Because of the large number of contacts, the overall resistance of the microphone through the granules averages out in such a way that it follows the waveform of the sound striking the diaphragm. When a steady source of voltage is applied to the microphone, the current passed through it will fluctuate in sympathy with the sound waves.

Microphone Directivity

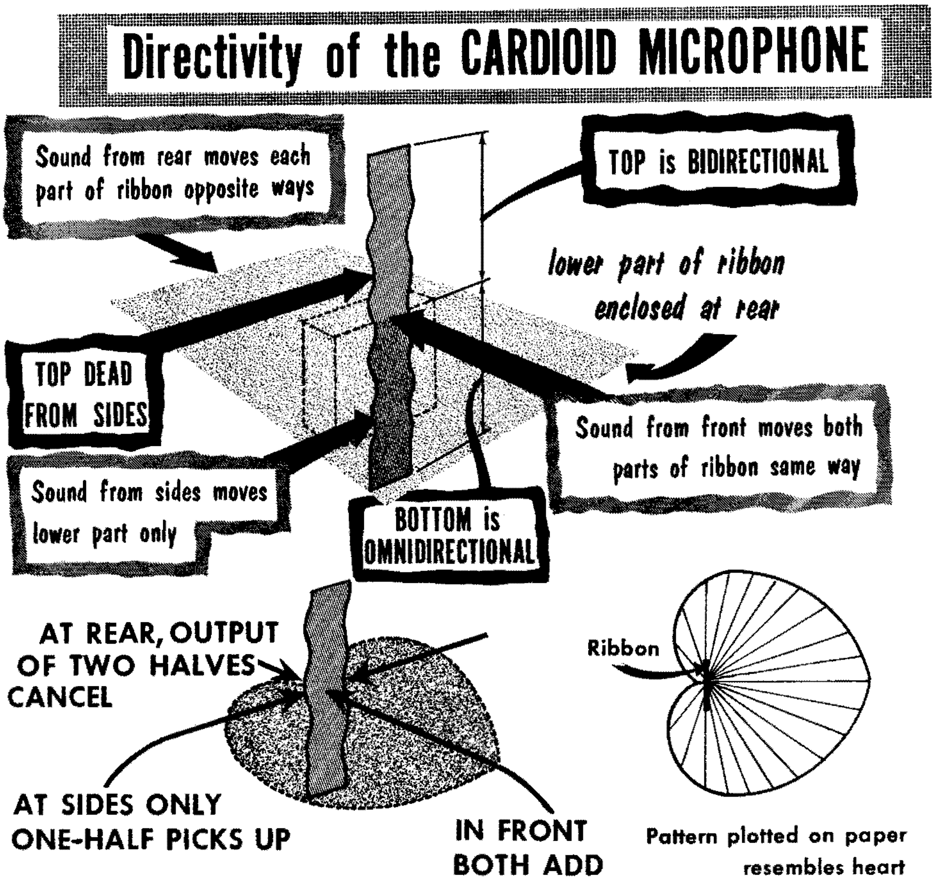


As well as differing between the way that they convert vibrations into electrical voltages or currents, microphones differ in the way that they pick up acoustic waves. For example, the ribbon moves with the air particles as the wave passes it. If a wave passes in a direction parallel to the flat surface of the ribbon, the ribbon will not move, and no sound will be picked up. This property of a ribbon microphone makes it *bidirectional.* This means it is sensitive to sound from two directions (back and front), but not sensitive to sound from other directions.

In most other types of microphones, the back of the diaphragm is shut off from access to outside air, so that sound waves reach it only from the front. This means that the microphone is sensitive only to pressure fluctuations, rather than to a pressure gradient or the air particle velocity. For this reason, such microphones are called *pressure* type. The pressure at the diaphragm varies in the same way, regardless of the direction from which the sound comes. For this reason, pressure microphones pick up sound from all directions, and are called *omnidirectional.*



Microphone Directivity (contd.)



Suppose that the lower half of a ribbon microphone is enclosed at the back so that this portion works like a pressure microphone; when sound comes from the front, both halves will move the same way and the microphone will give maximum output. Sounds from the back, however, will move the two halves of the ribbon in opposite directions, so that the resultant output cancels. Sound from the sides will only move the half of the ribbon that is enclosed at the back, but not the free, or velocity, half. This makes what is called a *unidirectional* or *cardioid* response.

Unidirectional signifies pickup from only one direction—the front. Cardioid describes the heart-shaped sensitivity of this microphone, when plotted on polar-coordinate paper. (To plot this kind of curve, the output from the microphone for a given sound intensity from various directions is marked along radii with corresponding directions.)

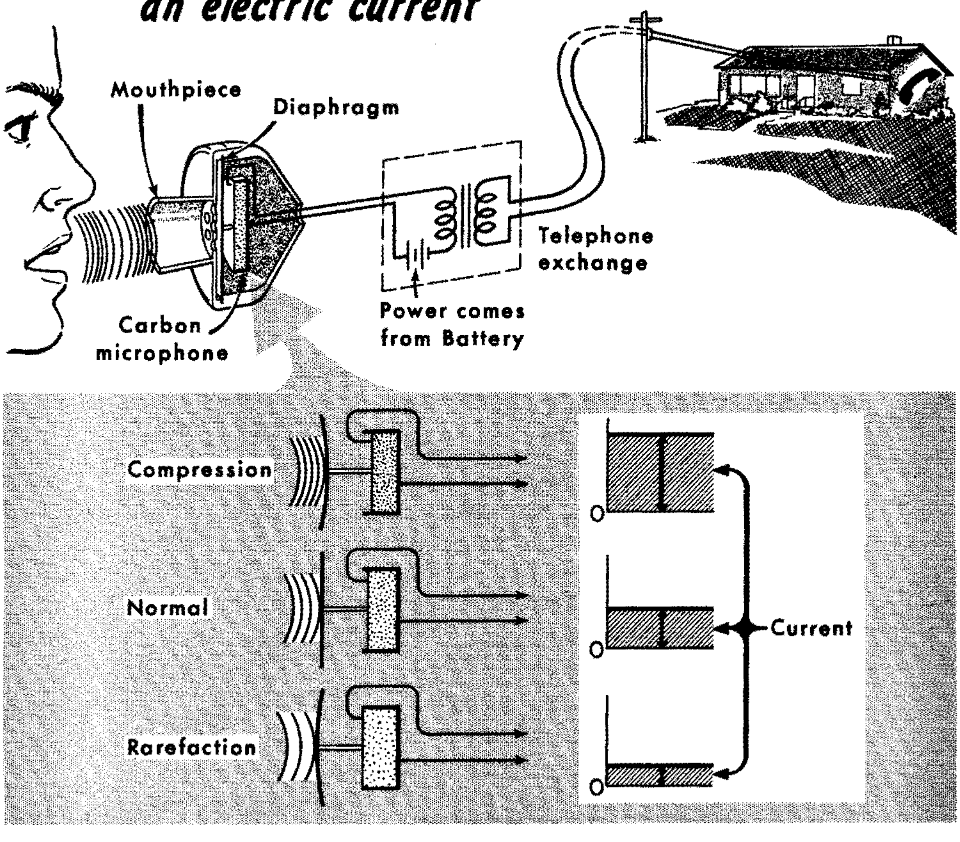
Microphone Sensitivity

A microphone does not amplify sound. The intensity of the waves is very small, whether the diaphragm is arranged to pick up particle movements, as in the ribbon type, or pressure fluctuations. The regular telephone microphone, which is of the carbon type, is effective only for two reasons:

1. It is provided with a mouthpiece to collect all the sound from the mouth of the person speaking into it.
2. It really works by "modulating" an electric current, *not generating one*. The current comes from a battery at the telephone exchange, and the vibrations of the diaphragm modulate this rather large current instead of generating small ones from the vibrations themselves.

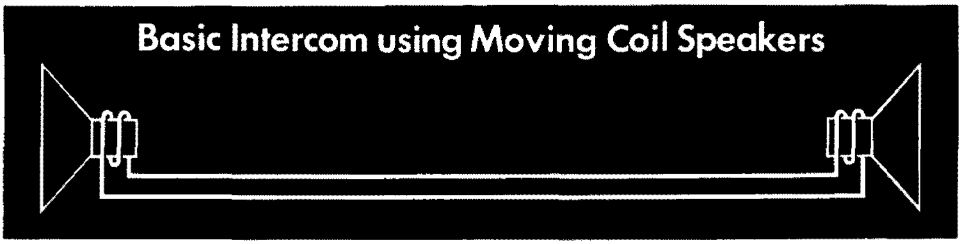
At the receiving end, too, advantage is taken of an earphone to get all the received sound energy right into the listener's ear. (The output of a telephone receiver is not great enough to drive a loudspeaker.)

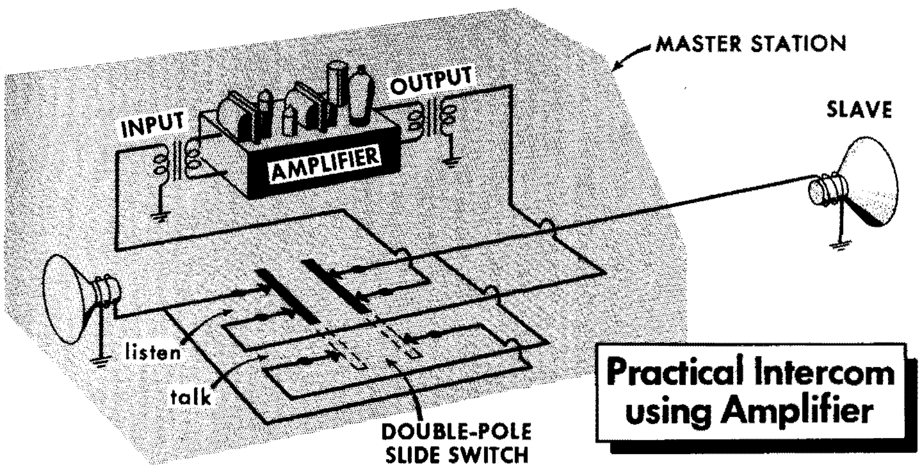
The Carbon Microphone modulates an electric current



Microphone Sensitivity (contd.)

A small loudspeaker can be used to make a microphone that is quite sensitive by ordinary standards. Sound waves impinging on the speaker cone move the voice coil and generate small currents. This type of microphone does not give as good quality as a properly designed one, but it is often used in intercommunication sets of the kind used in offices. If you try connecting two small loudspeakers in different rooms, you will only be able to hear by having someone speak very close to the "microphone" and putting your ear very close to the other loudspeaker.

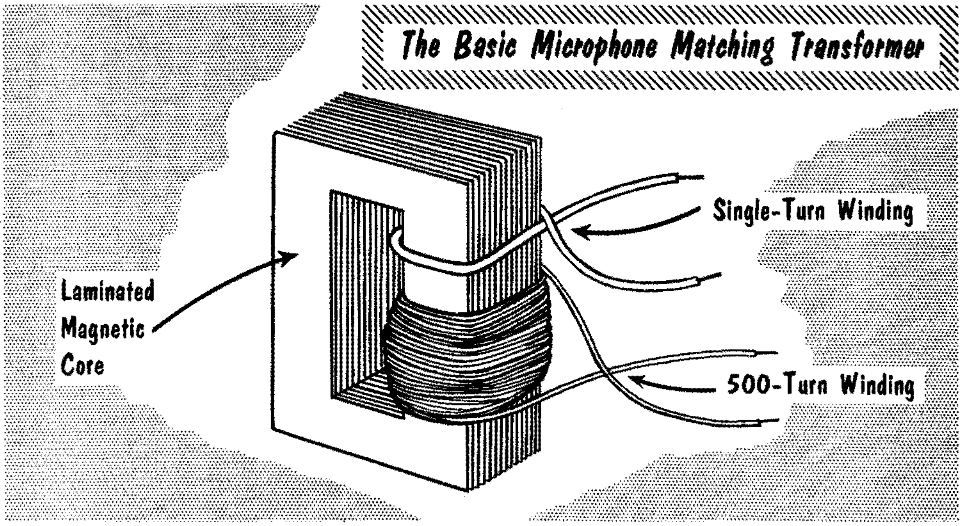




Practical intercommunication sets use an amplifier, between the two speakers, which makes communication much easier. Even then the quality is not good, but has the well-known shrillness associated with such "squawk boxes." The reason for their use is that they are much more sensitive than any other form of microphone, and so make the system less expensive, by requiring less amplification and less attention to special wiring. (If a loudspeaker were connected directly to a moving-coil *microphone*, it would be virtually impossible to hear anything, however close the speaker was to the microphone and the listener to the loudspeaker.) Thus, any practical form of microphone needs electrical (or, more properly, *electronic)* amplification, to get enough current to be of any use in driving a loudspeaker.

The Microphone Matching Transformer

A moving-coil or ribbon microphone directly converts acoustic power into electrical power; however, there is very little power to use, and the microphone must make the best use of what there is. (Electrical power is measured in watts, found by multiplying current in amperes by potential difference in volts; for microphones, the output is in fractions of a *micro*watt.)

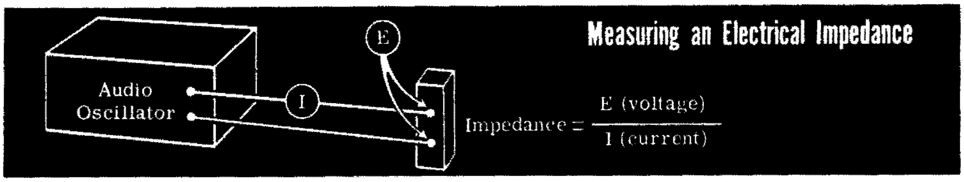


In the moving-coil microphone, there may be as many as 100 turns in the coil, each of which generates 5 microvolts for a particular sound intensity. The whole coil, therefore, gives a total of (100×5) or 500 microvolts. A ribbon does not generate much more than about 1 microvolt (open circuit) because there is not even one whole turn, and the magnetic field cannot be so intense, because of the wider spacing between the poles. However, the resistance of the ribbon is very small, about .05 ohm. Applying Ohm's law, we find that the short circuit current in the ribbon that would flow due to 1 microvolt is about 1/.05 or 20 microamperes. The voltage output of the microphone can be raised by a microphone transformer, which has a small number of turns in one winding, and a larger number of turns on another winding, both wrapped round the same magnetic core.

In a transformer, every one of the turns will have the same voltage "generated" in it; if the single-turn winding generates 1 microvolt, all the other turns will also give 1 microvolt, and a 500-turn winding will give 500 microvolts. The total power will, of course, be the same, and with 20 microamperes available from the microphone into the 1-turn, the 500-turn winding will only give 20/500 or .04 microampere. What the transformer does, in effect, is to replace the actual single ribbon with the equivalent of 500 very much thinner and lighter ribbons, all connected in series.

Microphone Impedance

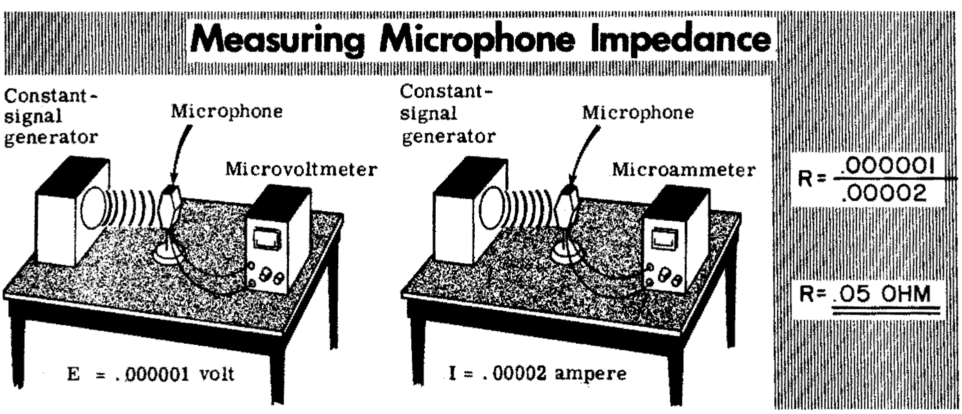
In most electrical circuits, impedance, or resistance, refers to the relationship between voltage and current flowing in some component. This is easy to measure because the voltage source is *external* and the current *through* the components can be measured.



In a microphone, the voltage source is internal or *inside* the component, so that we cannot measure it separately. But we can measure the open-circuit voltage and then measure current when we short-circuit the microphone, so that there is no voltage. The relationship between this voltage and current is the impedance of the microphone.

Going back to the theoretical case of the previous page: without the matching transformer, the open-circuit voltage was 1 microvolt, while the short-circuit current, determined by the resistance of .05 ohm, is 20 microamperes.

In the transformer secondary, the open-circuit voltage is 500 microvolts, while the short-circuit current is .04 microampere. Therefore the effective impedance on the transformer secondary is 500/.04 or 12,500 ohms. This is the impedance the microphone presents on the transformer secondary.



Thus the real impedance of .05 ohm has been multiplied by 12,500 .05, or 250,000. The impedance matching ratio 250,000:1 is the square of the turns ratio (500 $\times$ 500 = 250,000).

QUESTIONS AND PROBLEMS

1. Why is a microphone needed in an electrical amplifying system, and what are its two basic actions?

2. What is the essential feature of a dynamic microphone? Indicate which of the following types can be called dynamic: moving-coil, ribbon, condenser, crystal, carbon.

3. Explain the reason for the construction of a moving-coil microphone with special reference to (a) the shape of the magnet, (b) the shape of the coil, (c) the requirements of the diaphragm.

4. All the following designations may correctly be applied to the same microphone; however, some of them may also be applied to other types to which the remaining designations are not applicable: dynamic, ribbon, bidirectional, velocity, pressure gradient; explain these differences.

5. What feature do the ribbon and condenser type microphones share in common? In what respects do they differ?

6. What operation similarity is there between a condenser microphone and a carbon microphone?

7. How does a crystal microphone work, and what steps are taken to make its sensitivity comparable with that of other types?

8. What are the three basic directional characteristics of microphones?

9. What is meant by (a) pressure and (b) velocity microphones? Do these methods of operation have any connection with directional pattern? If so, into which group would you classify a cardioid pattern?

10. If someone asked for a "directional" microphone, what types could be intended? Explain.

11. Give two reasons why the type of microphone used for telephones does not need electrical amplification.

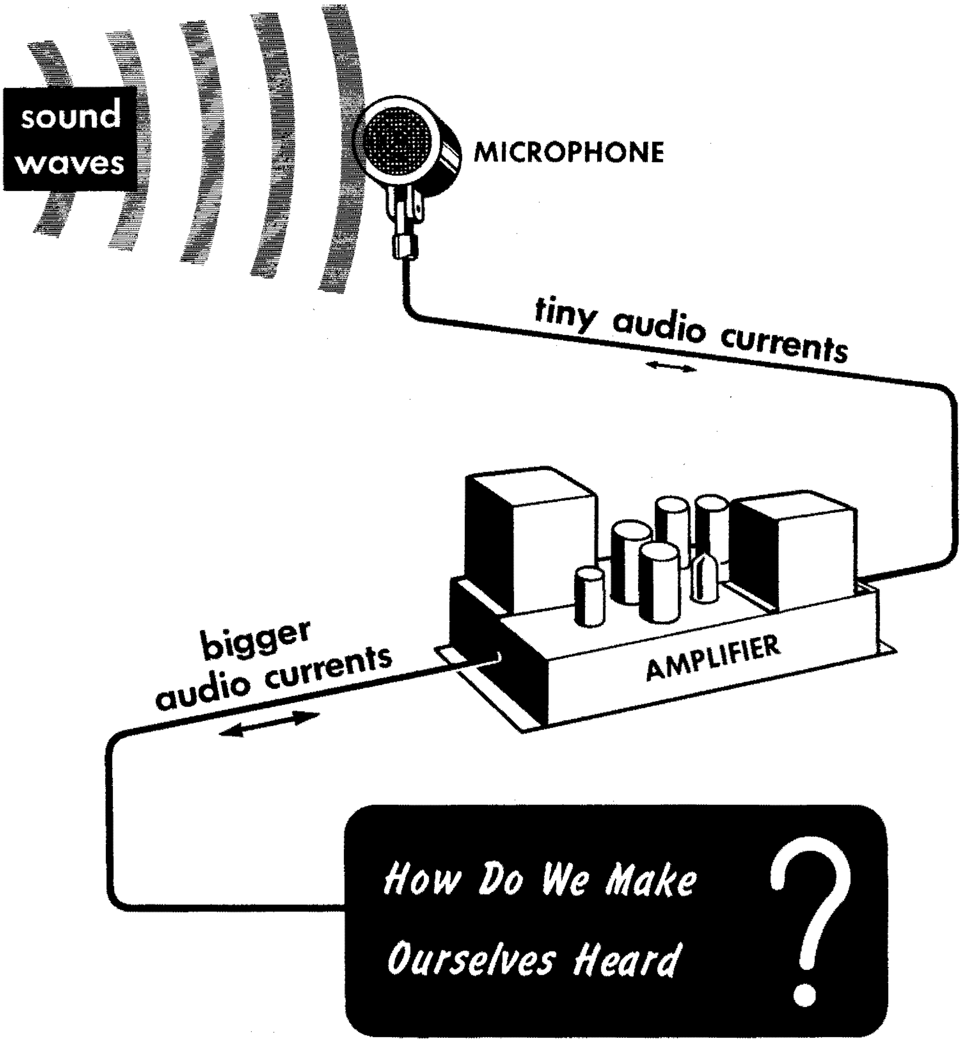
12. Would you expect a small moving-coil loudspeaker to make a good or bad microphone? Explain.

13. Why does any high-quality microphone need amplification?

14. When a ribbon microphone whose resistance is .05 ohm is used with a 500:1 step-up transformer, it produces 500 times the voltage output for a given sound wave. What is the effective resistance of the equivalent microphone with its transformer?

The Purpose of Loudspeakers

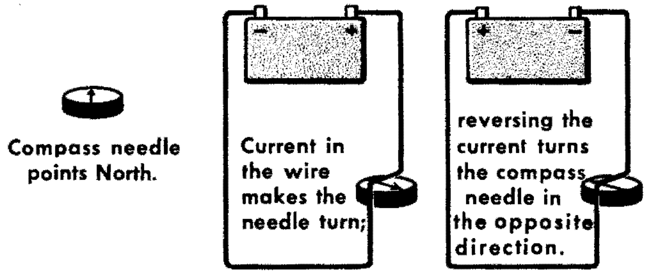
Amplifying the currents from the microphone is not enough—we cannot hear currents! (Birds do *not* sit on telephone wires to listen to the conversations.)



It is the purpose of loudspeakers to convert the varying currents back into sound waves. In fact, loudspeakers may be thought of as microphones in reverse. An example of the use of an actual speaker as a microphone was given earlier, and the principles of operation of the microphones already discussed are all used in making loudspeakers.

The Moving-Iron Speaker

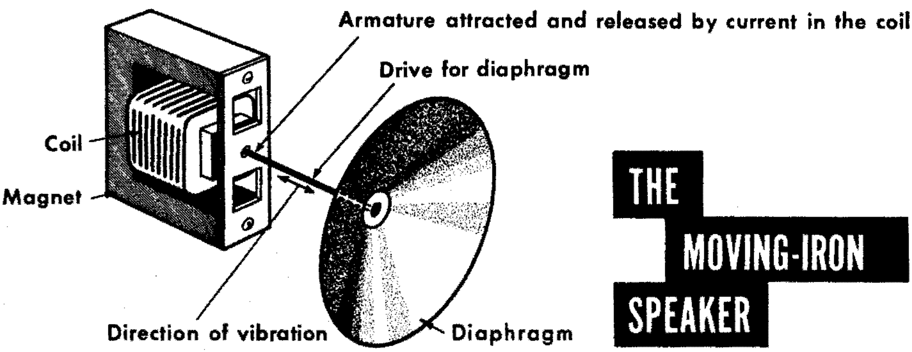
THE PRINCIPLE OF MOVING-IRON SPEAKER OPERATION



If you suspend a magnet or a compass needle over a wire connected to a battery, the needle will swing to one side. If you reverse the current in the wire, the magnet (or needle) will swing in the opposite direction.

The force given by the current in this setup is feeble—we could never produce appreciable sound output by passing the fluctuating sound currents through the wire (after attaching a diaphragm to the magnet). We can increase the force by using a stronger magnet, coiling the wire into many turns, and shaping the magnet so as to concentrate the effect of the coil. A strong magnet, however, must be heavy, and it would have difficulty moving at the speed of sound vibrations. For this reason, a large coil is used to provide the magnetism, but only a small piece of magnetic material (iron) is used to move the speaker cone.

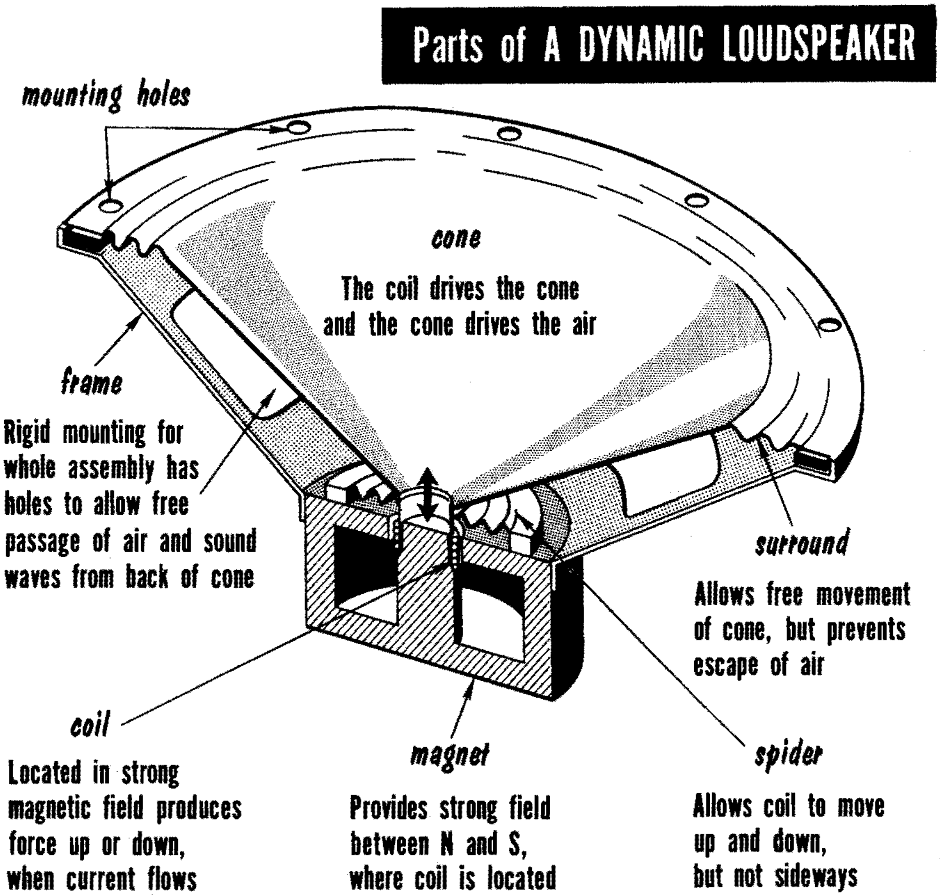
In the early days of radio, thousands, if not millions, of these loudspeakers were made. Because of their construction, however, they could not respond very well to the wide range of frequencies used in sound, from the lowest tone given by an organ or other musical instruments—about 32 vibrations per second—to the highest overtone or harmonic required to give the correct "character" to sound. The iron or steel armature, moved by the currents, was too stiff to move freely enough at low frequencies and too heavy to move as rapidly as needed at high frequencies, hence the extreme frequencies did not get reproduced satisfactorily.



THE MOVING-IRON SPEAKER

The Moving-Coil Speaker

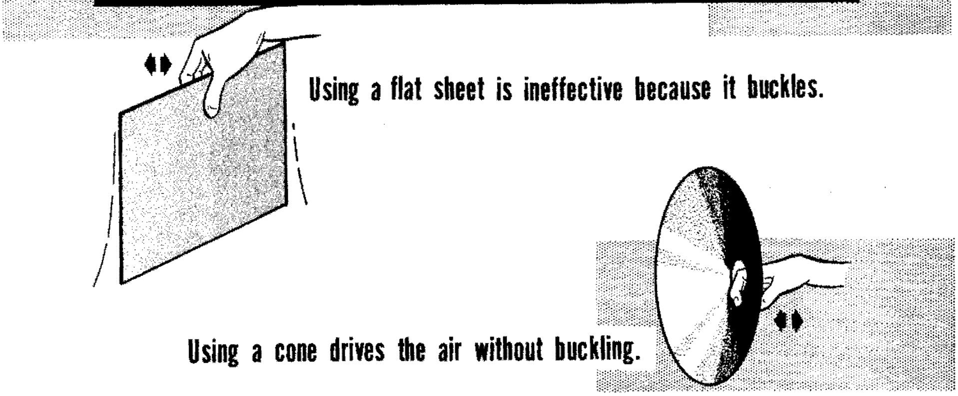
The *moving-coil* speaker was designed to overcome these defects. The magnet can now be as big and heavy as we choose, because it does not have to move. The coil is small and light, and is suspended by a light flexible material that will allow free movement at all frequencies. This makes a much better loudspeaker.



But why was the coil made round, or cylindrical? The requirements for getting maximum force to drive the cone (or diaphragm) are a strong magnetic field, and as much length of wire as possible in the field. The narrow cylindrical air gap permits a strong field that is easy to create. The attraction between the poles is uniform at all points, hence the center pole does not try to get out of position due to magnetic pull. By winding a thin, flat, cylindrical coil to work in this gap, the coil is kept light in weight, with maximum strength in the direction needed for drive.

The Speaker Diaphragm

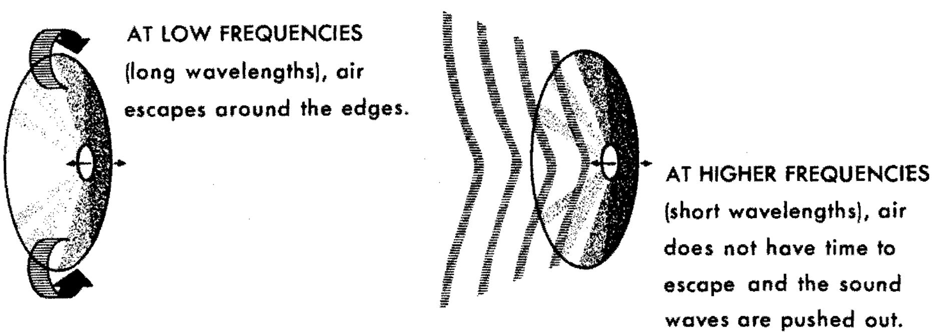
Why the Speaker Diaphragm is a Cone



Next, why the shape of the diaphragm—a cone? Again the requirement of rigidity with lightness of weight. Try waving a piece of cardboard or paper in a large sheet, and you will find it is impossible to move any air rapidly back and forth—wherever you hold the paper or card. But take a cone of paper and hold it by the center, where the coil drives a loudspeaker cone, and you will be able to push air back and forth with it much more easily, although it is just as light, as the flat sheet. The conical shape gives rigidity.

But how big should the cone be? This depends on the wavelength of the sounds we want to make. If the size of the diaphragm is smaller than a wavelength, the air tends to run round the edges instead of going back and forth *with* the diaphragm. If the diaphragm is large compared to the wavelength, the air will not have time to dodge around it during the passage of each wave.

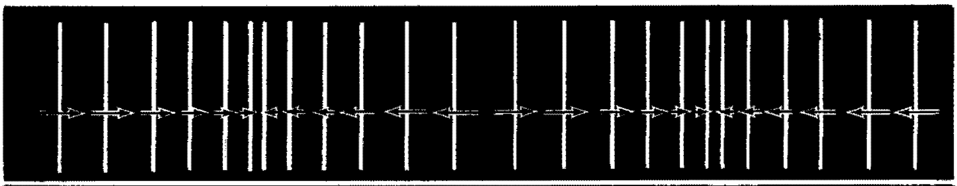
Cone Size and Frequency



The Speaker Diaphragm (contd.)

Wavelength varies inversely with frequency. The low-frequency tones have long wavelengths, whereas the high-frequency tones have short wavelengths. This means that we need a large diaphragm for the lower frequencies, while a smaller one will serve for the high frequencies. Of course, a large diaphragm will also move air at the higher frequencies, unless it is too heavy to be driven effectively at the higher speeds. For this reason, many installations use two or more speakers of different sizes, each of which handles only the band of frequencies that it serves best: large speakers (woofers) for the low frequencies, medium-sized speakers (squawkers) for the mid-range, and small speakers (tweeters) for the high frequencies.

Because of the "size" of the wave at low frequencies, a diaphragm has to move a lot of air to make sound. If you look at a loudspeaker diaphragm, you will see that it moves quite a long way for the low frequencies, although at higher frequencies, it radiates sound without visible movement. At these higher frequencies, however, the air load on the diaphragm is so great that the voice coil can hardly move. The high frequencies thus are not radiated.



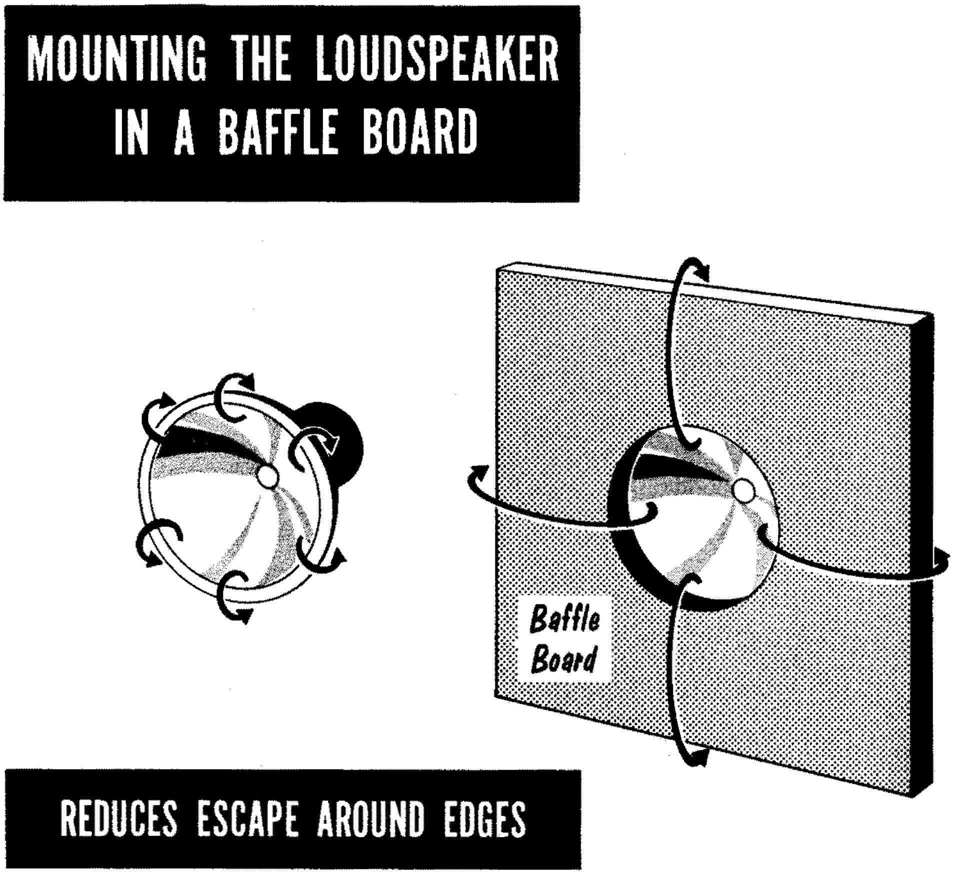
At low frequencies (long wavelengths), large volumes of air are compressed and expanded, requiring large air movement.

At high frequencies (short wavelengths), small quantities of air move small distances.

The reason for this difference in movement can best be understood by thinking of wavelengths. To create a pressure change, air has to be pressed into a given space. When the wavelength is small (as at high frequencies), the space momentarily requiring the additional air is small, and the pressure can be increased by very small (but very rapid) air movement. When the wavelength is large, the area of increased or reduced pressure is large and requires the movement of a bigger mass of air to achieve it.

The Use of a Baffle

To avoid air escaping round the edges, the diaphragm should be at least one-half a wavelength across. If we want to radiate a frequency of 32 cycles, for which the wavelength is 1100/32 or about 34 feet, the diaphragm would need to be about 17 feet in diameter. One way of avoiding this awkward size is to mount the loudspeaker in a baffle.



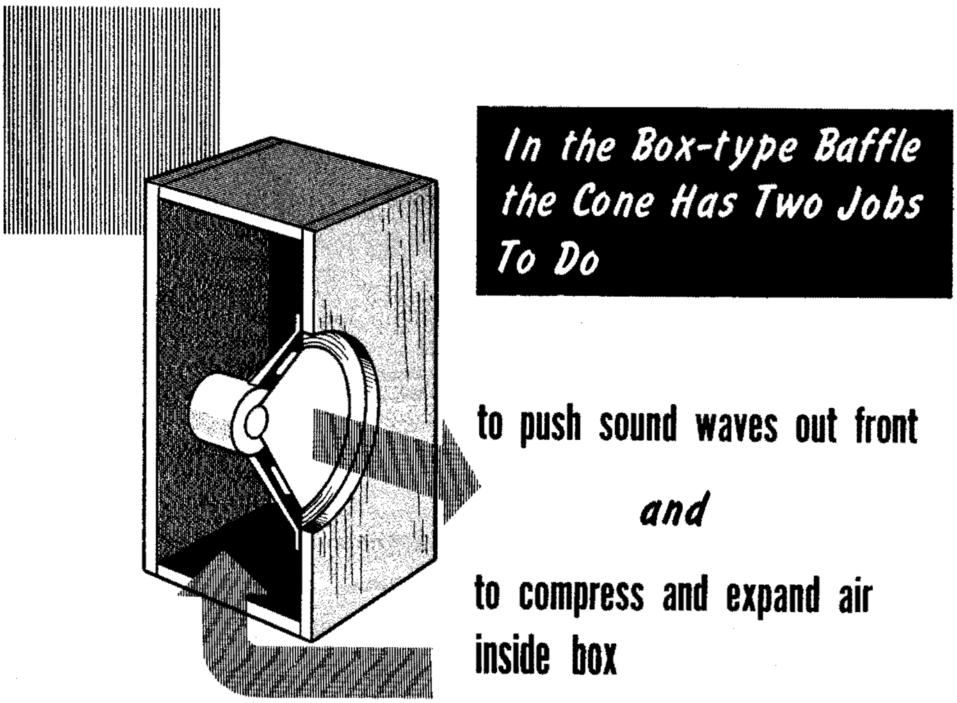
The loudspeaker diaphragm is continued by being joined to a solid, fixed board that extends out to the required size. This stops the air from escaping around the edges. Of course the diaphragm will have to move further to get the same amount of air movement than would one the full size of the baffle, because only a fraction of the surface is moving. But preventing the air from escaping around the edges of the diaphragm will also prevent the diaphragm from moving without radiating a wave at all. The situation is thus much better than without the baffle. Of course, a smaller baffle will do the same thing, except that it will not be effective to such a low frequency.

The Infinite Baffle

Next we consider ways to "tidy up" the baffle. A big board suspended in space, with a loudspeaker mounted in a hole at its middle, is not an ideal piece of furniture. And unless it is very big, it still has limitations in low-frequency response. Of course, if the baffle could be made infinitely large, by extending it up to the sky, it would be perfect, but this is obviously not practical.

The purpose of an infinite baffle is to prevent any access from back to front around the edges. Putting the loudspeaker in a box that is completely closed except for a hole for the diaphragm does the same thing. This works perfectly, as far as enabling the diaphragm to radiate frequencies right down to the lowest is concerned. However, this arrangement gives the diaphragm two jobs to do; as well as radiating sound from the front, it must alternately compress and rarefy the air inside the box.

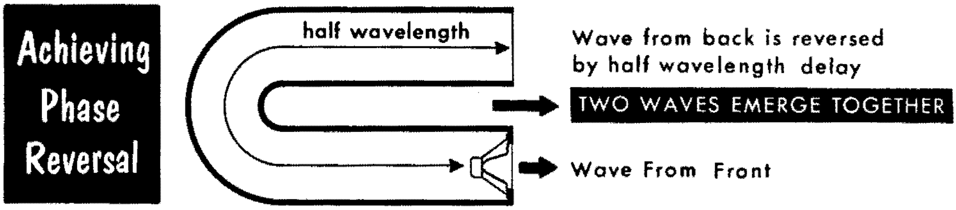
The air in front of the diaphragm moves much more readily than that behind it (which is restricted by being contained in the sealed box). For this reason, the air in the box places a greater "loading" on the diaphragm movement, and only a fraction of the driving force produced by the voice coil is used to radiate sound from the front; most of the energy is wasted compressing and rarefying the air inside the box. The smaller the box, the bigger the waste, and the less efficient the complete assembly is as a loudspeaker.



The Bass-Reflex Baffle

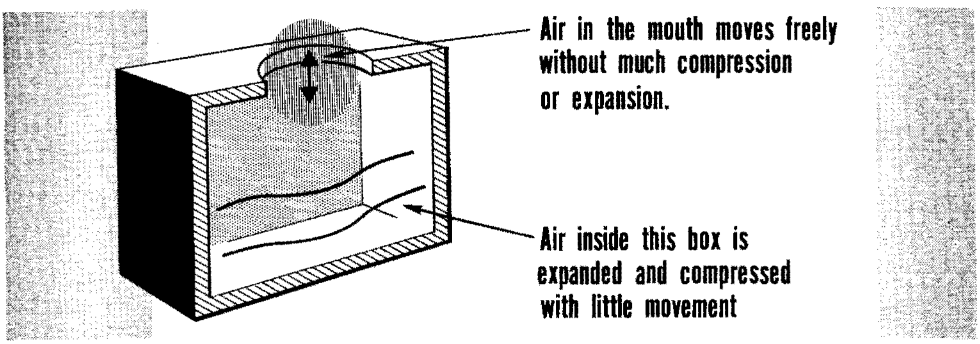
Two main facts bothered loudspeaker designers:
1. To get any radiation of power at low frequencies, the diaphragm has to move a lot of air.
2. The diaphragm moves air on both sides of it—so it would be an advantage to use *both* sides. The movements are, however, *in opposition;* the front pushes when the back pulls, and vice versa.

If only the movement from the back could be "turned around!" This is just what the bass-reflex does.

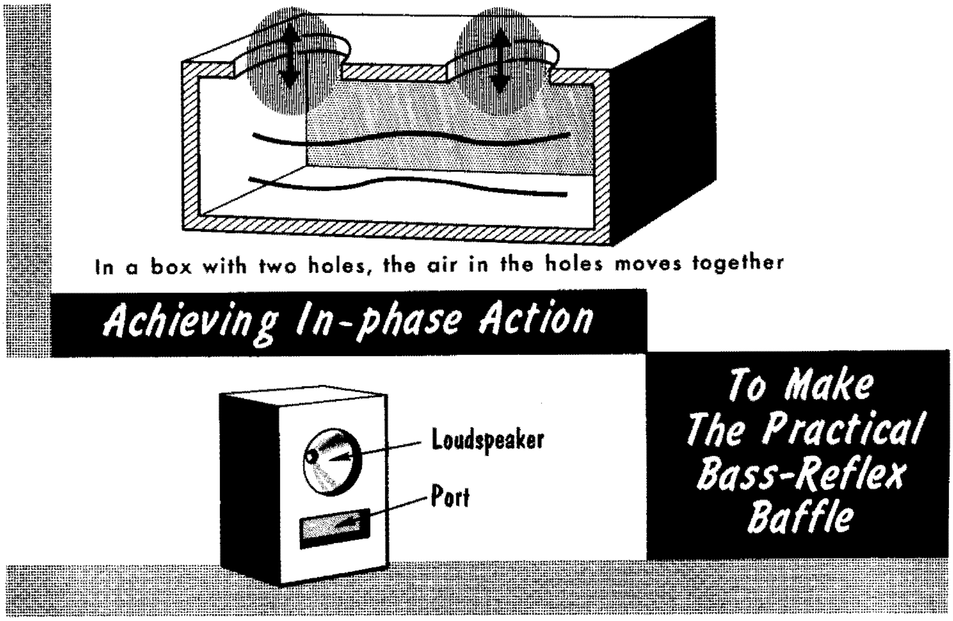


In normal propagation of sound, the only way to get the waves from the back turned around would be to take them for a half-wavelength longer journey than the waves from the front. To reproduce 50 cycles per second, this requires a propagation path of eleven feet, which can hardly be contained in a livingroom piece of furniture.

When sound travels through air, the wave maintains natural relationships between the pressure fluctuations and particle movements, because the air is free to move and allows the wave to develop. When air is confined in a space or an opening, however, the natural relationship no longer exists. Air in the mouth of the box is free to *move*; so it will not compress, but moves freely. When air at the mouth moves in and out, the air inside the box *compresses* and *rarefies*, but it does not move much. Hence *(relatively speaking)*, the air in its mouth moves without appreciable changes in pressure; the air inside the box changes pressure without appreciable movement. Of course, there is no definite line where the air changes from one condition to the other, but most of the air associated with the box is in either one condition or the other.



The Bass-Reflex Baffle (contd.)



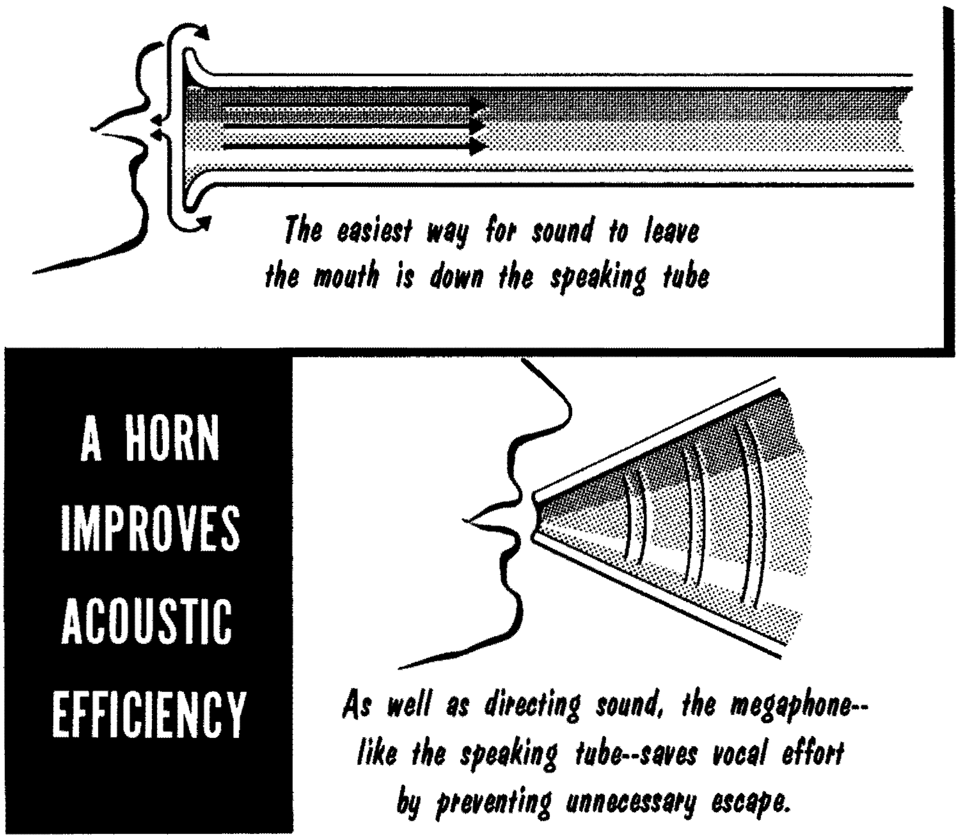
When a box has two openings of the same size, the fact we just discussed means that air movement in both mouths must be approximately *in-phase.* The air inside the box does not move much, but only compresses and rarefies, whereas that in the mouths moves without compressing and rarefying appreciably. If the air in the mouths did not move in and out at the same time, there would have to be considerable air movement inside the box which does not occur. It is easier, particularly at the resonant frequency of the box, for air to move with both mouths working together, or *in-phase.*

In a bass-reflex cabinet, or *enclosure* as it is called, one of the mouths is occupied with the loudspeaker diaphragm, which drives the air in that mouth, whereas the other is just an open *"port."* The dimensions of the second mouth, or port, are adjusted so that the air in the port is about equal in weight to the total of the diaphragm and the air it moves. This provides the correct condition for the two waves to emerge in phase at the resonant frequency of the box. The frequency is lower than could be obtained with the so-called infinite baffle of the same size, and helps radiate sound energy from both back and front of the diaphragm, without cancellation.

At higher frequencies, the box does not act in this way, but absorbs the movement from the diaphragm in the volume of air inside the box, without moving the air in the port mouth appreciably. Hence, at these frequencies, this arrangement works in the same way as the infinite baffle.

The Use of a Horn

The best diaphragm- or cone-type loudspeaker is a very inefficient transducer: only a small part of the electrical power delivered to it is converted into sound waves. An average efficiency figure for modern units is about 10%; a poor one may be 5% efficient or less; a good one *may* reach 20%. Thus even the best throws away four-fifths of the power that it gets. The main reason for this inefficiency is the difference in the density of the material of which the diaphragm is made and that of the air. Because of this big difference, the voice coil spends much more of its driving force moving the diaphragm than the air in contact with the diaphragm.



The easiest way for sound to leave the mouth is down the speaking tube

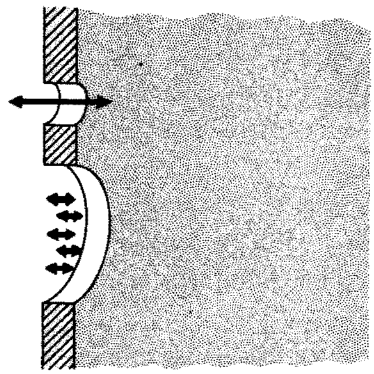
As well as directing sound, the megaphone--like the speaking tube--saves vocal effort by preventing unnecessary escape.

Using a *horn* is a way to improve this situation. The method by which a horn improves acoustic efficiency can be illustrated by two devices: a speaking tube and a megaphone. In a speaking tube, sound is propagated without the wave being allowed to expand. For this reason, the intensity does not fall off according to the inverse-square law, but reaches the other end of the tube almost undiminished.

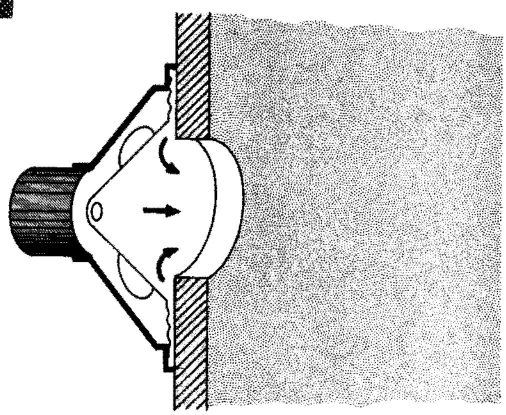
The Use of a Horn (contd.)

The megaphone, in addition to concentrating the sound into a narrower angle, intensifies the sound in that angle. Without the megaphone, sound comes from the mouth in a sudden transition from a narrow aperture (the mouth) to "free" space. The megaphone, by tapering off this transition, enables the mouth to radiate more sound with less effort.

To transmit a certain amount of air or sound wave through a small hole **requires the air to be forced further (and harder) than in a large hole**



Putting a diaphragm over a smaller hole **increases the effort to move the air and reduces diaphragm movement, improving efficiency in converting voice-coil force to air movement**

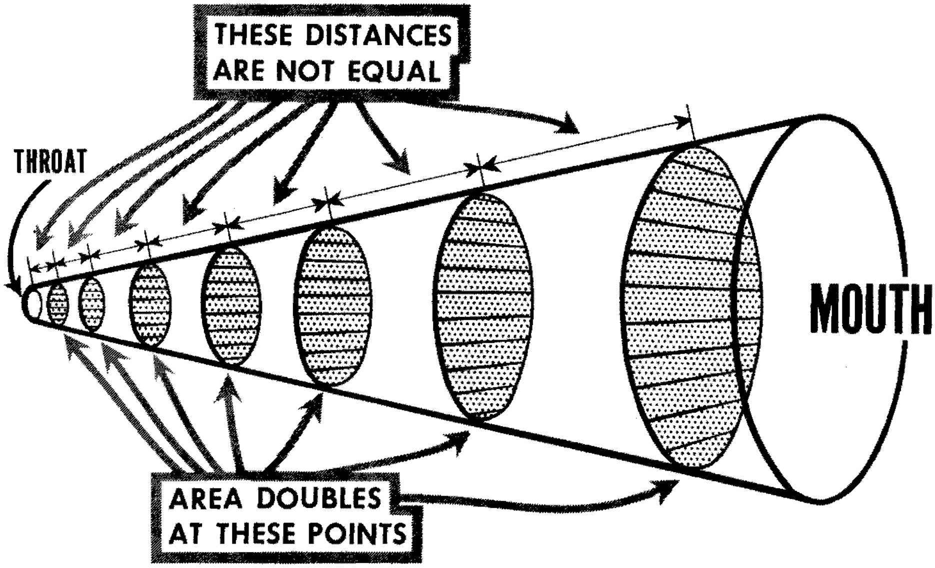


It requires more force to move air in a narrow channel than it does in a wider channel, because individual air particles in the narrow channel have to move farther to allow the same quantity, or volume movement. For this reason, placing a diaphragm opposite a channel that is smaller in cross section area will enable the voice coil to spend more of its energy in driving the air, wasting less in driving the diaphragm. This is like putting a speaking tube to the diaphragm. If we couple the tube to a megaphone, we shall have a more efficient sound radiator than the diaphragm by itself. The narrow opening makes the diaphragm move more air, and the megaphone effects a transition more gradually, so that this greater movement is used, instead of being lost.

Horn Shapes

A horn thus serves two purposes: it confines the sound into a narrower angle, but, more important than this, it improves acoustic efficiency by preventing the losses that occur at sudden transitions. This improvement in efficiency is sometimes the only reason that a horn is used, since some horns are designed to distribute sound in all directions.

Expansion in Conical Horn



Area inside a cone-shaped horn does not double at constant changes in distance along the cone. Therefore the sound waves do not expand at a constant rate.

The megaphone is a straight-sided cone, because this happens to be a convenient way to make it. Although it does improve acoustic efficiency, other shapes will do so even better. The purpose of a horn is to "stretch" the developing wave gradually. (The wave should not expand too rapidly suddenly, and making the expansion too gradual will make the horn unduly long.) In the conical horn, the rate of stretch varies. Near the *throat* the area doubles in only a short distance, whereas it expands more slowly near the mouth.

Horn Shapes (contd.)

A horn is effective in producing a *smooth* transition only if it takes not less than 1/18th of the wavelength corresponding to the lowest frequency required to double its area. (This relationship has a mathematical basis too advanced to give here, although it has been verified experimentally.) Suppose that the lowest frequency for a horn is to be 400 cycles per second (this is not very low, but is often used for practical equipment, because it results in a useable size): the wavelength for 400 cycles at 1100 feet per second is about 33 inches; so the horn should double its area in *not less than* about 2 inches (1-5/6 inches to be exact). In addition, the diameter at the mouth of a horn should not be less than about half a wavelength for the lowest frequency that is to be radiated. For a 400-cycle horn, the diameter at the mouth should not be less than about 16½ inches.

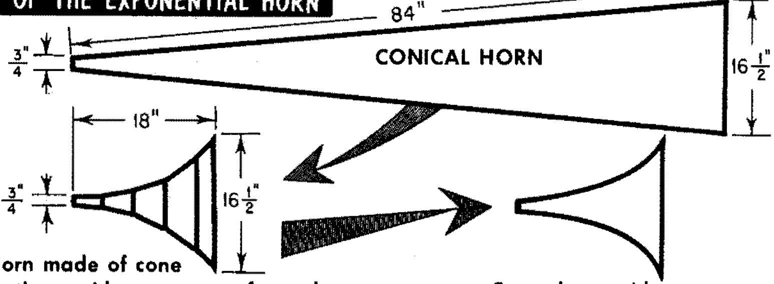
To get the best acoustic energy match into the throat of the horn, a diaphragm about 1½ inches in diameter works into a hole about ¾ inch in diameter. The horn thus has to stretch the ¾-inch diameter to about 16½ inches. Doubling the diameter will quadruple the area, and this should never happen in less than 4 inches of horn length.

If it takes 4 inches to build up from ¾ to 1½ inches, it will require $4 \times (16\frac{1}{2} - \frac{3}{4})/(1\frac{1}{2} - \frac{3}{4})$ or 84 inches to reach 16½ inches with a conical horn. This is a horn 7 feet long, to handle frequencies only down to 400 cycles per second. But it is only this long to get the correct rate of stretch near the throat. If we make a horn of a group of cones so that the diameter doubles every 4 inches, it will be:

Distance from throat (inches)	0	4	8	12	16	20
Diameter (inches)	¾	1½	3	6	12	24

The required diameter is reached between 16 and 20 inches from the throat. Even this shape is not ideal, because the rate of stretch is not quite uniform —it changes suddenly where the conical sections join. The best shape smooths out these joints, so that the stretch is uniform. This shape is called *exponential.*

DEVELOPMENT OF THE EXPONENTIAL HORN

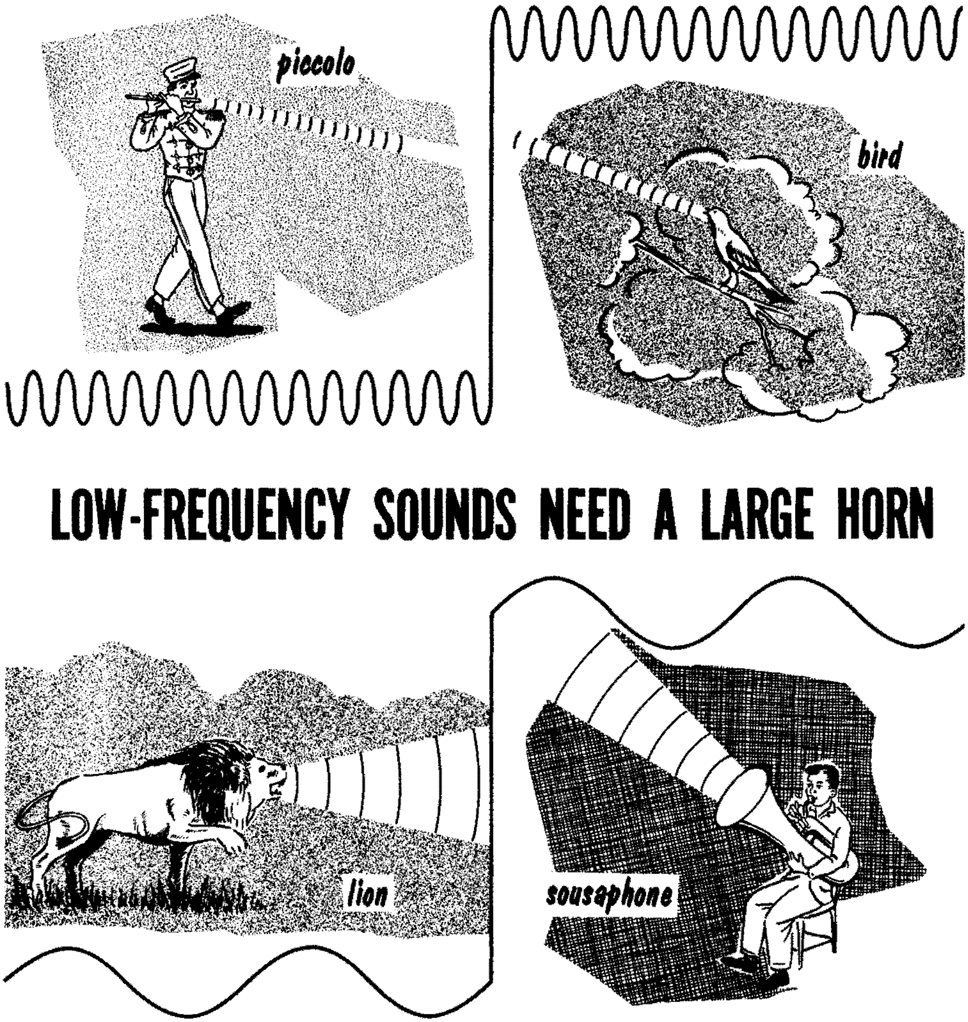


Horn made of cone sections with same rate of stretch of wave as the start of the cone

Same horn with smooth exponential curve

Horn Shapes (contd.)

If a horn is required only for frequencies above 800 cycles the rate of flare could be increased. It could double its diameter every 2 inches of length instead of every 4 inches. In addition, the mouth could be half the size.

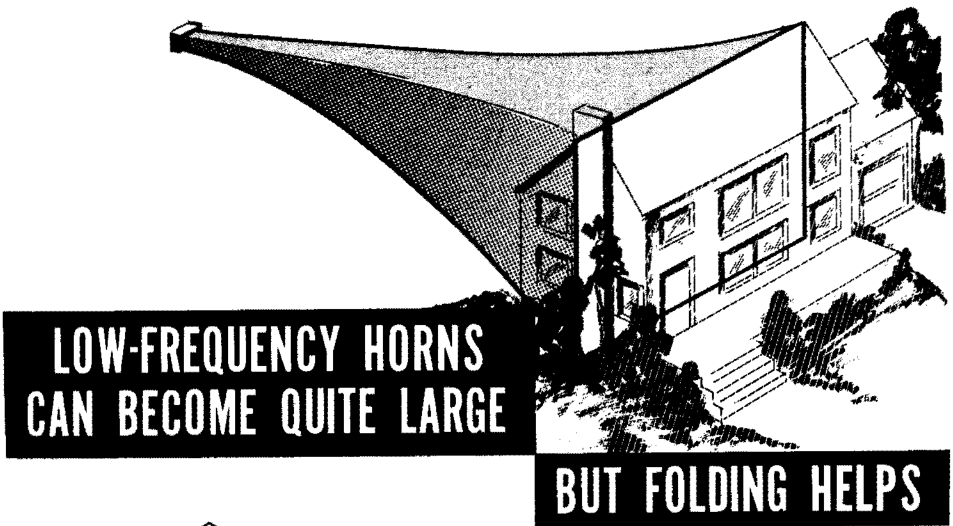


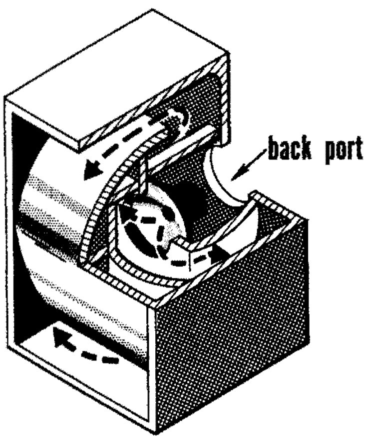
From this we can see a pattern emerging: low frequencies need a big thing to radiate them successfully, but high frequencies can come from quite small objects or sources. Which explains why small birds chirp at higher frequencies, and only larger animals, like lions, can produce a deep-throated roar.

The Folded Horn

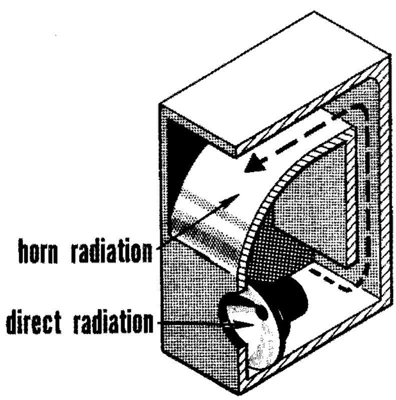
The horn is a wonderful way to improve acoustic efficiency, but where we really could use this help—at the low frequencies—it has to be very large. For example, to get down to 100 cycles, the mouth needs to be 5 feet across, and the length, starting from a throat 4 inches across, needs to be 64 inches, (over 5 feet). To work from a ¾-inch throat, using a regular horn drive unit, we would need a horn over 9 feet long.

To show that the principle works, horns about 30 feet long have been built, getting down to 40 cycles. But these are hardly practical for everyday use. The 100-cycle horn however is not so very big, although it is somewhat awkward. This can be overcome by "folding" it.





Double-Folded Horn

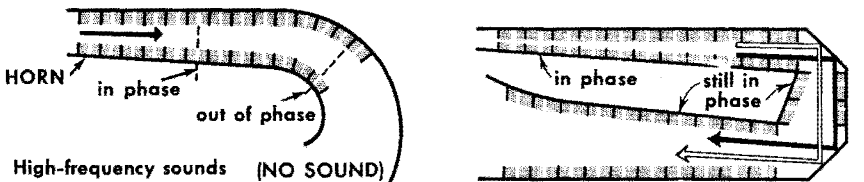


Back-Loaded Folded Horn

The Folded Horn (contd.)

For low frequencies, the folding of a horn has no disadvantages—in fact, any horn can be folded without disadvantage to frequencies near the lower end of its frequency range. However frequencies that are well above the cutoff frequency (the lowest it will handle) tend to get "lost" due to folding. Different paths down the horn have somewhat different lengths, and the higher frequencies can get to a condition in which the wave following a longer path cancels one taking a shorter path.

Avoiding High-Frequency Cancellation

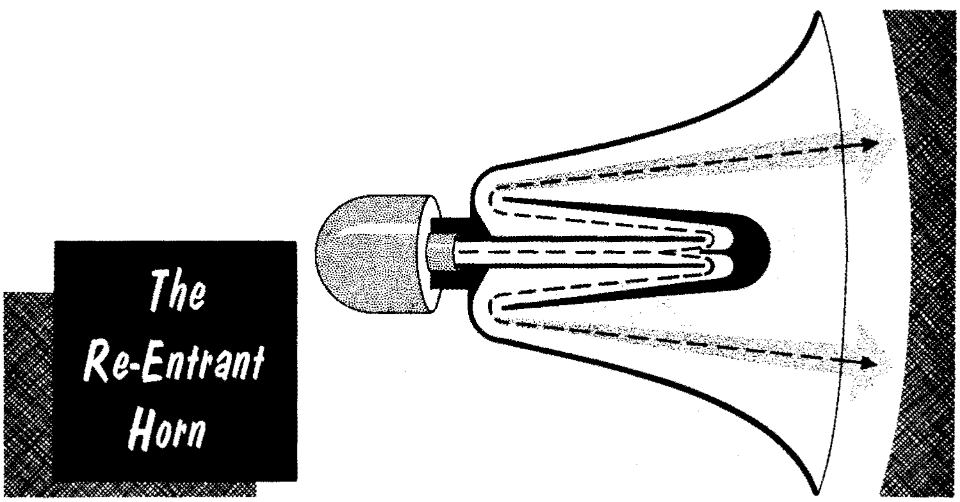


High-frequency sounds get no further than the first bend due to cancellation.

By bevelling corners high frequencies are reflected instead of being cancelled

One way in which to compensate for this problem is to design the bends of the horn so as to "invert" the waves at higher frequencies, thereby equalizing the path lengths. Instead of following a smooth contour at the corners, the shape is arranged to reflect the complete wave.

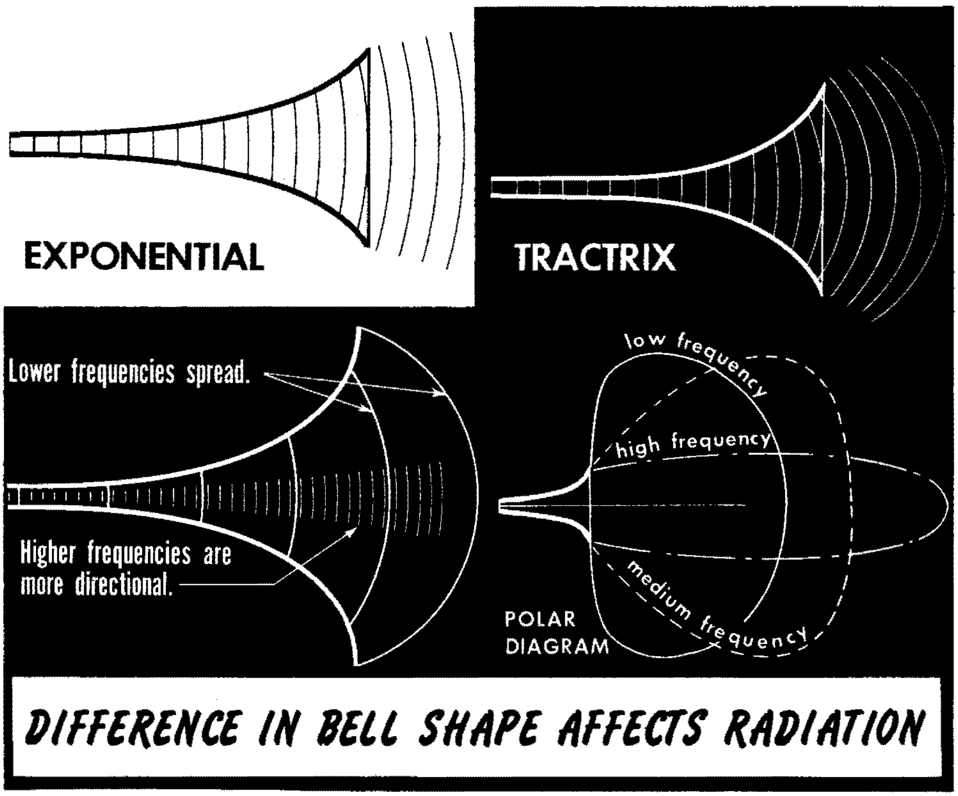
For horns designed for higher frequencies, a very convenient form of construction is the re-entrant type horn, in which the whole structure is concentric.



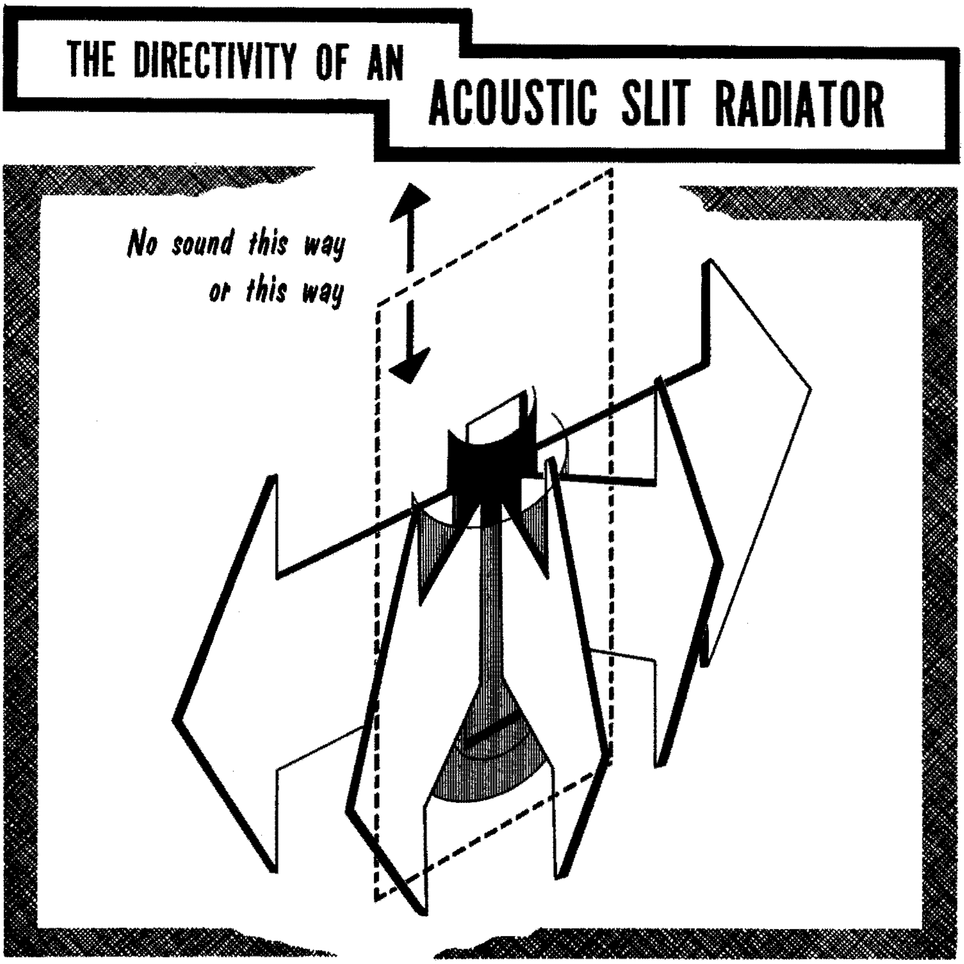
Horn Directivity

Like the original megaphone, horns tend to be directional in their radiation of sound. The strongest part of the wave is the section formed by the horn development. Radiation outside this angle is due to a kind of spill-over.

The directivity of a horn is dependent on its precise shaping. For example, a shape called the *tractrix,* which is nearly identical with the exponential shape, except at the mouth, is designed to produce hemispherical waves; in other words, to cover the entire 180° in front of its mouth. This ideal distribution from the tractrix only exists at the lower frequencies it radiates. At higher frequencies, the bell has progressively less effect.



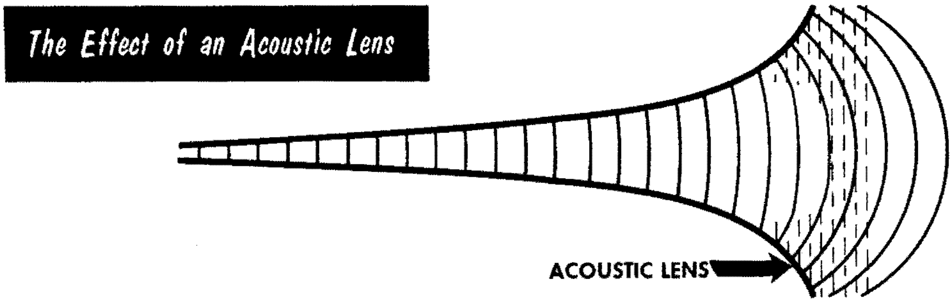
The Acoustic Slit



The only way to improve the uneven distribution at different frequencies is to use a different principle altogether. One such is an acoustic-slit radiator. Because the whole length of the slit "puffs" and "sucks" at the same time, it generates a cylindrical wave, with the slit as a center line. The slit will work well provided that its length is greater than half a wavelength, and its width is less than half a wavelength. Thus its length and width fix the range of frequencies for which it is effective.

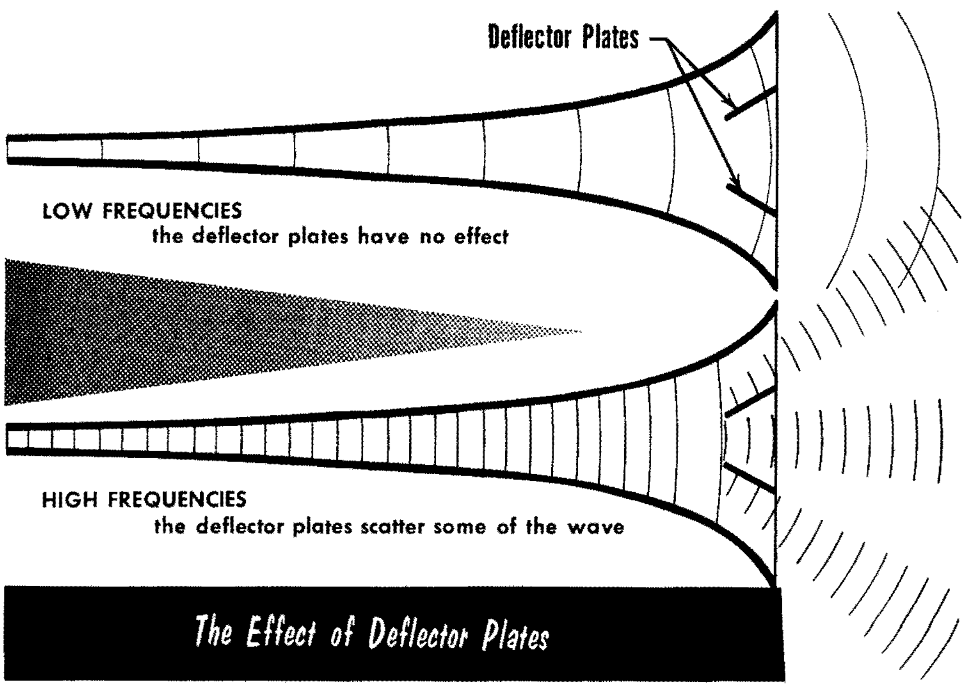
As an example, a slit 10 inches long and 1 inch wide will be effective between about 600 and 6000 cycles. Below 600 cycles it will be ineffective because of acoustic mismatch (unsatisfactory transition from the slit to free space); above 6000 cycles the radiation begins to concentrate into a beam.

The Acoustic Lens



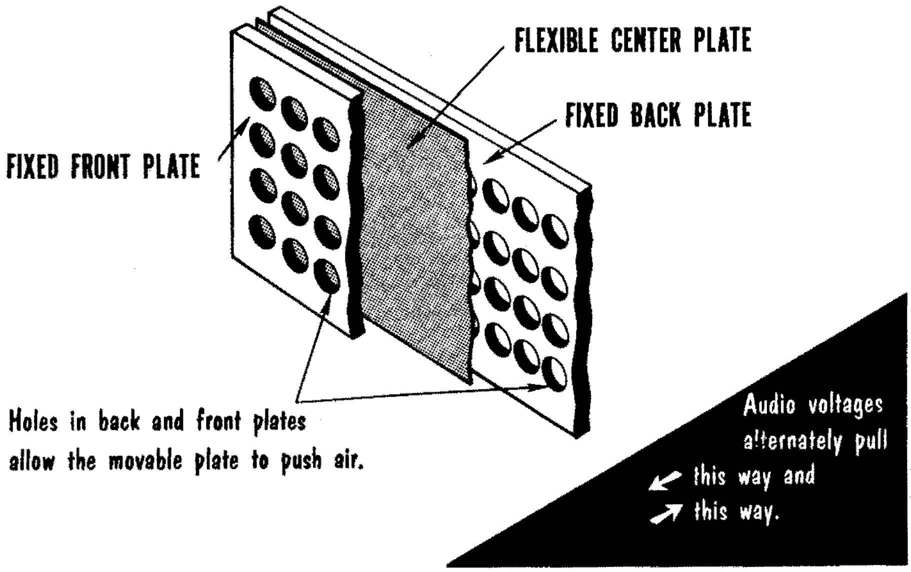
The best solution for wide ranges of frequency is the acoustic lens (a series of perforated or parallel plates). Like its counterpart in waveguide radio, it is effective *above* a certain limit frequency. As the ordinary horn construction, using a tractrix or similar mouth, can be effective *up to* a frequency where the bell ceases to have effect, the lens can be arranged to take over from there and effectively scatter all frequencies above that point.

Another system is based on the fact that high frequencies are subject to reflection by surfaces a wavelength or more in size. This method uses "deflector" plates inside a horn. The bell distributes the lower frequencies, and the deflectors take care of the higher ones.



The Electrostatic Speaker

THE ELECTROSTATIC SPEAKER DIAPHRAGM



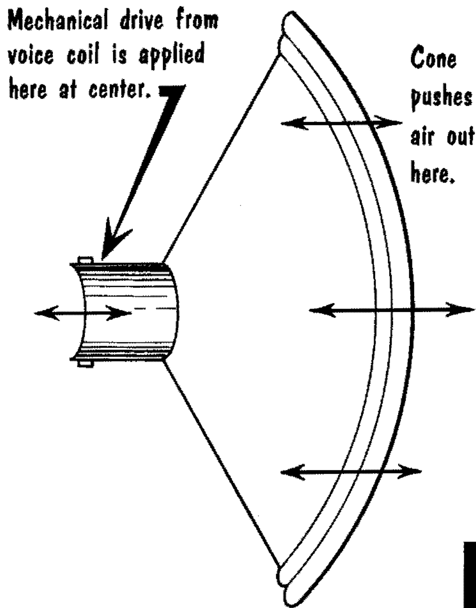
The dynamic or moving-coil transducer has had a very long period of popularity. Much has been done in a great variety of ways to perfect it. The reason for its popularity has been that it has given extremely good results at quite a low cost. For various reasons, however, it is impossible to make it a *perfect* transducer. Therefore some workers in the field have been trying to find a substitute that may do better. One of these is the electrostatic unit.

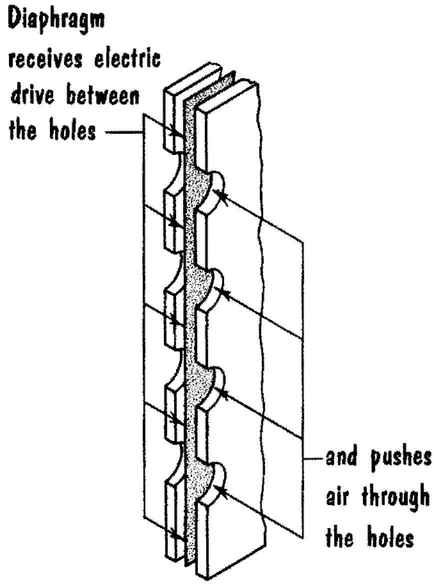
When a voltage is applied between any two surfaces, it sets up an electric field. This field causes attraction to take place between the surfaces. By varying the voltage, the force of attraction is varied. This is the basis of the electrostatic loudspeaker.

This system has its problems—to get a large force from the applied voltage, the spacing between the surfaces must be very small; making the spaces small means the moving surface has little room in which to move. This means that the electrostatic loudspeaker is either limited to use at high frequencies, where only small movement is required, or else a very large diaphragm size is needed to move enough air for the low frequencies. Some have even suggested building the loudspeaker into a room and having it occupy one entire wall. Between these extremes, some quite useful units have been made, which radiate frequencies above about 800 cycles, using a moving coil woofer for the lower frequencies.

The Electrostatic Speaker (contd.)

Unlike the Moving-Coil Speaker's Diaphragm,



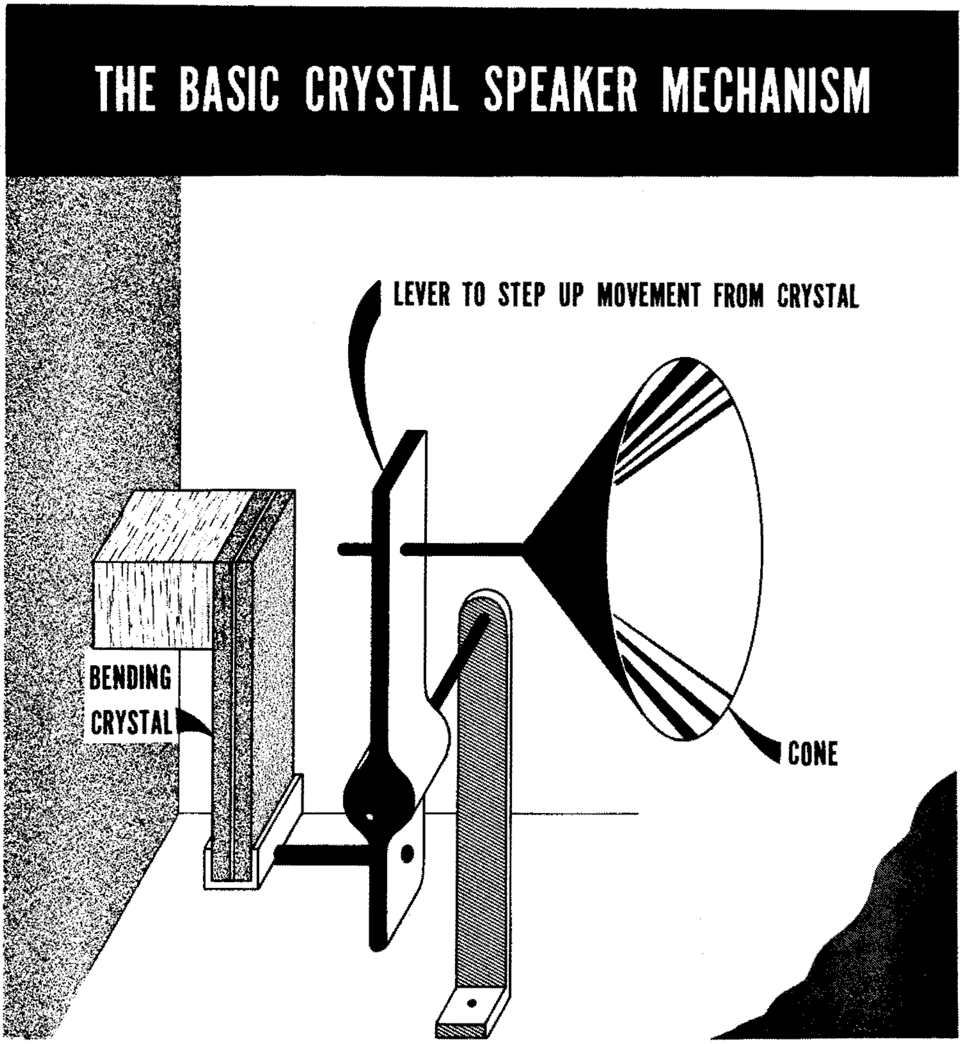


the Electrostatic Speaker's Diaphragm Is Driven Almost Uniformly

The big thing in favor of the electrostatic speaker is the fact that the driving force is distributed over the diaphragm surface. The dynamic loudspeaker is driven from a separate coil. The disadvantage of the coil drive (as any loudspeaker manufacturer knows only too well) is the problem of getting an absolutely rigid mechanical coupling between the drive coil and the diaphragm that is not so heavy that it prevents movement at high frequencies. Choice of material and construction is always a compromise for this reason. The electrostatic units avoid this problem by driving the diaphragm over its entire surface in addition to producing sound from the entire surface. This means that the mechanical properties of the diaphragm (so critical in the moving-coil speaker) become relatively unimportant.

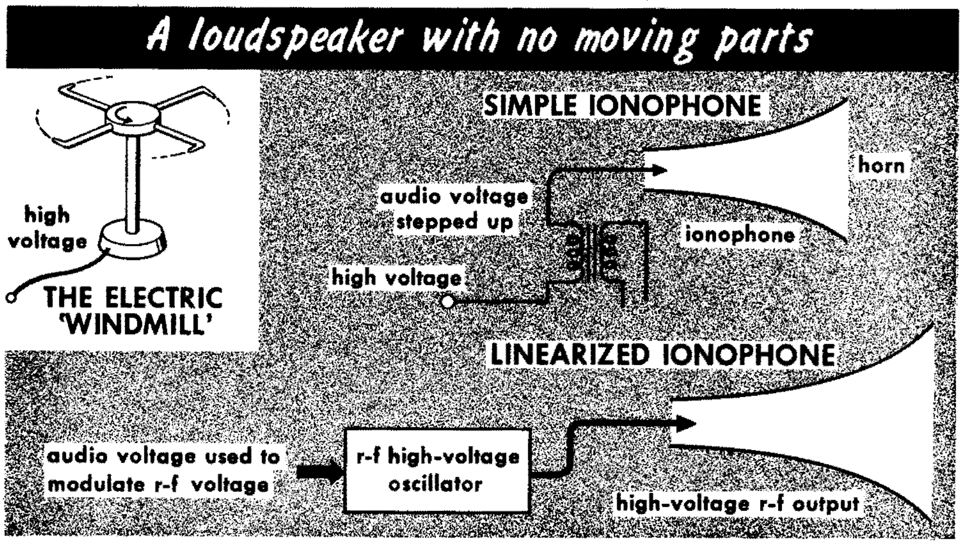
Another problem involved with the use of the electrostatic unit will be better understood in a later section of this course; it concerns matching from the output of the amplifier. The moving-coil speaker is much easier to use from this standpoint.

The Crystal Speaker



The piezoelectric transducer can also be used for loudspeakers. Again it needs the lever principle to get enough movement to the diaphragm. If the moving-coil speaker type suffers due to the mechanical problems in the drive from a simple coil attached directly to the diaphragm, it is obvious that this more complicated drive is going to have problems. Nonetheless, the crystal speaker enjoyed some popularity as a very cheap unit during a period when the moving-coil speaker was relatively expensive. Improved materials and production techniques for moving-coil speakers have reduced their cost and made the crystal-type loudspeaker virtually obsolete.

The Ionophone



With every kind of loudspeaker, there has always been the problem of mechanical parts which can add effects of their own to the sound radiated. It has long been a dream among idealists in high-fidelity circles to find a loudspeaker which had no moving parts, but converts electrical waveforms directly into sound waves. For this reason the *ionophone* has a strong appeal.

Its operation is based on the electric wind principle. When a high voltage is connected to a point, a discharge takes place from the point, in the form of wind. By having the high voltage vary in accordance with the audio waveform, the wind leaving the point will vary in velocity and produce a sound wave to correspond.

The velocity of the wind is not strictly proportional to the applied voltage. Quite a high voltage is needed before the wind starts, and then its velocity rises quickly. A way of avoiding the distortion this would cause is to use high voltage alternating at a radio-frequency and modulate it with the audio.

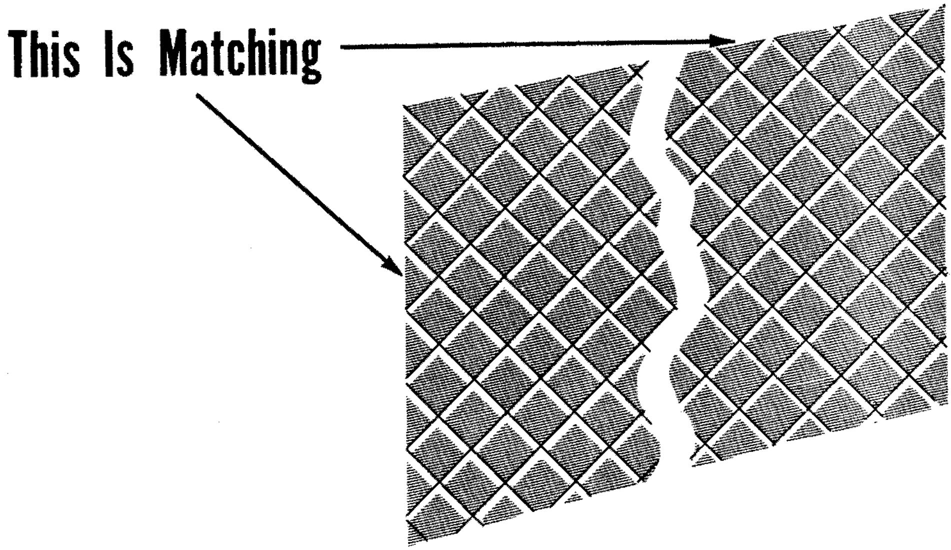
Its popularity has been delayed by the problem of getting enough wind to "generate" the lower frequencies. Many points could be used instead of one, but then there are apt to be acoustic problems as severe as the mechanical ones we want to avoid. Another acute problem is the wind itself. It is much more difficult to get a stream of air quiet enough not to obtrude into the quieter program passages than it is with an electron stream in an electronic tube or transistor.

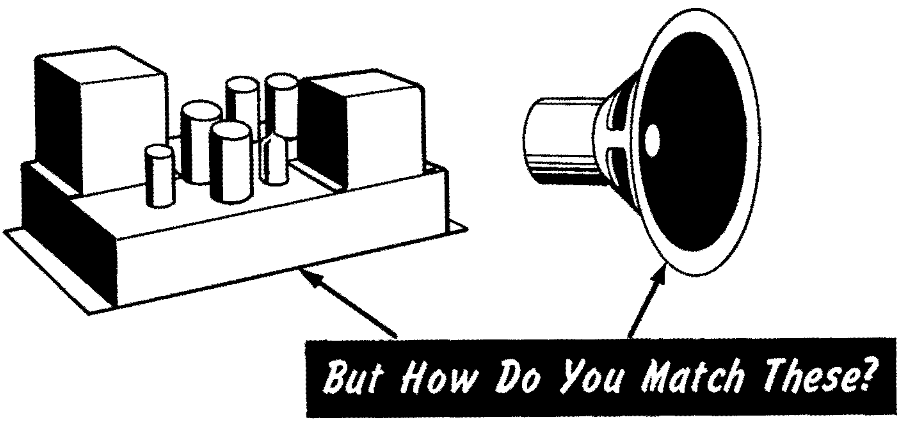
One advantage to the thing, if these difficulties are solved, is that no detector is needed in handling radio signals. The radio signal input itself is simply amplified and fed directly to the ionophone.

QUESTIONS AND PROBLEMS

1. Compare the action of a moving-iron and moving-coil loudspeaker, showing why the latter have come to be almost universally used.
2. Explain the purpose of the cone, coil, spider, and surround of a moving-coil speaker. Why does a moving-coil speaker have to be so much larger than a moving-coil microphone?
3. Why is some sort of baffle needed for a loudspeaker? Is there any alternative to using a baffle?
4. What is an infinite baffle? Explain why it reduces loudspeaker efficiency if made too small.
5. What two facts are utilized in a bass-reflex baffle?
6. Explain how a reflex baffle reverses the phase of the radiation from the back of the speaker diaphragm. Does this happen at all frequencies?
7. How is the size of the port in a bass reflex enclosure determined?
8. Explain why use of a horn improves the efficiency of a loudspeaker. Why is a simple cone not a good shape for a horn?
9. Show by a simple derivation why the exponential form is the best horn shape. What limits the frequency of a horn loudspeaker (a) at the low-frequency end and (b) at the high-frequency end?
10. Assuming two horns that have the same throat diameter (say, ¾ inch), what would be the approximate relationship between their lengths if one works from 100 cycles and the other from 200 cycles?
11. Space can be saved by folding a horn, but what problem does this introduce, and how may it be partially solved?
12. What part of a horn is principally responsible for its directional characteristics in radiation? Give examples and show limitations.
13. Describe briefly (a) the acoustic lens and (b) the acoustic slit, as a means for modifying the radiation properties of a loudspeaker, showing what properties control the frequencies at which their effect becomes active.
14. What are the principal attraction and the main limitation of an electrostatic loudspeaker?
15. What was the principal appeal of the crystal loudspeaker? Why has it become obsolete?

Matching

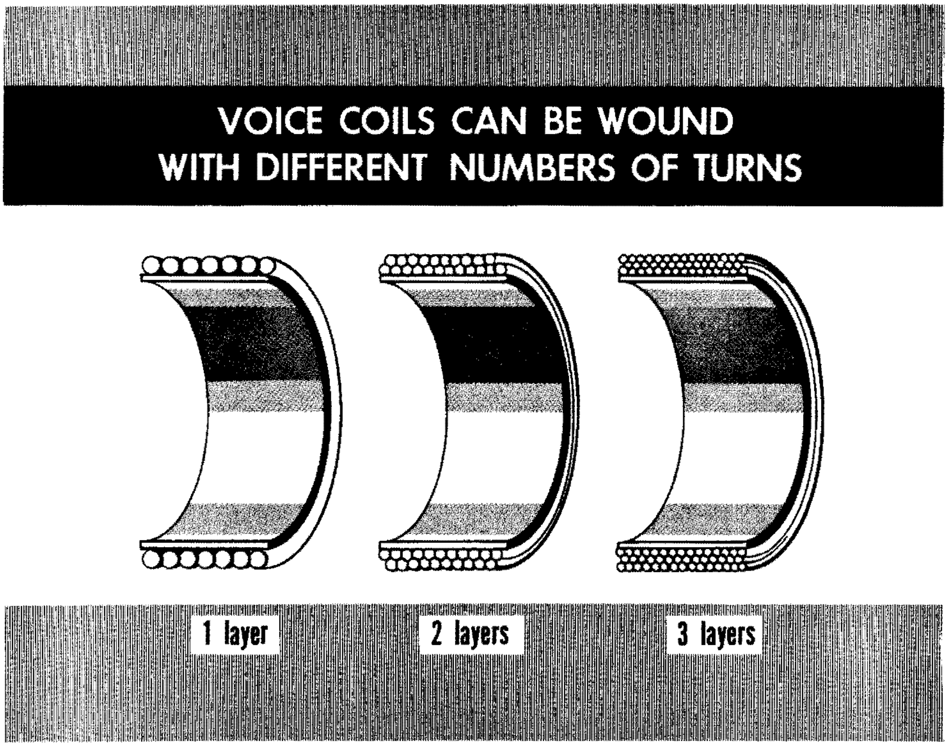




When the word "matching" is used about paint, fabric, or in a repair job, its meaning is obvious; however, when we speak about matching a loudspeaker to an amplifier, the meaning can at first be mystifying! The word is used to refer to electrical properties. An amplifier performs properly when the right kind of electrical impedance is connected to it. If, by some lucky chance, the speaker voice coil impedance is just what the amplifier needs, it is said to *match* it. This does not usually happen so conveniently, and some circuit finagling, known as *matching*, is required to achieve the best performance.

Matching (contd.)

Some variation in the electrical impedance, or resistance, of the loudspeaker can be made by winding the voice coil differently. A single layer of turns may have a resistance of 1 ohm. By halving the diameter of the wire, twice as many turns will go in the same layer length, and we can get two layers in the same thickness. The wire has one quarter the cross section, will be four times as long, giving (4×4) or 16 times the resistance: 16 ohms. Using wire of one-third the diameter originally used, we can use nine times as many turns at 9 times the resistance per turn, which yields ($9 \times 9 \times 1$) or 81 ohms.

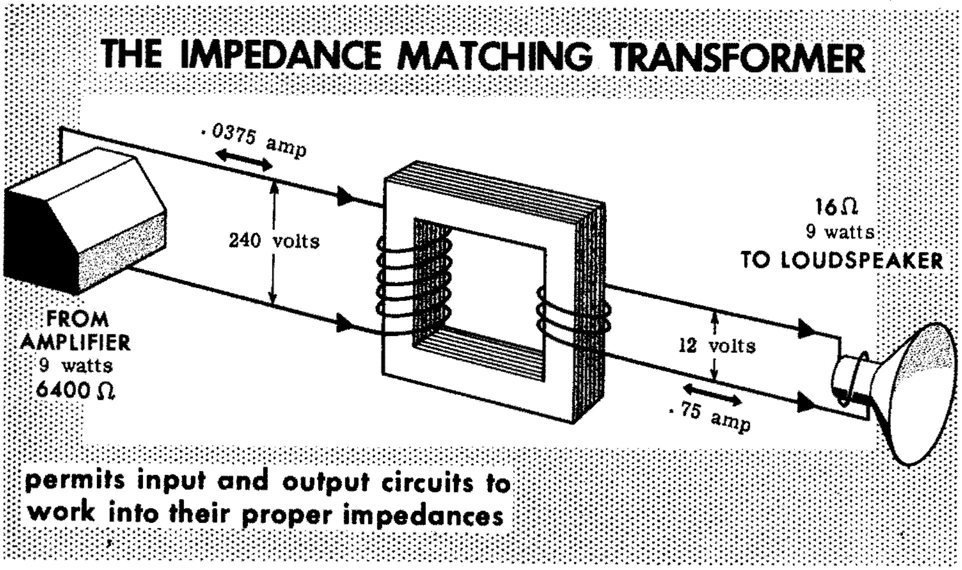


In this way the voice-coil resistance can be changed to some extent by choice of winding. Common commercial values for voice-coil resistance are 1 ohm, 2 ohms, 4 ohms, 8 ohms, and 16 ohms. Some, for special purposes, have a resistance of 45 ohms, and coils have been wound to resistances as high as 500 or 600 ohms. (These coils have a great many turns of extremely fine wire—too fine for a robust coil, and they tend to give trouble in service.)

Even 500 or 600 ohms is rather too low a resistance to match most amplifiers, hence a matching arrangement is still required. Once this is the case, it is as easy to match from 1 ohm as it is from 600 ohms (or any value between, of course), and there is no longer any point in making such fragile coils.

The Matching Transformer

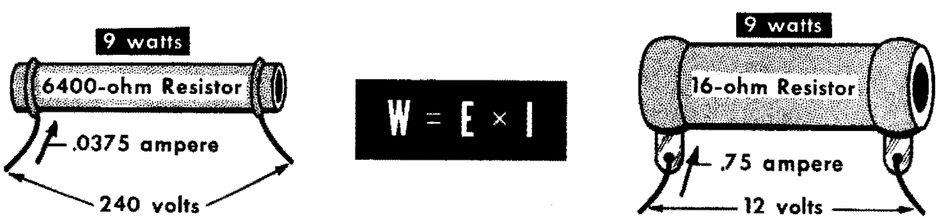
A matching transformer for a loudspeaker is just like one used to step up the output from a microphone, except that it is bigger, and designed to handle much more power. Suppose that a transformer is to match a 16-ohm loudspeaker to an amplifier requiring, not 16 ohms, but 6400 ohms, to which it is to deliver its power.



Suppose 9 watts are to be transferred. The formula for power is $W = EI$, where W is power in watts, E is voltage in volts, and I is current in amperes. (Remember that the relation between voltage and current in a resistance is $E = IR$, where R is the resistance in ohms.)

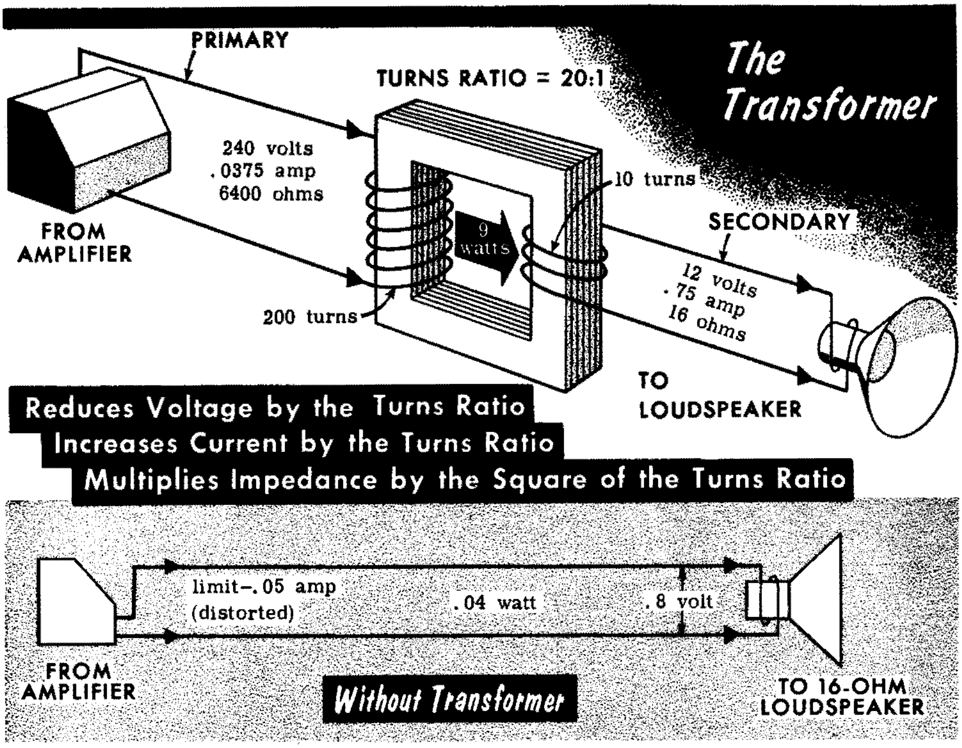
Combining the two formulas, $W = EI = E \times E/R = E^2/R$. Then by multiplying both sides by R, $WR = E^2$. In the example, $WR = 9 \times 6400 = 57{,}600 = E^2$. Hence, $E = 57{,}600 = 240$ volts. Also $I = E/R = 240/6400 = .0375$ ampere.

At the voice-coil resistance, $WR = 9 \times 16 = 144 = E^2$. Hence, $E = \sqrt{144} = 12$ volts. $I = E/R = 12/16 = 0.75$ ampere.



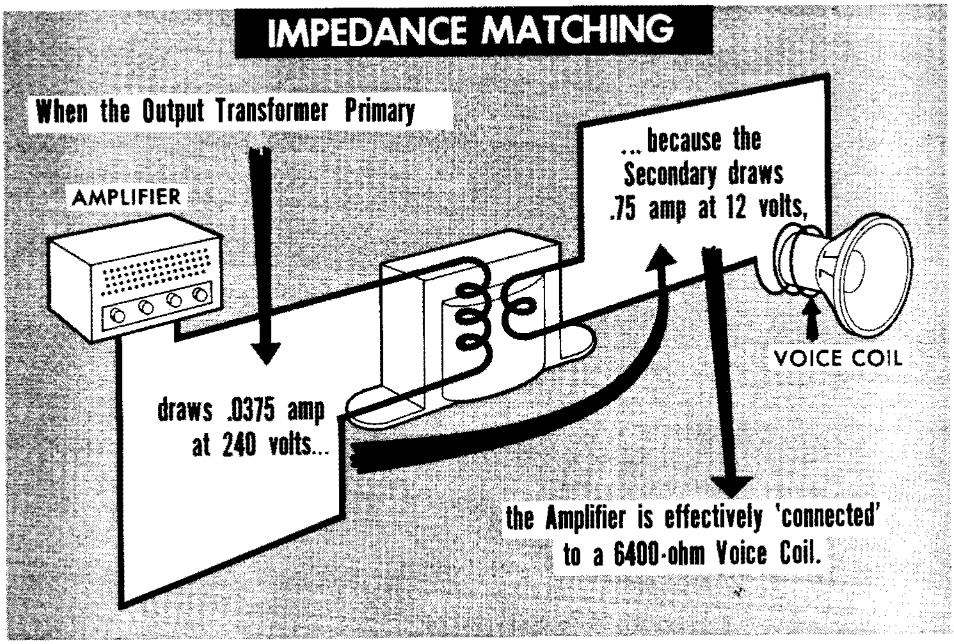
The Matching Transformer (contd.)

But what would happen if a matching transformer were not used? The amplifier is only designed to deliver .0375 ampere into a 6400-ohm resistance or *load* as it is called. If it is connected to 16 ohms, it will probably not deliver much more than .0375 ampere—maybe .05 ampere—and the waveform will be very distorted. When .05 ampere is delivered to 16 ohms, the voltage is only $E = IR = 16 \times .05 = 0.8$ volts, and the power delivered is only $W = EI = 0.8 \times .05 = .04$ watt, in place of the expected 9 watts!



Notice what the transformer does: it reduces the voltage by the ratio of turns in the windings (called the *turns ratio*) and it also increases the current by the same ratio. It thus effectively multiplies the resistance (or impedance) connected to the secondary by the *square* of the turns ratio. In this case, the ratio was 20:1. The impedance connected to the secondary is multiplied by $20 \times 20 = 400$ ($16 \times 400 = 6400$ ohms). The input voltage is 240 volts and the output is 12 volts; the input current is .0375 ampere and the output 0.75 ampere. The input and output power are the same (except for any losses due to the inefficiency of the transformer, which we have conveniently ignored; in practice, an output transformer would be more than 90% efficient, so this is no very great error).

The Matching Transformer (contd.)



When the amplifier supplies 240 volts to the "high" winding of the transformer, the core will be magnetized, and, due to its high inductance, very little current will be drawn from the amplifier, unless the voice coil is connected to the low winding.

The high winding must have 20 times as many turns as the low winding. This way, 240 volts induction in the primary will cause 12 volts in the secondary.

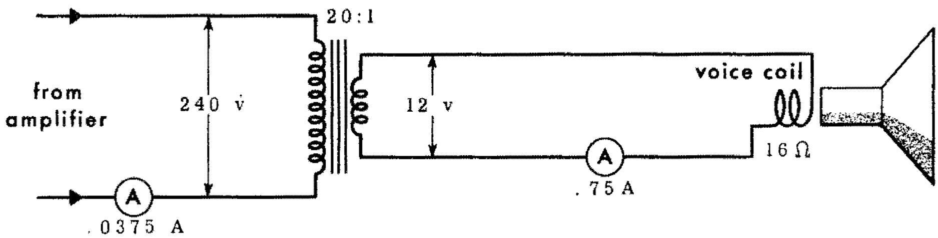
When the voice coil is connected across a 12-volt source, it will draw 0.75 ampere. If no current flowed in the primary of the transformer, this secondary current would destroy the induction by saturating the core, and the 12 volts (as well as the 240 volts) would disappear. To sustain the 12 volts, the amplifier must supply current to the primary to neutralize the effect of the 0.75 ampere in the secondary. As the primary winding has 20 times as many turns, it will only require one-twentieth the current, or .0375 ampere, to have the same effect and neutralize the effect of the secondary current.

Thus the transformer causes the primary winding to take .0375 ampere from the amplifier at 240 volts, when the secondary is connected to a voice coil of 16 ohms that takes 0.75 ampere at 12 volts. To the amplifier, it is the same as connecting a voice coil with a resistance of 6400 ohms, which it "wants." This is matching.

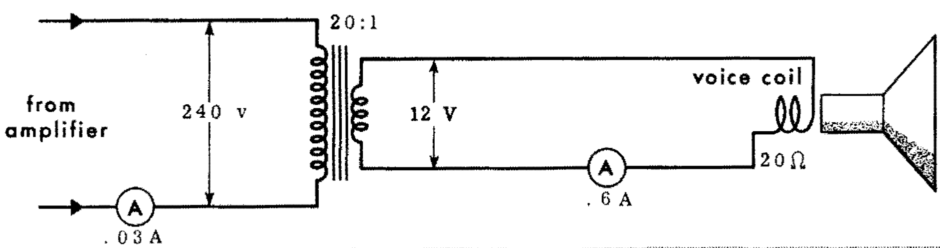
The Matching Transformer (contd.)

What would happen if the voice-coil resistance were 20 ohms instead of 16 ohms? If the transformer secondary voltage were still 12 volts, the voice coil would only take 0.6 ampere in place of 0.75 ampere. The turns ratio would still produce 240 volts across the primary winding, but the primary current required to balance the new secondary current of 0.6 ampere will be 0.6/20, or 0.03 ampere, in place of 0.75/20, or 0.0375 ampere. This is the same as if a load of R = E/I = 240/0.03 = 8000 ohms were connected to the amplifier directly. 8000 ohms is just 400 times the 20 ohms connected to the secondary winding of the transformer. The 20:1 turns ratio of the transformer thus always multiplies the resistance, or impedance, connected to its secondary winding by a factor of 400, or 20 squared (20×20).

Impedance matching permits maximum power output



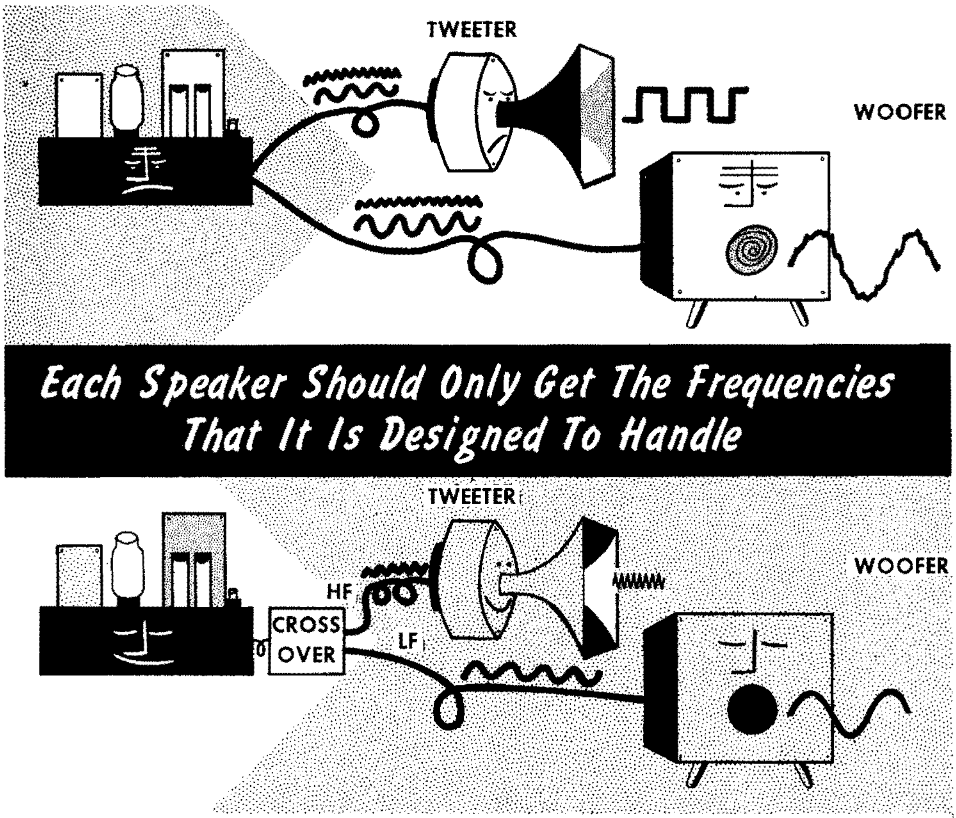
Amplifier is MATCHED to 16-ohm speaker
Amplifier 'sees' correct 6400 ohms
Output power is 9 WATTS



Amplifier is NOT MATCHED to required 16-ohm, but 20-ohm speaker. Amplifier now 'sees' 8000 ohms – Output power is 7.2 Watts

Dividing Networks

When a number of loudspeakers—two or three—are used to cover the frequency range, one handles the low frequencies, another the high frequencies, and sometimes a third one handles the middle frequencies. When the two-speaker idea was introduced, the low-frequency speaker was called a *woofer* and the high-frequency speaker a *tweeter* (for obvious reasons). The addition of a middle-to-upper range unit led to the name *squawker* (a rather doubtful description). Some prefer to call the mid-range speaker (which is still "high" in the musical sense) the tweeter, and the unit that handles the extreme highs is called a *super-tweeter*. In any event, whatever the units are called, each should only get the frequencies that it is supposed to handle.

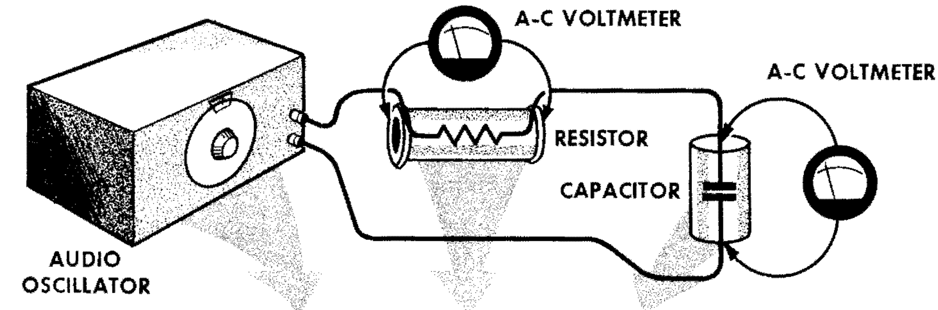


If the low frequencies are fed to a unit not designed to accept them, it will rattle and distort badly. Feeding frequencies higher than a unit is intended to handle will not cause any particular distortion, but will result in loss of power at these frequencies, because the voice coil will accept the power, although incapable of delivering corresponding sound waves. For this reason, *dividing networks* are needed so each unit gets "what is coming to it."

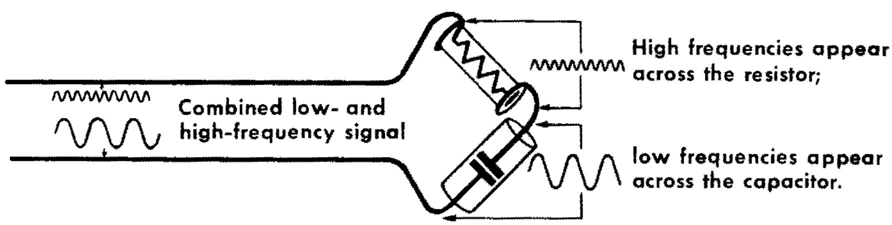
The Series R-C Network

Dividing networks are circuits capable of discriminating (to some extent) between frequencies that are higher or lower than a certain value. Although this may sound quite an "intelligent" action on the part of a mere circuit, there is nothing basically difficult about it.

ACTION OF THE SERIES R-C DIVIDER



Frequency	Voltage Across Resistor	Voltage Across Capacitor
100 cycles	2 volts	10 volts
500 cycles	7 volts	7 volts
2500 cycles	10 volts	2 volts

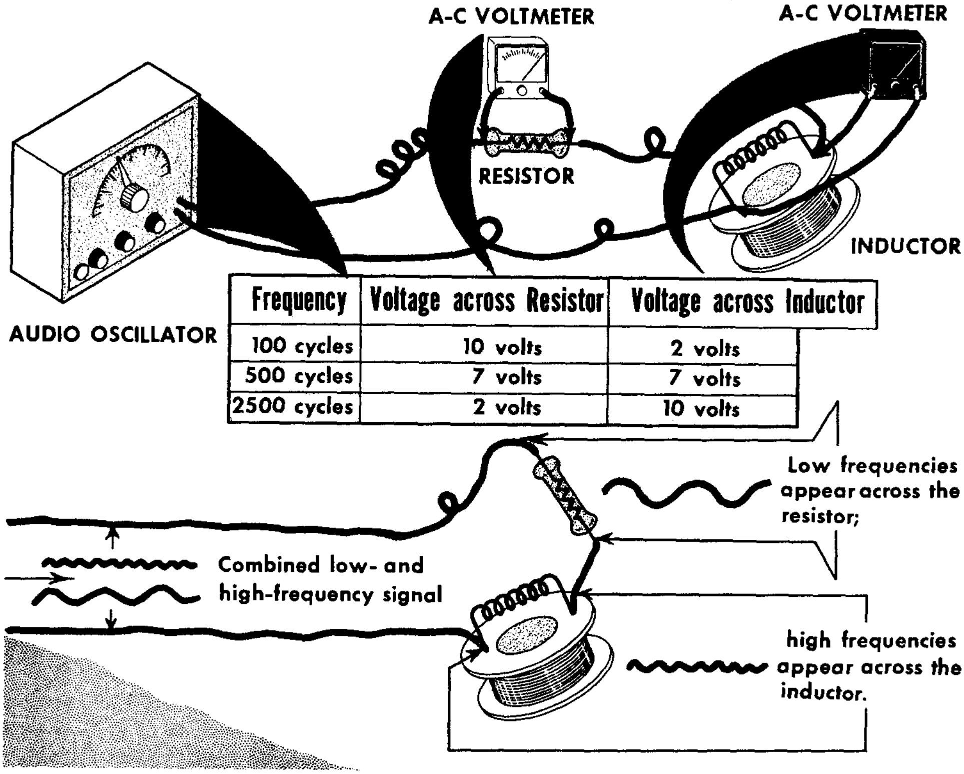


At a low frequency, the charge on the series capacitor follows the fluctuations, and the voltage across it very closely follows the input voltage, only a small current flows, and very little fluctuation appears across the resistor. At a high frequency, the charge on the capacitor does not have time to follow each fluctuation, and the whole of the applied voltage appears across the resistor.

At a medium frequency, some of the voltage appears across the capacitor and some across the resistor due to the current required for charging the capacitor. When a combined program is applied across the resistor-capacitor combination, the low frequencies get concentrated across the capacitor, and the high ones across the resistor.

The Series R-L Network

ACTION OF THE SERIES R-L DIVIDER



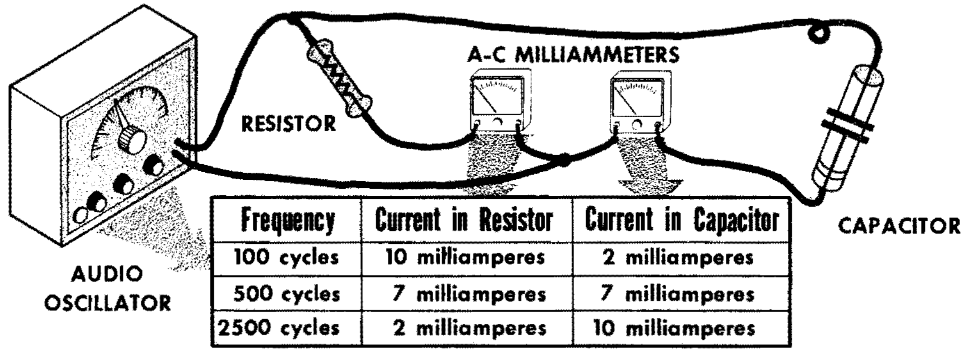
Frequency	Voltage across Resistor	Voltage across Inductor
100 cycles	10 volts	2 volts
500 cycles	7 volts	7 volts
2500 cycles	2 volts	10 volts

A combination of resistance and inductance will also separate frequencies. Whereas a capacitor takes a flow of charge (current $\times$ time) to change the *voltage* across it, an inductor requires that energy be used to change the magnitude of the *current* passing through it. At low frequencies, the current has plenty of time to change, hence the voltage across the inductor does not fluctuate appreciably. (Most of the fluctuation appears across the resistor.) At high frequencies, the rate of fluctuation does not allow time for current in the inductor to change, and the resistor voltage and current are almost constant. (Most of the input voltage appears across the inductor.)

At a middle frequency, some of the voltage appears across both components, because the current changes, requiring a voltage across the inductor to make it do so. When a combined signal is applied across the resistor-inductor combination, the low frequencies get concentrated across the resistor, and the high ones across the inductor. Either of these arrangements would be quite satisfactory to separate *voltages* of different frequencies.

Parallel, R-C, and R-L Networks

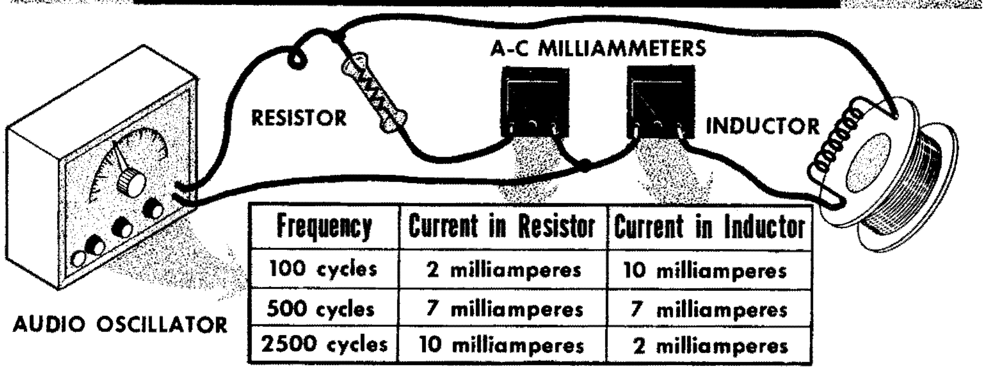
ACTION OF THE PARALLEL R-C DIVIDER



Frequency	Current in Resistor	Current in Capacitor
100 cycles	10 milliamperes	2 milliamperes
500 cycles	7 milliamperes	7 milliamperes
2500 cycles	2 milliamperes	10 milliamperes

Currents of different frequencies can be separated by connecting the same components in parallel. With a resistor and capacitor in parallel, the resistor draws the current at low frequencies, the capacitor at high frequencies, and in the middle range, the current is shared. With a resistor and inductor in parallel, the resistor draws the current at high frequencies, the inductor at low frequencies, and in the middle range, the current is shared.

ACTION OF THE PARALLEL R-L DIVIDER

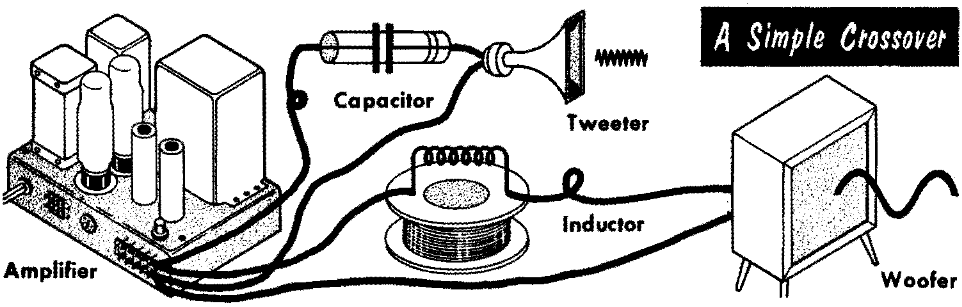


Frequency	Current in Resistor	Current in Inductor
100 cycles	2 milliamperes	10 milliamperes
500 cycles	7 milliamperes	7 milliamperes
2500 cycles	10 milliamperes	2 milliamperes

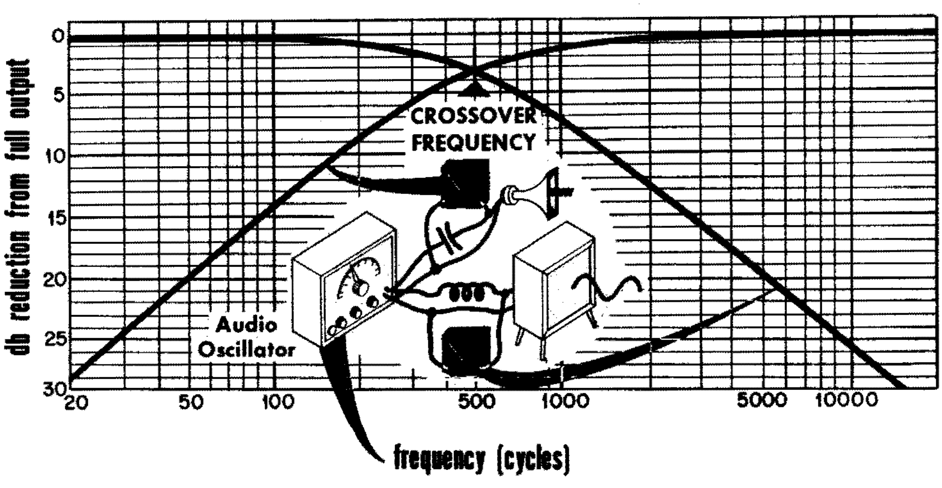
Thus series arrangements separate voltages of different frequencies, and parallel arrangements separate currents of different frequencies. *Power,* however, is absorbed only by resistors, hence all the power supplied to these circuits is dissipated in the resistor, regardless of its frequency. Because loudspeaker units require power rather than merely voltage or current, a simple voltage or current divider of the type that we have discussed thus far will not suffice for sorting out the input to the various loudspeakers in a two- or three-way system.

The Parallel-Fed Crossover

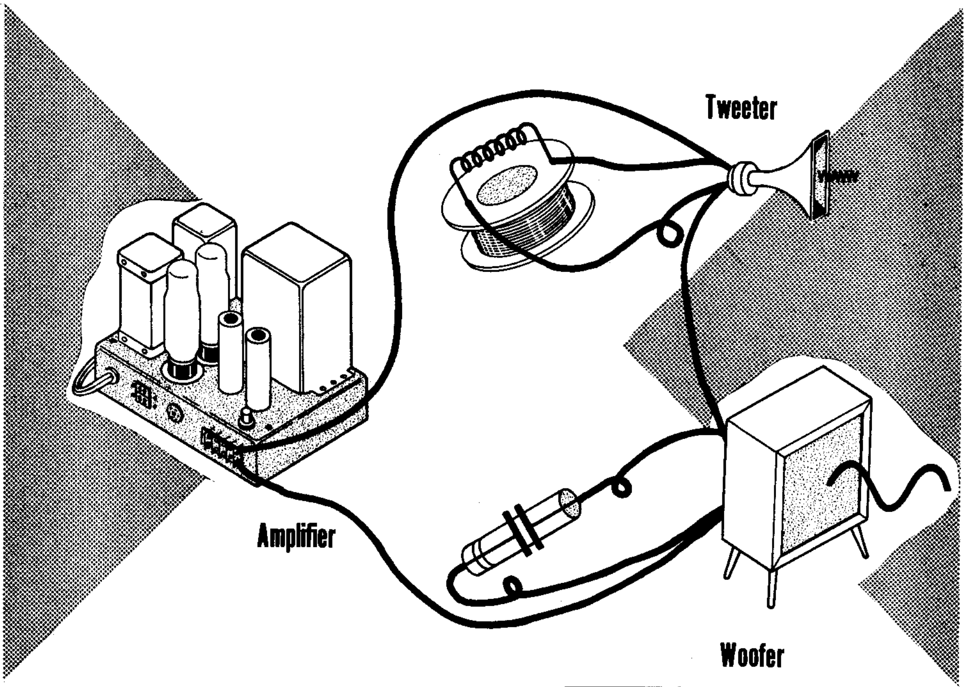
To obtain power division, both inductance and capacitance must be used, for even the simplest network. Here the dividing elements do not absorb any power and all the power goes to the loudspeakers. If a capacitor is connected in series with a speaker, it will block the lower frequencies and allow only the high frequencies to get through to the speaker. The loudspeaker with an inductor in series with it will receive only the lower frequencies, because the inductor will block the higher ones.



If the right values of inductance and capacitance are used, both loudspeakers will get half the power at a middle frequency, called the *crossover point.* At frequencies below this, more of the power goes to the speaker with the series inductance, whereas above the crossover point, progressively more of the power goes to the speaker with the series capacitor. The *current* between the two loudspeakers is divided in this way. Because each branch of the circuit delivers its current to a load resistance (one loudspeaker), the current is accompanied by a proportionate voltage, and a *power* division results. We can also use these components to effect a power division by dividing the voltage.



The Series-Fed Crossover



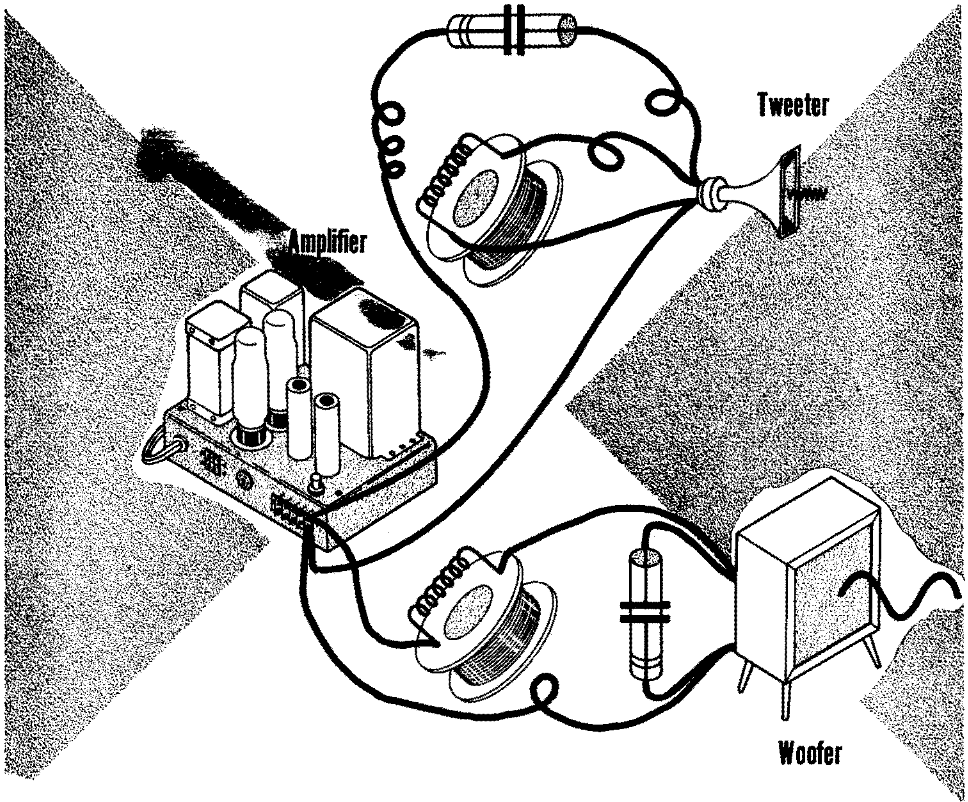
Another Arrangement of the **Simple Crossover**

For a voltage divider, both loudspeakers are in series, but one has a capacitor in parallel with it, and the other has a parallel-connected inductor. At low frequencies, the inductor bypasses the loudspeaker across which it is connected, so that the other one gets all the low frequencies. At high frequencies, the capacitor bypasses its loudspeaker, and the other speaker gets them. As before, at a middle frequency, the energy is equally divided between both units. The response would be exactly as plotted for the other arrangement. The only differences are the series or parallel connection of the speakers with respect to the amplifier output and the connection of the reactive elements. (In the series arrangement, the capacitor is connected to the woofer and the inductor to the tweeter; in the parallel arrangement, this association is reversed.)

Although the circuits divide power, each of them resembles a simple divider, one for current, the other for voltage. The power delivered to each of the speakers is proportional to this voltage or current division. It is possible to get better frequency separation by using more components in the networks.

Two-Element Crossovers

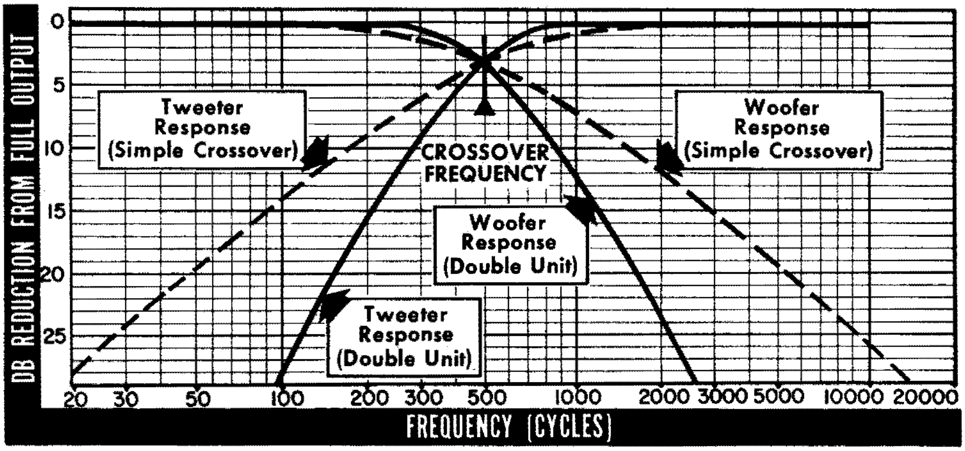
This arrangement works by progressive current division that accentuates the separation. The first division is by series components, connected to the output in parallel. Then each unit has a further component in parallel with it to bypass the unwanted frequencies.



More Complicated Crossover Using Two Capacitors And Two Inductors

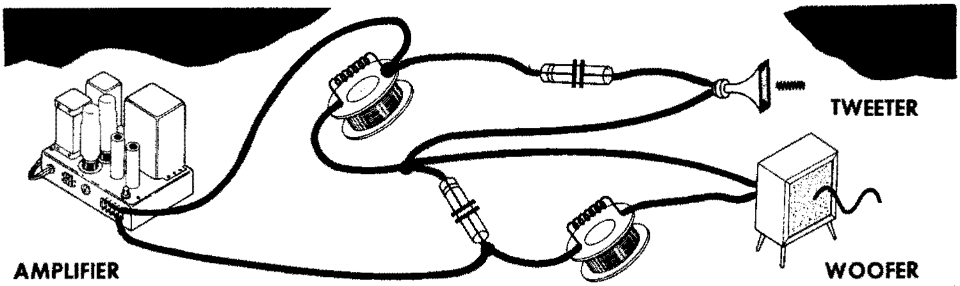
The low-frequency loudspeaker is bypassed at higher frequencies by the capacitor connected across it, then the inductor in series with it further blocks higher frequency voltage. However, the action is a little more complicated than this. The inductance and capacitance interact, rather like a tuned circuit in a radio, with the result that the energy is delivered to the low-frequency unit more efficiently at frequencies near to cutoff (the crossover point). This can be shown by plotting the frequency responses and comparing them with the simple circuits.

Two-Element Crossovers (contd.)



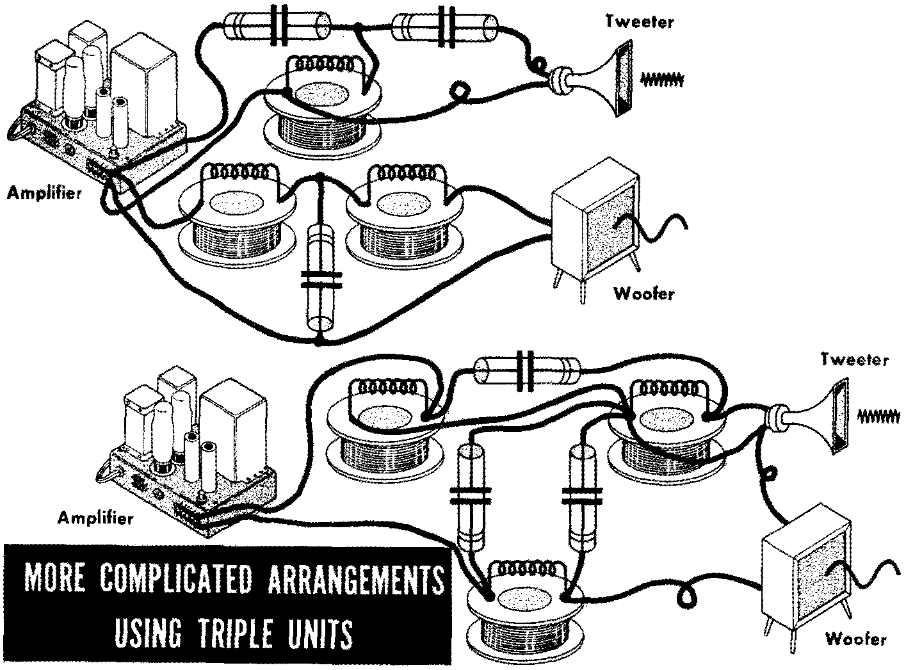
Notice that below the crossover point the low-frequency unit is fed more efficiently with the improved network, and that the response falls off more rapidly above crossover, due to the double action. The same thing happens with the feed to the high-frequency unit: the inductor in parallel with it bypasses the low-frequency currents, and the capacitor in series blocks the low-frequency voltage. The interaction again works like a tuned circuit to improve the efficiency just above crossover point, and to make the response fall off more rapidly below it. The first pair of components divide current, as in the simple current-dividing network, and the component connected in parallel with each unit accentuates the action by redividing the current.

A similar development can be made, using progressive voltage division. Here the voltage-blocking components are in series with each unit, then the current bypass components are in series across the amplifier output to cause voltage redivision. The values of the components are different from those in the other circuit, but the results are the same.

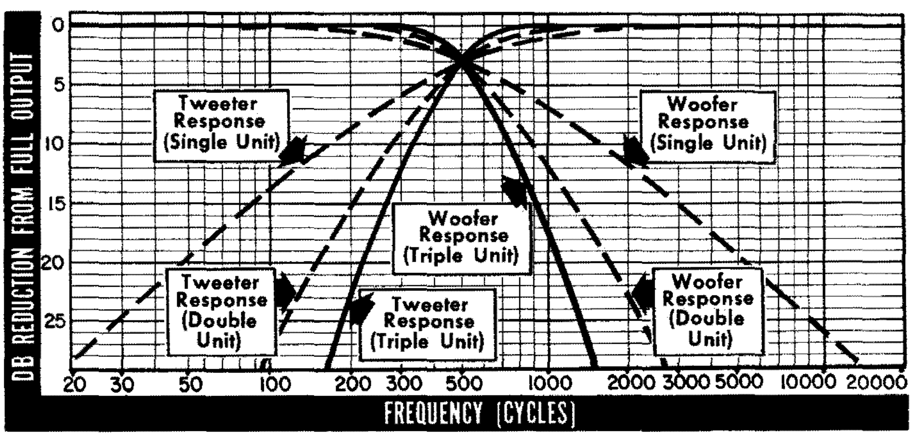


ANOTHER ARRANGEMENT USING TWO CAPACITORS AND TWO INDUCTORS

Three-Element Crossovers

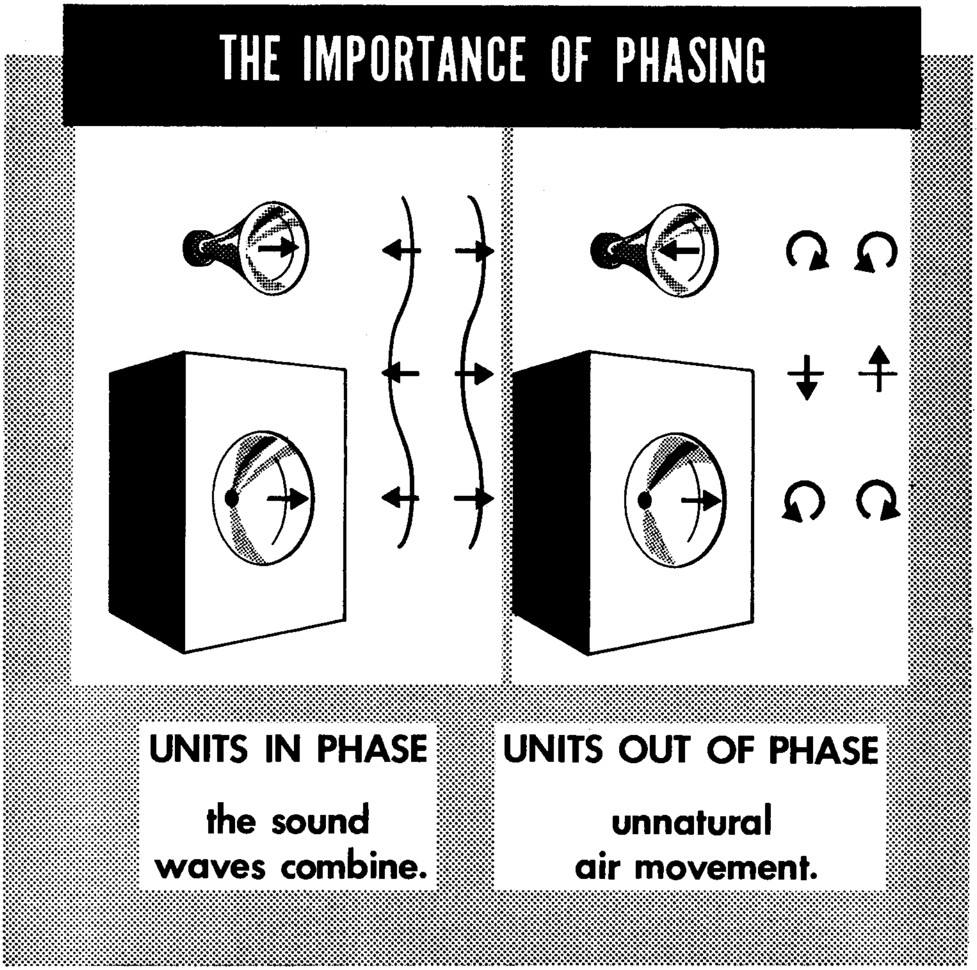


The process of adding components may be continued until there are as many elements or components as desired, although three, as shown here, is usually the limit. Using a greater number of components leads to complications in design, if the response and frequency division are to be correct. The more components used, the more effectively the frequencies are separated; frequencies near crossover point are transmitted more efficiently to the correct unit, and those beyond the crossover point are rejected better.



Phasing

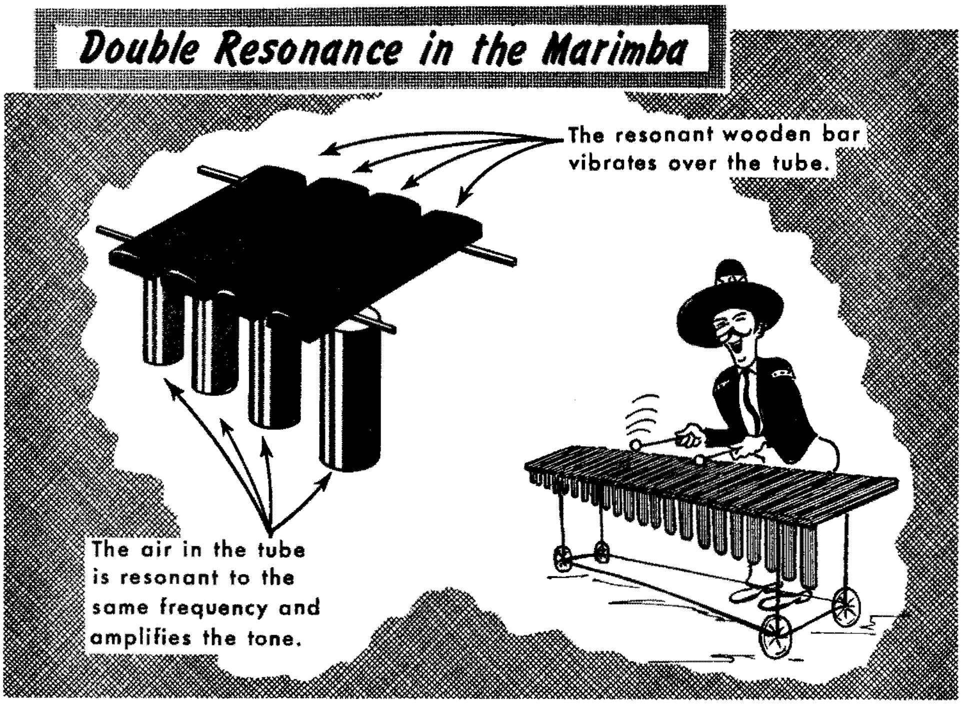
An important aspect of connecting these networks is called *phasing*. At the one frequency—crossover point—both loudspeaker units are delivering about an equal amount of the total sound energy. It is important that both diaphragms should "push" and "pull" together in *phase*. If one speaker pushes when the other pulls, and vice versa, they are said to be *out of phase*.



When the units work correctly (in phase), the combined sound wave builds up into the same kind as would be produced by one big loudspeaker unit. When they are out of phase, however, the motion of the air particles is crossways. This makes quite an unnatural kind of sound, and the realism of the reproduction is destroyed. For this reason care should be taken to connect the units so that they are in phase.

Resonance

The property of resonance plays an important part in audio. For music, it is a property much sought after, but for musical *reproduction,* it is something to be avoided. Every musical instrument has its own variety of resonator. In stringed instruments the strings vibrate at a resonant frequency that fixes the note played, and the resonances in the body of the instrument control the overtone structure. Pipes and horns use a resonant column of air, in which the weight and elasticity of the air particles themselves (and the dimensions of the pipe or tube) control the frequency of vibration. Reed organs use vibrating reeds where the natural frequency of the tongue fixes the tone with resonators that give the desired timbre.

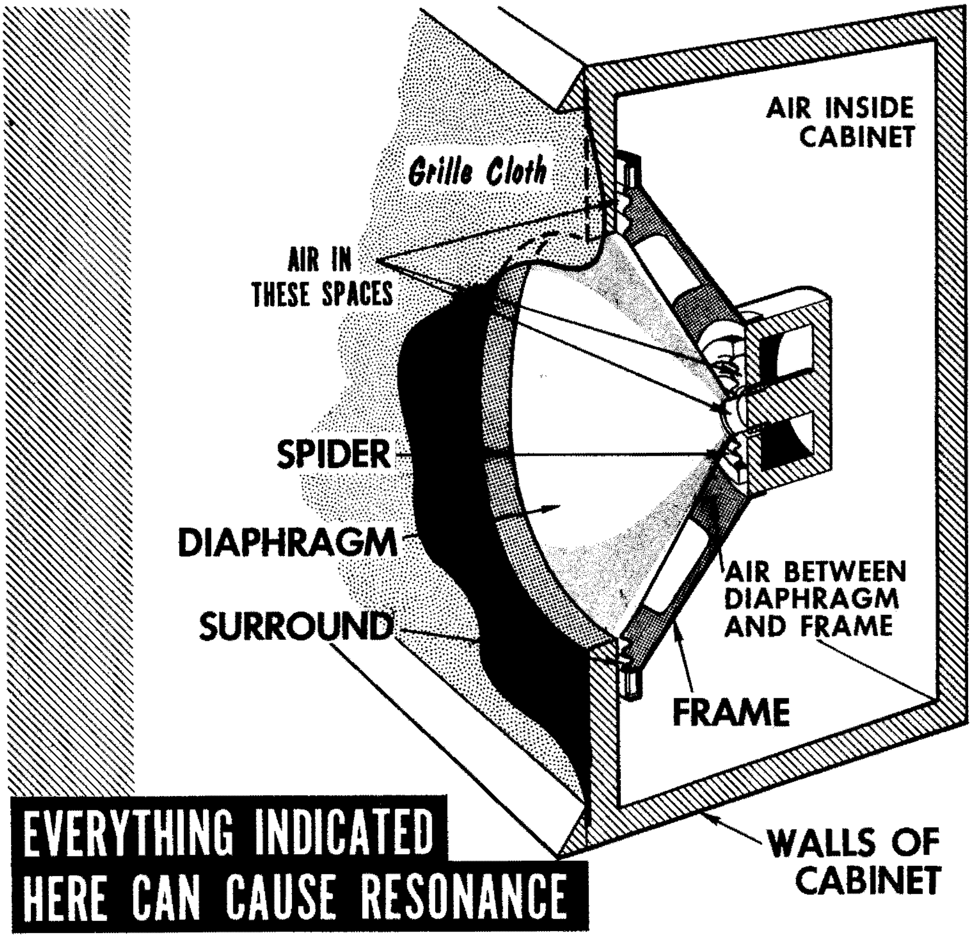


With all these musical instruments, it is a family of resonance patterns that control the production of tones peculiar to that instrument. In the xylophone or marimba, a vibrating piece of wood initiates the tone, which is reinforced and given its fullness of tone by a resonant tube, in which an air column is set in vibration by the feeble vibrations from the wood bar.

In audio, however, resonance is taboo. We want to hear the original instruments' tones and timbres, not some extra effects contributed by the audio system. Unfortunately, it is not possible completely to eliminate resonance from anything with moving parts, such as a microphone, loudspeaker, or phonograph pickup.

Resonance (contd.)

A telephone diaphragm is a simple steel disk of metal that vibrates according to the currents in the earpiece. It also has its own simple resonance (much like a miniature drum). This tone tends to predominate, giving the quality of speech peculiar to telephone reception. Choice of a suitable resonant frequency (in the construction of the instrument), makes for maximum intelligibility on speech, but music does not sound very good over a telephone.



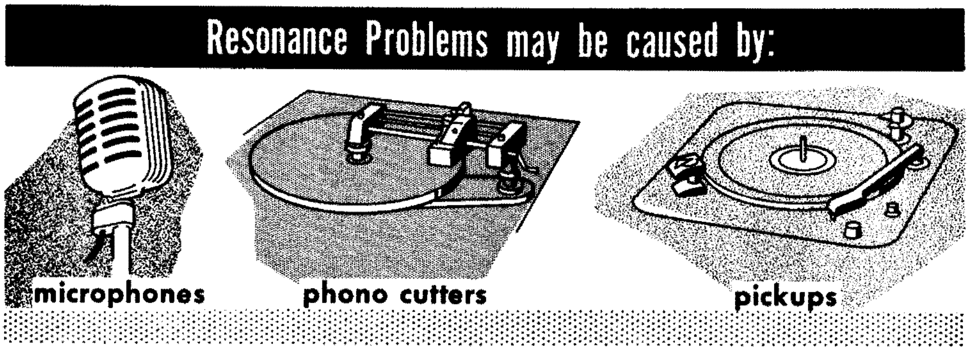
In loudspeaker design, every effort is made to minimize resonance effects, but there are still plenty left. The best that can be done in a good design is to spread the resonances so they all have as little effect as possible. Some manufacturers, from time to time, make the claim of having produced the perfect loudspeaker, with no "tone" of its own. So far, none of these claims has been true, and future claimants should be checked very carefully.

Resonance (contd.)

The loudspeaker is not the only part of an audio system that is subject to resonance problems. They afflict microphones, phono cutters, pickups, and even some electrical circuits. As a result, it is difficult to be certain that the very simple statements that we have made are readily true.

We said that a microphone converts acoustic vibrations into electrical waves. It does. But how *faithfully* does it do so? Are the electrical waves an accurate copy of the acoustic vibrations? This is very difficult to be sure of, because resonances in almost any microphone will augment some frequencies and diminish others. We know this happens, but by how much?

Determining this is complicated by the fact that rooms in which we try to measure things also have a resonance of their own—standing waves have the same effect as resonances in microphones or loudspeakers; they emphasize some frequencies, and by contrast they diminish others. Thus even if we know about the microphone, how much of the acoustic vibration picked up is due to the original sound wave and how much is augmentation due to standing waves?



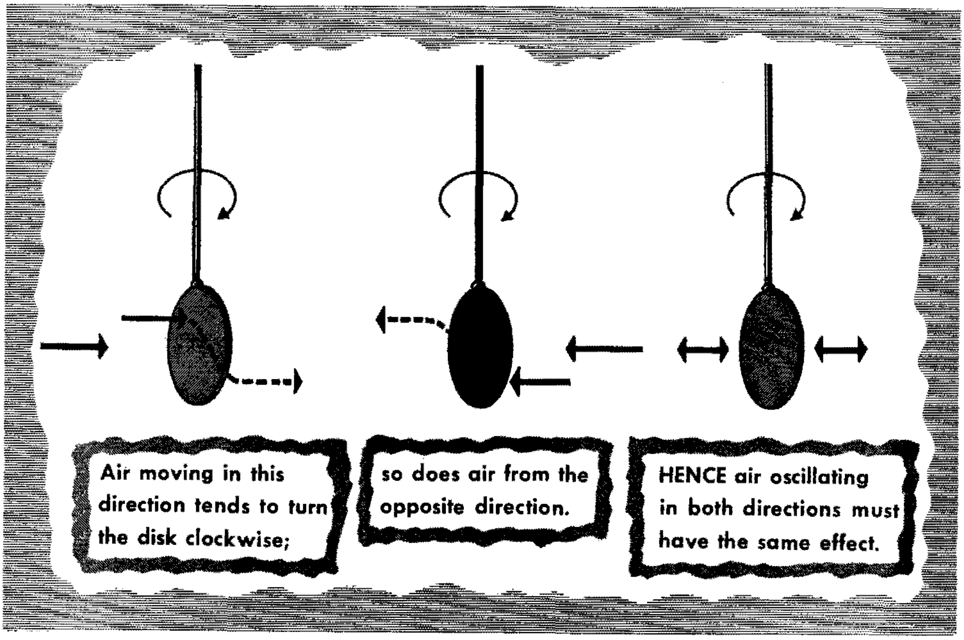
In view of the extreme difficulty in getting rid of undesirable resonances, it is difficult to make measurements against an *absolute* standard. It is relatively easy to compare two pieces of equipment, either with instruments or by listening. When you know how much they differ, the question remains as to which one is nearest right. Listening can offer a guess, but different people's guesses do not always agree.

The thing we need to establish is a *standard sound field* as a basis for comparison. This means an area in which the intensity of sound vibration itself is accurately known. Anything that involves the use of mechanical movement to measure movement of air particles brings in the problem of resonance, which will make the method give different answers at different frequencies. For this reason, *standards* of wave measurement use devices that do not depend on transmitting the movement to mechanical parts to measure it.

The Rayleigh Disk

The principal difficulty in measuring the intensity arises due to the very small amount of energy in a sound wave. One method, discovered by Lord Rayleigh, uses the *Rayleigh disk.* It is still used today.

The Action of the Rayleigh Disk

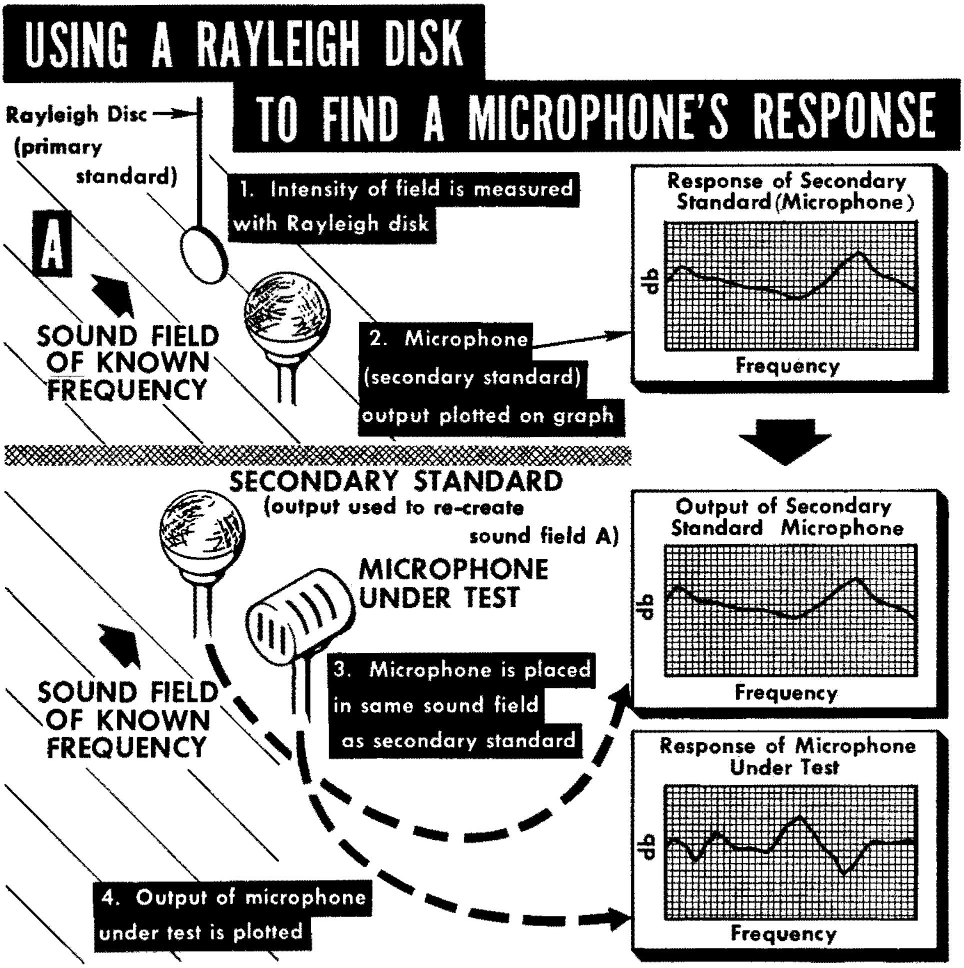


When a small disk is suspended in air, any air movement striking it tends to set the disk "head on" to stop the air. This happens whichever side the movement comes from. In a sound wave, the movement comes alternately from opposite sides, as air particles move back and forth. Both movements tend to turn the disk the same way, hence the torsion on the disk will depend on the *amount* of air particle movement, not its direction or frequency.

To use this delicate instrument, the disk is carefully hung at a definite angle to the approach of the sound wave to be measured, and shielded from any other air currents. Then the thread by which it is suspended is twisted by a calibrated instrument, to see how much torsion is needed to maintain the same angular position against the force caused by the wave. This information can then be used to calculate the air particle movement in the wave by measuring the torsion on the disk when it is held in a steady stream of air that is moving at carefully controlled velocity.

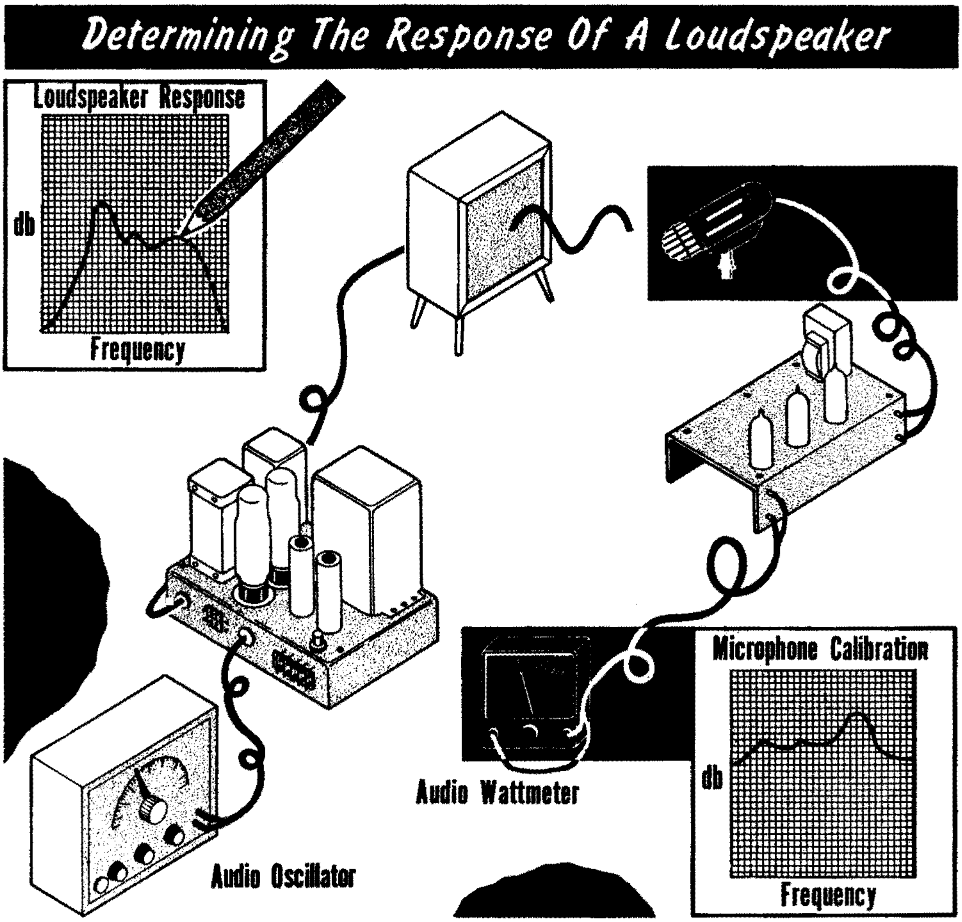
Calibrating a Microphone

From the measured properties of air, the pressure fluctuations in a wave can be calculated, once the particle movement is known. This makes it possible to "calibrate" a microphone of any reliable type by placing it in this known sound field and measuring its electrical output with suitable instruments.



In this way we get a calibrated microphone—one with which we can measure a sound field, by making measurements and referring to its calibration chart. From this we can measure the performance of any other item of audio equipment. This calibrated microphone is called a *"secondary-standard."* Another microphone in the same sound field can be compared against the secondary standard, and an *absolute* response obtained for it by using the secondary standard's calibration chart.

Microphone and Speaker Response Curves



The response of a loudspeaker can also be found by using a calibrated microphone to measure the sound field that it radiates at different frequencies. For all these purposes, it is obvious that the calibrated, or secondary standard microphone should be as near perfect in response as possible, so that little correction will be necessary from its chart. High-quality condenser microphones and a special form of crystal microphone without any lever system, are favored for this purpose.

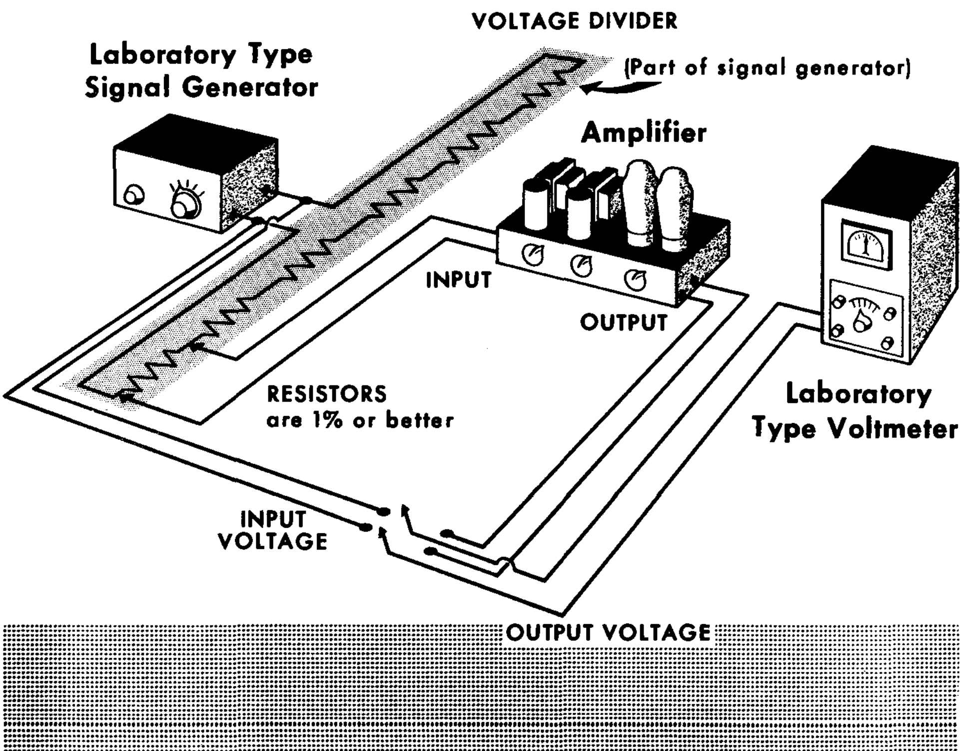
All this kind of work is very informative in finding out how good equipment *is*, but we are left with the question of how good it *should be.* To know this, we want to know something about people's hearing faculties. After all, the main purpose of audio equipment is to produce a satisfactory listening illusion.

Calibrated Amplifier

But how do you measure the output from a microphone, when this is only a few microvolts. Most voltmeters will not read anything smaller than a few volts. The voltage from the microphone needs amplification. This means we need an amplifier whose amplification is accurately known.

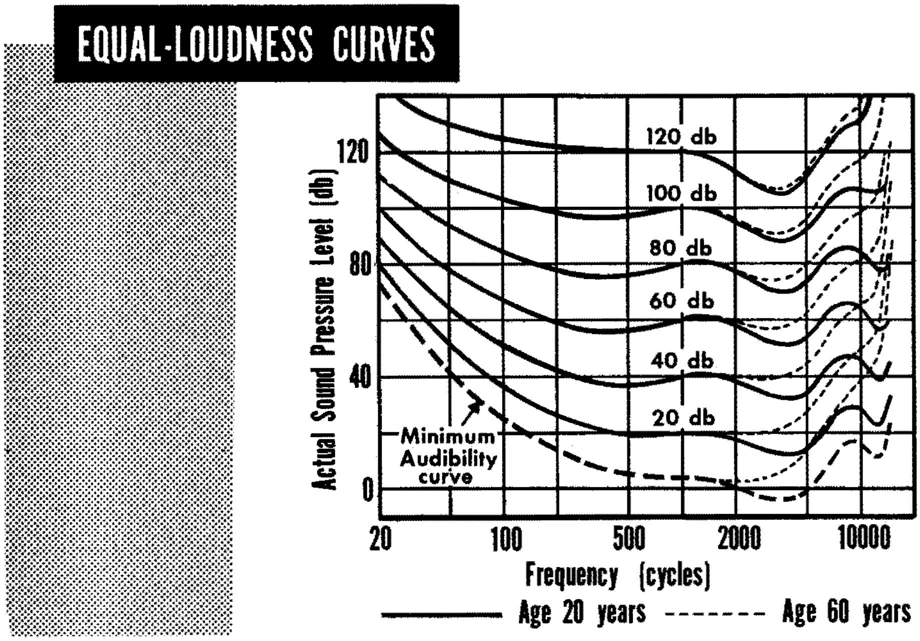
Such an amplifier is called a *calibrated amplifier*. To calibrate an amplifier we need to apply very carefully measured input voltages and then measure the output voltages. To get the known input voltage, this must be big enough to measure directly, so a voltage about the same as the amplifier output is used. Then it is passed through a resistance divider arrangement called an *attenuator*. The values of these resistors are accurately measured, so that the voltage applied to the amplifier input will be a very accurate fraction of the larger input voltage.

Calibrating an Amplifier



The Ear's Response Curve

To make these basic hearing tests, a number of people were used as subjects. Each person was asked to tell when tones of different frequency sounded equally loud. Using a frequency of 1000 cycles as a reference, the response of individual ears could be plotted in this way for different volume levels. When this was done, an average was taken for a large number of people, and the resulting curves are shown here. These curves are not a close representation of the response of your ears, nor indeed of any one person's. Careful analysis of the results showed that hardly anyone has an "average" pair of ears. Individual hearing may differ widely (by 10 decibels or more) from the average response.



This fact shows us that "taste" in music or quality is not solely responsible for differences in opinion about audio equipment—individual people's ears give different impressions for them to judge by. Added to this difference is the fact that we do not hear a musical program as a number of groups of frequencies. In the composite of frequencies presented, our ears have the ability to recognize individual instruments in an orchestra (if the reproduction is good) even though the frequencies each instrument uses overlap.

These facts can readily be recognized by anyone with a little listening experience. They are mentioned here to complete the picture of what comprises "Audio." The explanation must come later, and will appear in one of the later volumes in this course.

QUESTIONS AND PROBLEMS

1. What are the natural limitations to obtaining a voice coil of specific resistance? If a voice coil, wound in 4 layers, has a resistance of 10 ohms, what would be the approximate resistance of a coil of the same dimensions, wound of a wire whose gage would permit 5 layers?

2. What is matching (a) as applied to microphones, (b) as applied to loudspeakers? If a loudspeaker has a voice-coil resistance of 8 ohms and an amplifier output needs a load of 5000 ohms, what transformer turns ratio is needed?

3. If the voice-coil resistance in Question 2 were 10 ohms instead of 8 ohms, what load would be provided for the amplifier by the same transformer?

4. What led to the use of multiple loudspeaker units to cover the audio range?

5. Why are dividing networks used with multiple loudspeaker systems? What might result from failure to use a suitable network?

6. Networks consisting of resistance and capacitance or resistance and inductance produce division of voltage or current according to frequency. Why cannot networks of these types be used to feed power to loudspeakers?

7. What is basic difference between networks in which the elements are (a) in series, (b) in parallel?

8. In the simplest type crossover, suppose the crossover frequency is 800 cycles. How much of the total amplifier power of 10 watts goes to each unit, when the frequency is (a) 400 cycles, (b) 800 cycles, (c) 1600 cycles?

9. What is the effect of using more inductors and capacitors in crossover networks? Explain how this difference in response is achieved.

10. What is meant by phasing? How would you tell if two loudspeaker units were connected to provide correct phasing?

11. What is resonance? State in which of the following it is a desirable feature, and in which it is undesirable: piano, microphone, violin, marimba, loudspeaker, organ pipe, flute, oboe, phonograph pickup.

12. A moving-iron loudspeaker, when compared with a moving-coil type, actually seemed louder, although its quality was poor, particularly in the bass. Explain this with reference to the effect of resonance.

13. Explain how a simple device can be used to obtain calibration of the particle velocity of air in sound wave. Does a Rayleigh disk discriminate between waves from the back or front? Is it uniformly sensitive to sound waves from all directions?

14. Is human hearing uniformly sensitive to all audible frequencies, and does its sensitivity to different frequencies depend on their intensity?

INDEX TO VOL. 1

INDEX

basic audio

by NORMAN H. CROWHURST

VOL. 2

JOHN F. RIDER PUBLISHER, INC.
116 West 14th Street • New York 11, N. Y.

Library of Congress Catalog Card Number 58-13425

Printed in the United States of America

PREFACE

Audio is like Topsy: it wasn't born, it just growed. Whatever Topsy may have been like, Audio has grown like a gawky child—not always in proportion! Originally associated with radio and later with high fidelity, audio now finds application in many other places—to name a few: computors, automation, ballistics and guidance for missiles, sonar detection for navigation, ultra- and infra-sonics for medicine, both diagnostic and therapeutic, as well as geophysical and other work. In fact Audio is now one of the largest and most basic divisions of electronics.

Courses in audio were nonexistent not too many years ago. Since then, textbooks and courses have appeared. But their approach follows the principle of many professors: "I learned it the hard way—you'll have to!" It's like learning watchmaking from a bridge-building man.

My wide experience in various aspects of audio has shown the need for a better way. In industry, in academic education, and particularly in working with graduates from college, this need is evident. My extensive technical writing for magazines and consultant work in the industry have also shown me audio's educational needs.

Many competent "practical men" find themselves hindered by lack of academic background in the subject. They can do their job in their own established "groove." But they do not have—and find it impossible to acquire—the background to enable them to expand outside this groove. These people need help in closing the gap between "theory" and "practice."

Engineers are conversant with the accepted "technical language," but they read the literature with only an "intuitive comprehension" (or should it be apprehension?). Their education dragged them past many "awkward spots" about which they have never felt really "comfortable." Like the King of Siam in "The King and I," they find many facts of which they wish they were more certain they are sure.

Very important are the new students, technicians, and audiophiles. They will need a basic education in audio to enable them to add their contribution to progress (and to earn themselves a living!). Why make it difficult? They'll do much better if they can get a good start.

All-in-all, it is time that certain roundabout approaches to this key subject were eliminated. We need a direct, meaningful way to take the place of the difficult detours. Then each of our three groups can not only "learn audio," but also understand it! This three-volume book results from the author's extensive education research. The finished arrangement achieves a completely new directness.

Let me give just one example: how many understand the behavior of a coupling capacitor, particularly its contribution to amplifier transient performance, and what sometimes happens to feedback? This has always been based on the concept of capacitive reactance, which does not adequately *explain* all the effects. We have adopted a practical "what happens?" approach.

As a result, someone who learned this the old way may miss the familiar landmark of the reactance concept—when he expects it; a closer examination will reveal the reason for postponing it: the whole presentation has been arranged to avoid the "dead spots" left by the traditional approach.

Inevitably such a change of approach will mean a change of stress. I make no apology for this. I know from practice that it is far more successful in getting Basic Audio "across."

It would be impossible to acknowledge the very many who have, knowingly or unknowingly, contributed to my experience, making this book possible. But I would like to express my thanks to the John F. Rider staff for their cooperation in "packaging" it in a form that interprets my intentions so well.

NORMAN H. CROWHURST

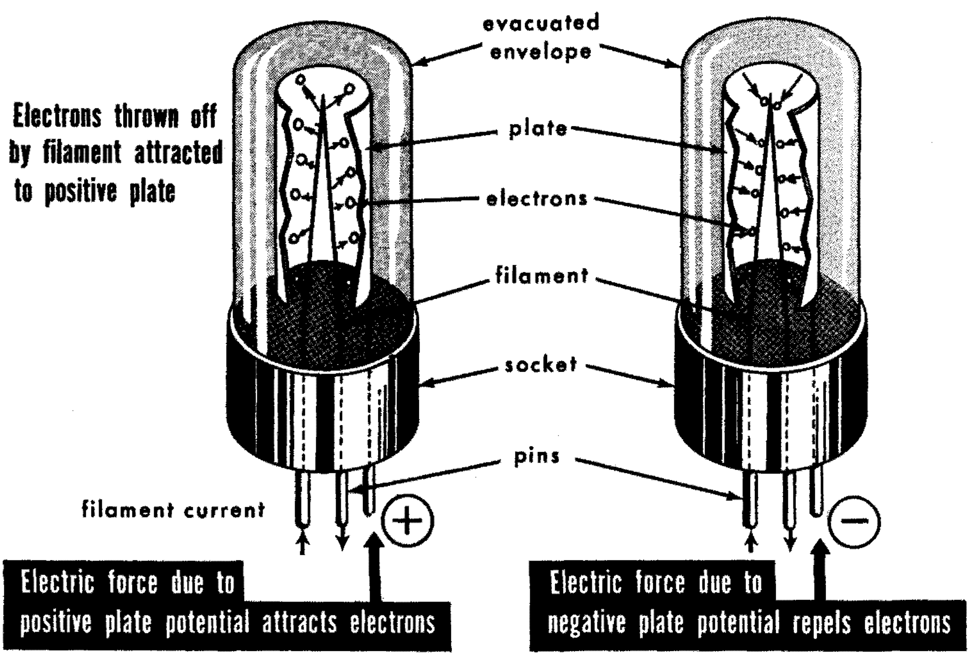
New York, N.Y.
August 1959

CONTENTS

VOL. 2—BASIC AUDIO

The Diode

CONSTRUCTION AND ACTION OF A DIODE



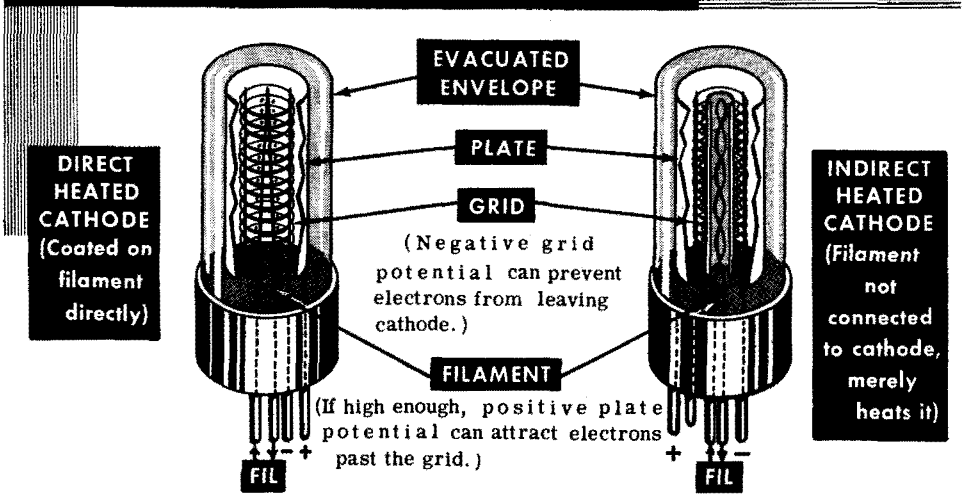
The first electronic amplifiers used a tube type called a *triode* because it has *three* active electrodes. The earliest triodes started with an addition to the *diode (two-electrode)* tube which used a filament of wire heated by passing a current through it and a piece of cold metal called a *plate*. These elements were placed in a space from which all the air was evacuated, usually enclosed by a glass *envelope*.

The diode *rectifies* current (allows it to pass only one direction). By electron theory, this action is explained as follows: Heating the filament causes the electrons in it to become very agitated, and they leave its surface freely in a vacuum. Electrons have a negative charge, hence if a positive charge is present in the vacuum, the electrons will be drawn to it, reaching the positive electrode. If, however, the potential on the plate is negative, it will repel the electrons (like charges repel, unlike charges attract), so they will stay near the filament, and no further charge can be taken by the plate. Therefore, when the plate is negative, no charge passes, which means that no current flows through the tube.

The Triode

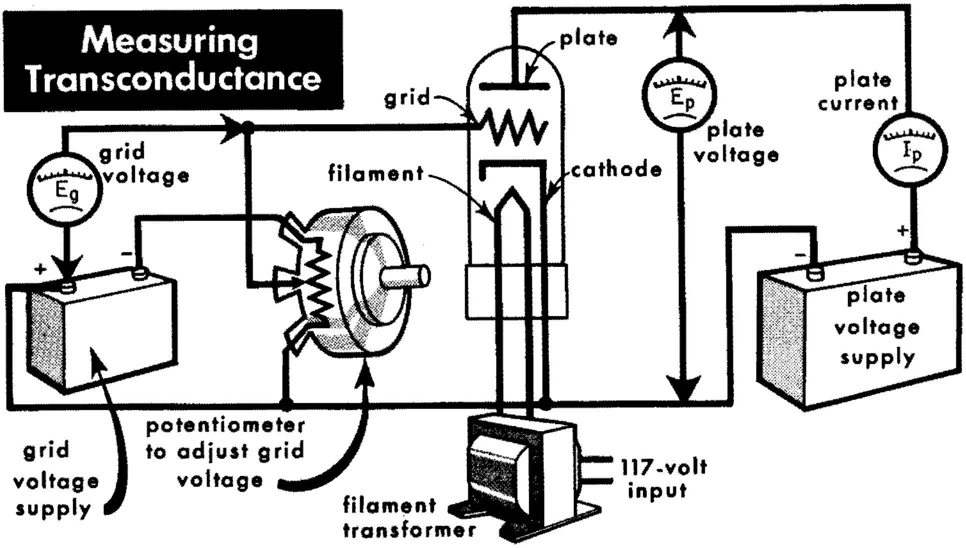
Whether electrons come away from the filament or stay close to it depends on the electric field at the filament surface. This force can be made negative, even though the plate is positive, by putting a negatively charged wire mesh, or grid, between the *plate* and the filament. If the grid is sufficiently negative, it will keep all the electrons close to the filament; if it is less negative, the positive force from the plate will overcome it to some extent, and some of the electrons will get through. Thus, the number of electrons that pass from the filament to the plate depends on how negative the grid is and how positive the plate is.

Construction and Action of a Triode



Beside its modern counterpart, the early triode was crude. The filament had to have current flowing through it to keep it hot, so that electrons would be thrown off *(emitted)*. The only essential thing, however, is to have a hot metal emitter, or *cathode*. By using a separate *heater* wire, to which current is supplied to provide the heat, the cathode does not need to have this supply connected to it. This makes the tube much more versatile. Improvement in materials and better design has enabled smaller tubes to be made, and now space is further saved by putting two or more "tubes," or complete electrode assemblies, into one evacuated "envelope." (Strictly speaking, the envelope—made of glass in most cases—is the *tube*, but as each electrode assembly in one envelope *can* be used separately if desired, it has become the practice to call each assembly a "tube" because it acts as one.)

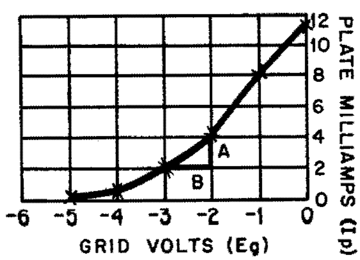
Transconductance



How is it that a small change in voltage at the grid can be converted into a larger one at the plate? Two things control how many electrons get away from the cathode of the tube: the plate voltage and the grid voltage. The number of electrons flowing inside the tube fixes the current in the outside circuit. Using a fixed plate voltage, the grid voltage can be changed and plate current measured each time it is changed.

Each time the grid voltage is changed, the plate current changes, until the potential on the grid is so negative that no electrons leave the plate, at which point the current becomes zero. The amount of change in plate current for each volt change in grid voltage is called the *mutual conductance* or *transconductance* of the tube. It is, of course, different for each change, but the figure usually quoted for a tube is the change that occurs when the voltage and current are nearest to the practical operating values used in a circuit.

Plate Current versus Grid Voltage Curve used to obtain Transconductance



EXAMPLE:

TRANSCONDUCTANCE $= \frac{A}{B}$

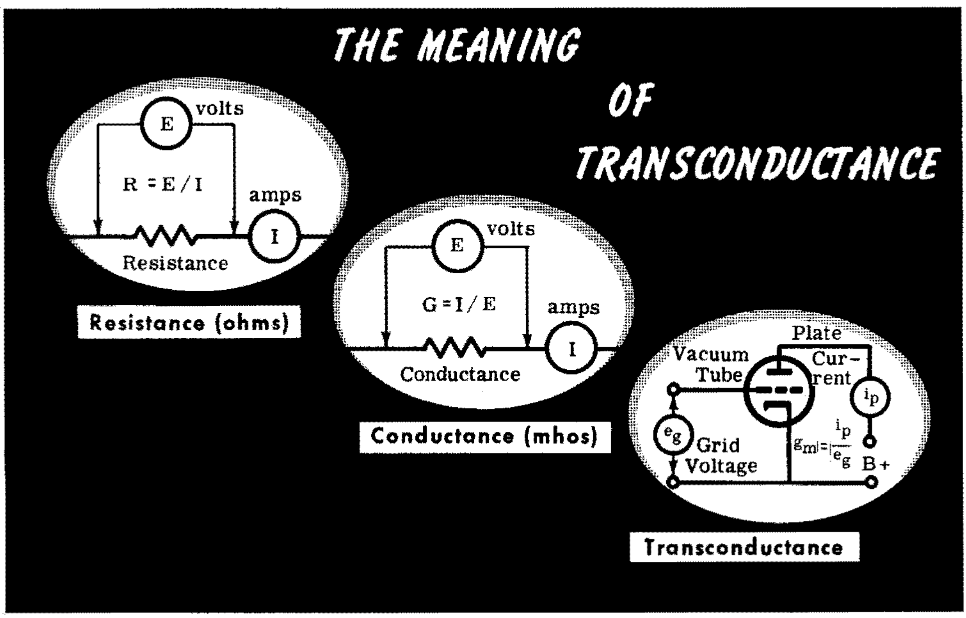
IF Eg VARIES FROM −3 TO −2 VOLTS, Ip VARIES FROM 2 TO 4 MA

Transconductance (contd.)

Resistance is the opposition that a circuit presents to current flow—the larger the resistance, the smaller the current at the same voltage. *Conductance* is the opposite of resistance—a measure of the *ease* with which current can flow through a circuit. A higher conductance means that more current can flow at the same voltage, or that the same current will flow at a smaller voltage. *Transconductance* is a measure of the ease with which current in one circuit (the plate circuit) can be affected by the voltage in another (the grid circuit).

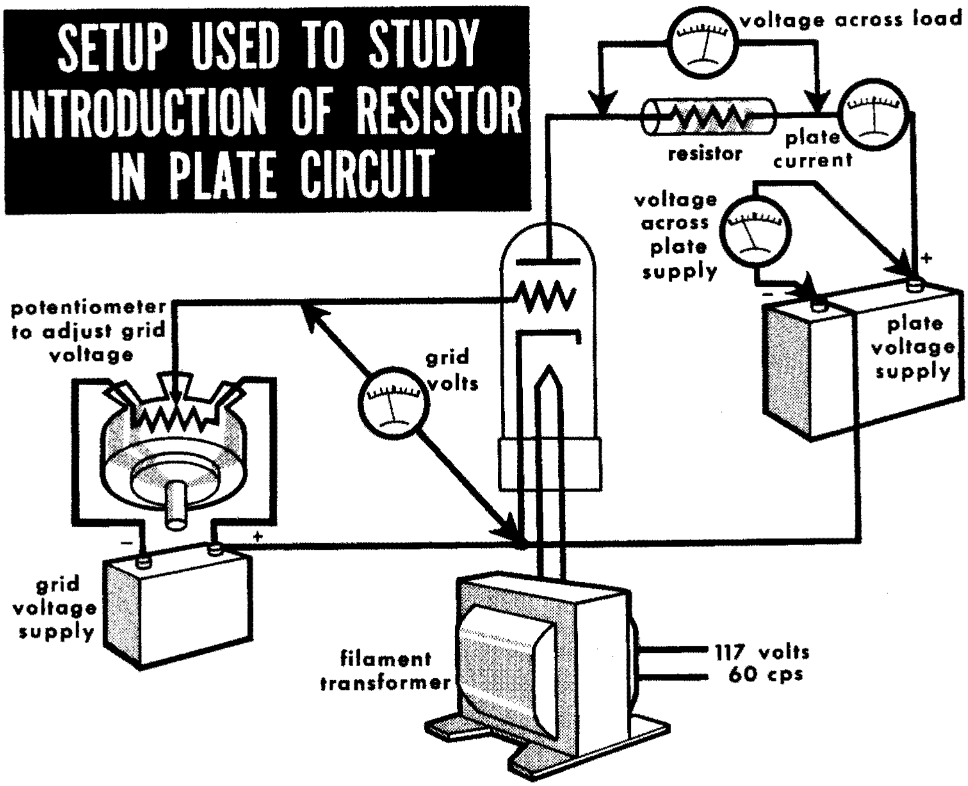
The relationship for resistance is the well-known *Ohm's law,* which can be expressed in symbols as $R = E/I$. The unit for conductance is the reciprocal of that for resistance: instead of being *volts per ampere,* it is *amperes per volt.* This unit is the mho (ohm spelled backwards). The conductance form of Ohm's law is $G = I/E$.

In tubes, a change of a volt at the grid only produces a change of milliamperes at the plate. (A milliampere is one-thousandth of an ampere.) For this reason, transconductance (symbolized as g_m) is generally quoted in *milliamperes per volt.* (We use g rather than G, because *conductance* applies to *direct* readings of current and voltage, whereas the tube works on *changes* in voltage and current.) Milliamperes per volt could also be called *millimhos.* But this terminology is never used; instead, the unit of transconductance is the *micromho,* which is a *microampere per volt.* (A microampere is a millionth of an ampere.) Thus, 2 *milliamperes per volt* is called 2000 *micromhos.*



Dynamic Gain Measurements

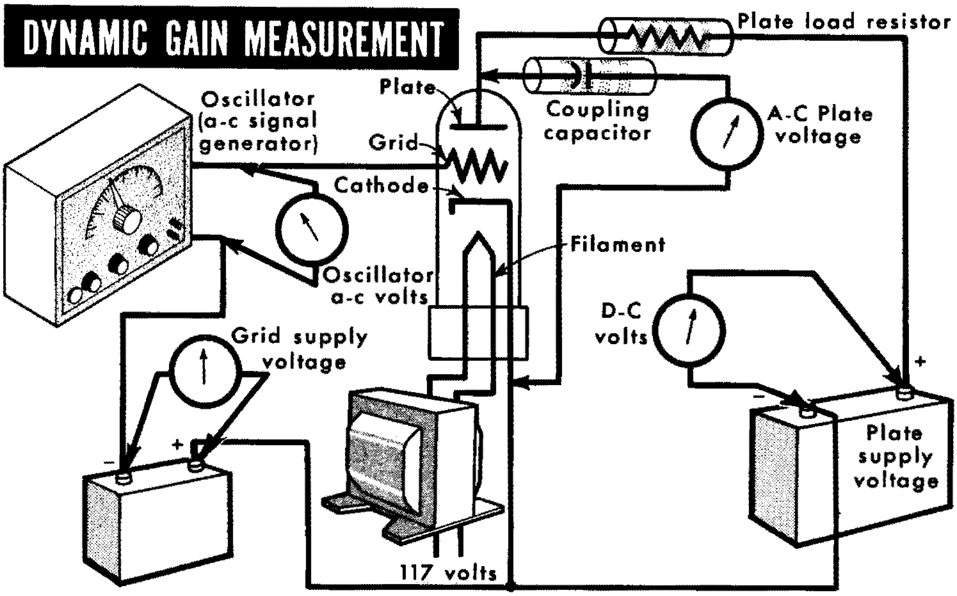
In the setup we used, the plate voltage was kept constant, and a change in grid voltage caused only a change in the *current* in the plate circuit. This is not the kind of amplification needed. To be useful, a small input voltage fluctuation must give a larger output *voltage* fluctuation.



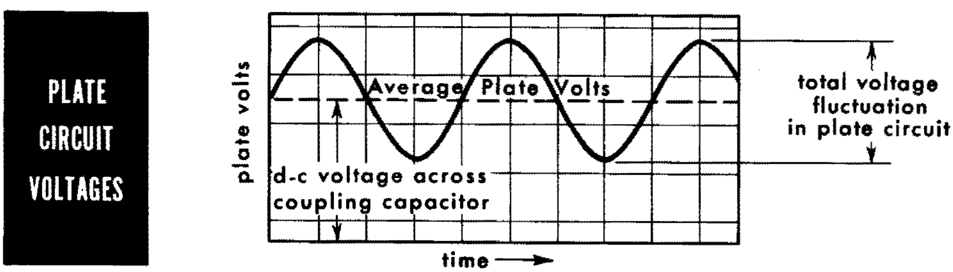
We can make the fluctuating plate current produce a voltage fluctuation by passing it through a resistor. The higher the value of the resistor, the greater the voltage drop that a given current will cause, and hence the greater the *voltage fluctuation* a given *current fluctuation* will produce.

The plate current, however, is dependent on *plate* voltage as well as on *grid* voltage. When the grid swings more positive (actually, *less negative*), the plate current tends to increase; but when a resistor is in series with the plate, the increased current causes the plate voltage to drop. A drop in plate voltage by itself causes a drop in plate current, hence the drop in voltage due to the resistor means that the current will not rise as much as it did for the same grid-voltage change, when the resistor was not in the circuit.

Dynamic Gain Measurements (contd.)



We could take a lot of time making adjustments of this kind, but there is a quicker way to get the answers we need. This is the *dynamic measuring* method. It enables us to measure the voltage fluctuation continuously, instead of having to take a whole series of individual measurements and then combine them. We apply a fixed d-c voltage, or *bias* at the grid together with a fluctuating voltage from an oscillator. (These voltages can be measured separately.) This produces a fluctuating voltage in the plate circuit that can be broken into a fixed component and an alternating component.



A capacitor placed between the plate and the a-c voltmeter takes on a charge corresponding to the steady or d-c component of the voltage, but does not have time to alter that charge with the alternating component. The voltmeter, therefore, measures only this alternating component.

Dynamic Gain Measurements (contd.)

Using a larger value of coupling resistor increases the output fluctuations for a given input fluctuation (the *gain* of the tube) up to a point. Eventually, the increase flattens off, and further increase in the value of the resistor causes the output voltage to decline somewhat. This decline occurs because the larger resistor values cause the *average* (steady d-c) voltage on the plate to drop, so that the tube gets "strangled." This can be overcome by adjusting the supply voltage so that the average or *operating* voltage, measured at the plate, is the same for each resistor used.

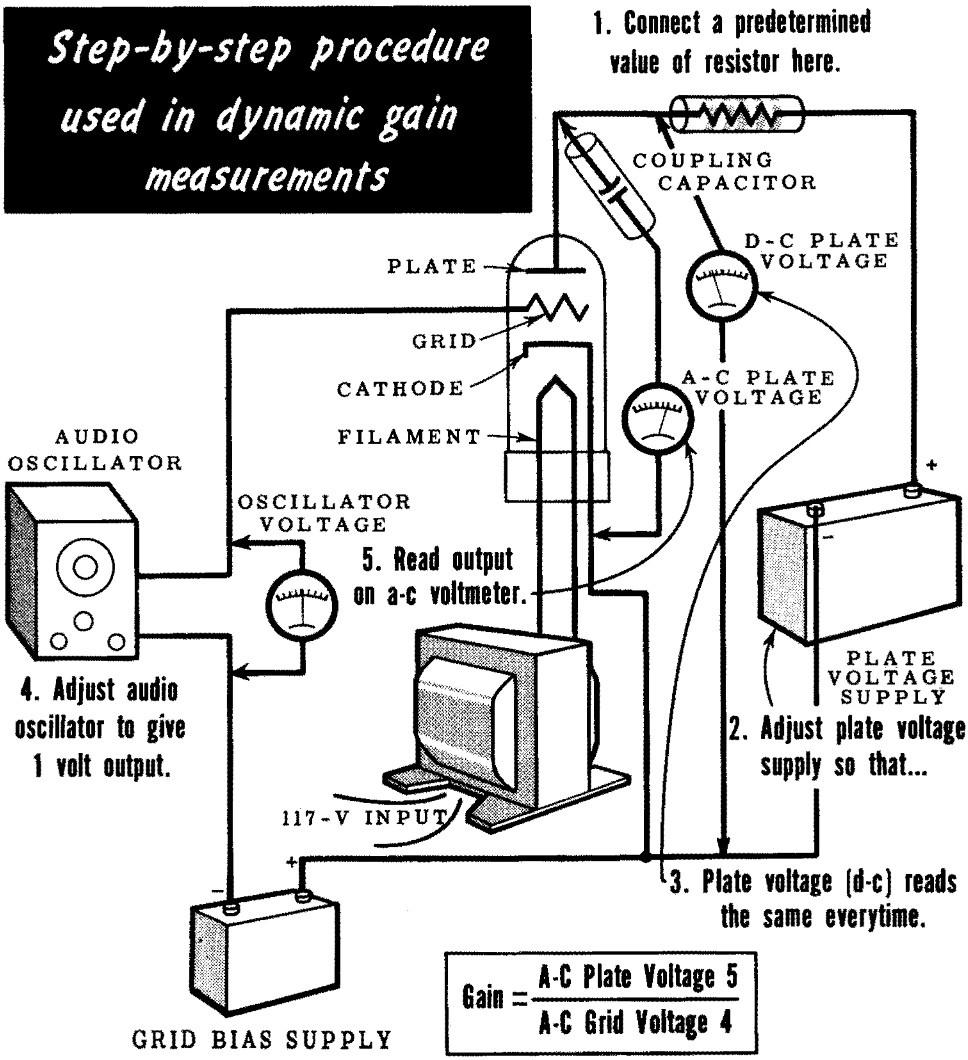
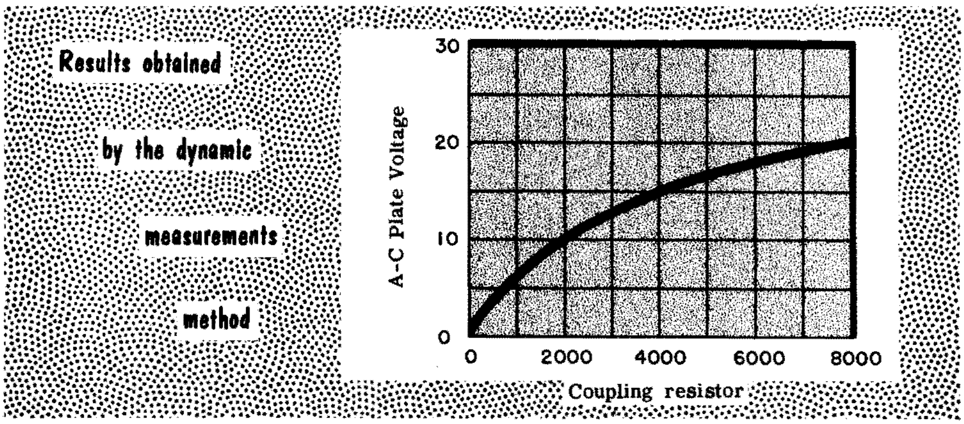


Plate Resistance



When the tests are made this way, the output, or gain of the tube, goes up indefinitely, until we have difficulty in finding supply voltages high enough to maintain the same plate voltage.

A curve almost exactly like this can be made using only resistors without any tube. (A larger alternating voltage is needed, because there is no amplification, and the proper fixed resistance has to be chosen to get the curve the same shape.) From this we see that the tube is somewhat like a generator that produces an a-c voltage and has an *internal* resistance, the value of which is that used in the equivalent circuit that produces the same curve.

The number by which the a-c grid voltage has to be multiplied to give the input voltage in the equivalent circuit, is called the *amplification factor* (or *magnification factor)* of the tube. The value of the fixed resistor in the equivalent circuit, is called the *plate resistance* or *a-c resistance* of the tube.

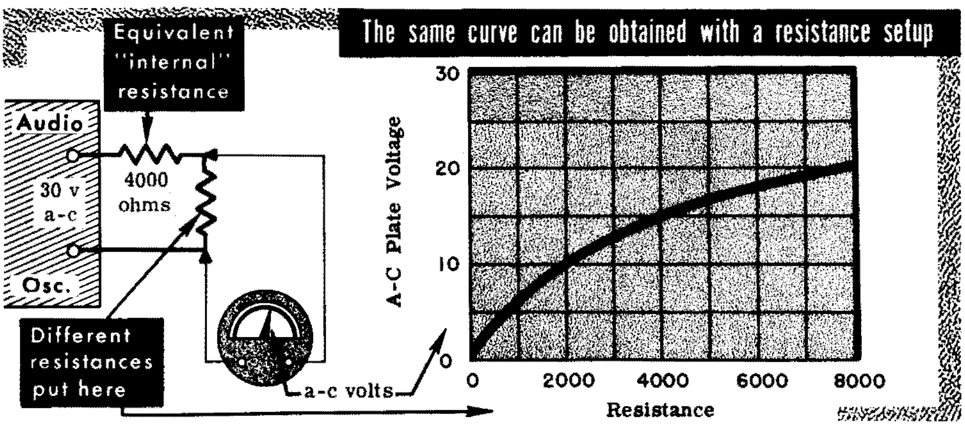
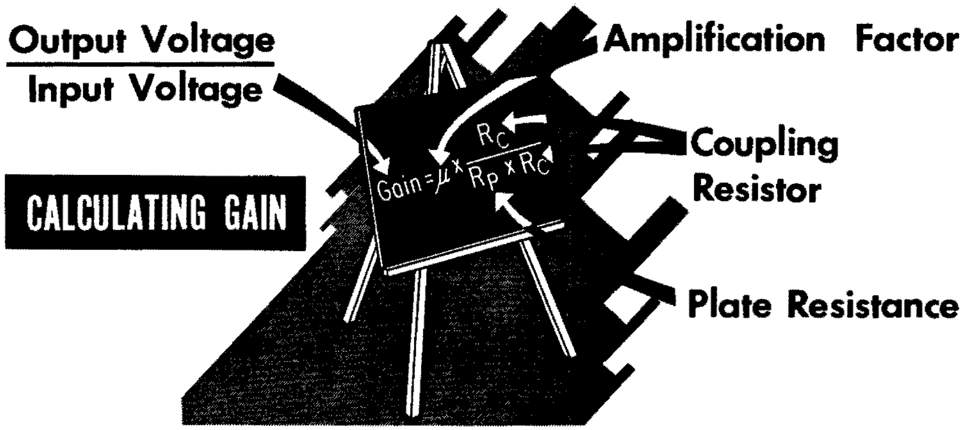
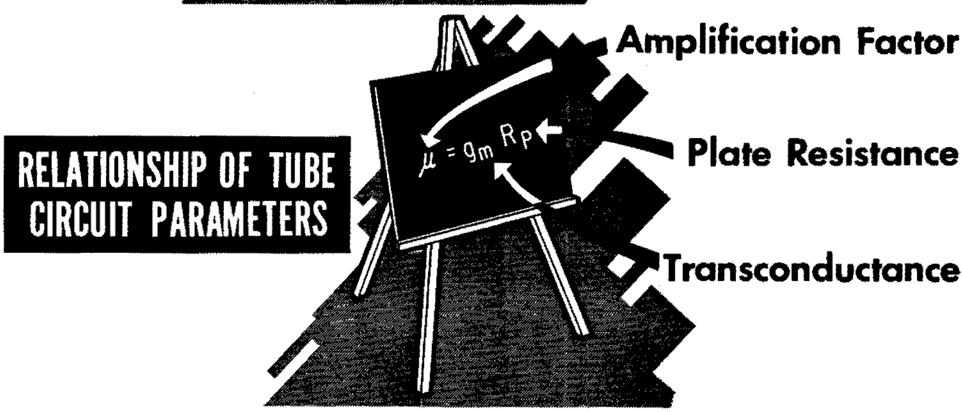


Plate Resistance (contd.)

Because voltage in a series circuit is proportional to the total resistance in the circuit, we can deduce that the output voltage from the tube is given by $\mu \times R_c/(R_p+R_c)$. The bigger the coupling resistor between the supply voltage and the plate of the tube, the less the plate *current* must fluctuate to produce a given plate *voltage* fluctuation. If, in imagination, we could make the resistance and the supply voltage infinitely large, only the plate voltage would vary. The gain of the tube in this imaginary circuit is its amplification factor.

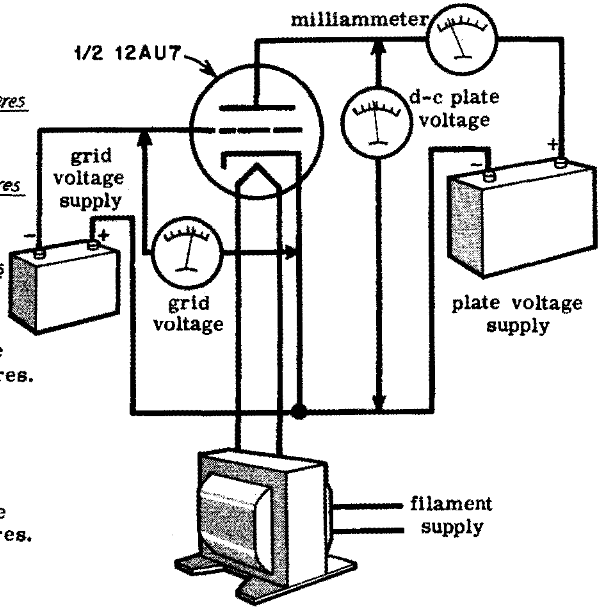


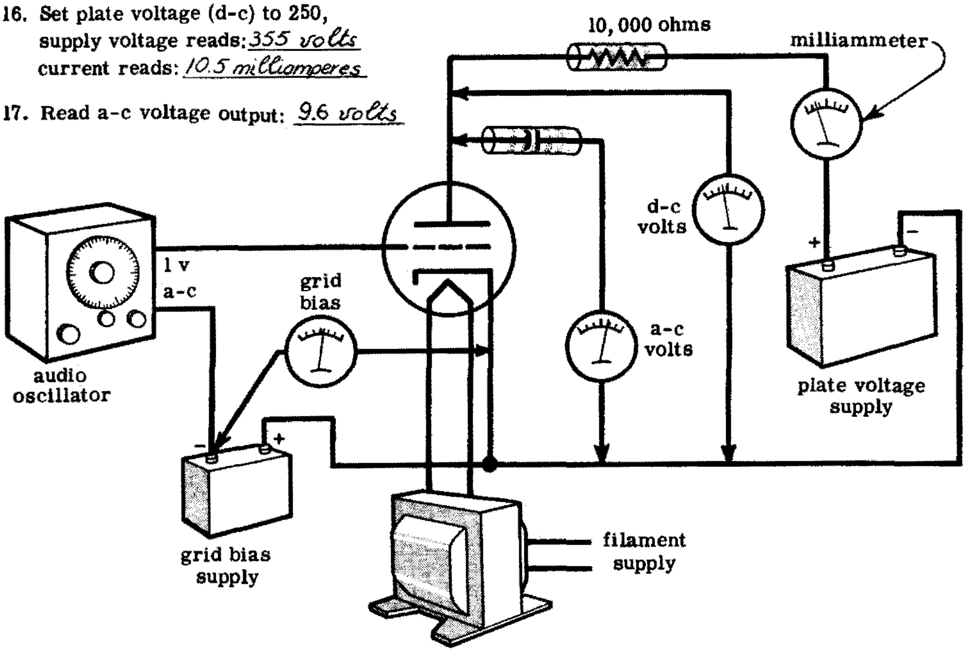


If there is no coupling resistor present (this was the case when we first examined the circuit), the voltage remains constant, and the current fluctuates. The plate current fluctuation for each volt of grid fluctuation is the transconductance of the tube. The relationship between these two conditions (which have no practical use, but help no end in figuring out the operation of practical tube circuits) is determined only by the parameter that we have called plate resistance.

Demonstration and Calculation of Gain

1. Set plate voltage to 250.
2. Set grid voltage to 8. 5.
3. Read plate current: 10.5 milliamperes
4. Set grid voltage to 7. 5.
5. Read plate current: 12.7 milliamperes
6. Set grid voltage to 9. 5
7. Read plate current: 8.3 milliamperes
8. Set grid voltage to 7. 5.
9. Adjust plate voltage so that plate current is again 10. 5 milliamperes.
10. Read plate voltage: 233 volts
11. Set grid voltage to 8. 5.
12. Adjust plate voltage so that plate current is again 10. 5 milliamperes.
13. Read plate voltage: 267 volts
14. Add 10, 000-ohm resistor and oscillator.
15. Set bias to 8. 5 volts, a-c input to 1 volt.
16. Set plate voltage (d-c) to 250, supply voltage reads: 355 volts current reads: 10.5 milliamperes
17. Read a-c voltage output: 9.6 volts





Demonstration and Calculation of Gain (contd.)

Let us take some figures for a real tube, the 12AU7, which is a general-purpose miniature twin triode with two sets of elements in one envelope. Since both sets have the same characteristics, we need consider only one of them. Using a 250-volt plate supply and a bias of —8.5 volts, a 1-volt change at the grid produces a 2.2-milliampere change at the plate, from 10.5 milliamperes at —8.5 volts to 12.7 milliamperes at —7.5 volts, or 8.3 milliamperes at —9.5 volts.

Of course, we cannot put the imaginary infinite resistance into the circuit, and we should have considerable trouble finding a source of infinite voltage, but we can get the desired effect by changing the grid voltage and altering the plate voltage enough to keep the plate current constant. (See steps 8 through 13.) Using this method, the current is 10.5 milliamperes with 250 volts on the plate and —8.5 volts on the grid. Changing to —7.5 volts on the grid, the plate potential must be dropped to 233 volts to keep the current at 10.5 milliamperes. Changing the grid potential to —9.5 volts, necessitates raising the plate potential to 267 volts.

From the first measurement (steps 1 through 7), the transconductance is found to be 2.2 milliamperes per volt (2200 micromhos). From the second measurement (steps 8 through 13), the amplification factor is found to be 17, because the change in plate potential required for constant plate current with a 1-volt change in grid potential is 17 volts each way. From this we can find the plate resistance (g_m must be in *amperes per volt*, to give R_p in *ohms*): therefore, $\mu = g_m R_p$ and $R_p = \mu/g_m = 17/.0022 = 7700$ ohms.

Now assume that we are using a 355-volt supply and a load resistor of 10,000 ohms, with the bias still at —8.5 volts. With 10.5 milliamperes of current passing through it, the drop will be 105 volts, still leaving the average voltage at the plate at 250 volts. By calculation, the gain should be

$$\frac{e_o}{e_i} = \mu \times \frac{R_c}{R_p + R_c} = 17 \times \frac{10{,}000}{7700 + 10{,}000} = 9.6$$

Each volt change on the grid will cause 9.6 volts change at the plate. This checks with the experiment (steps 14 through 17), using a 10,000-ohm resistor and a 335-volt supply.

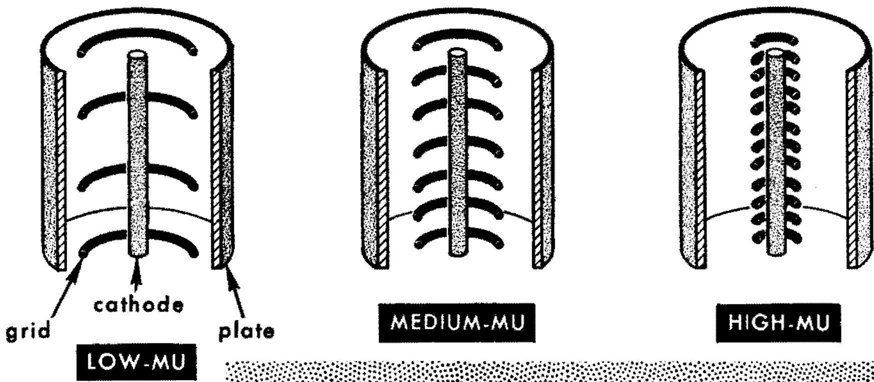
With this tube, a practical amplification of about 10 can be obtained, but regardless of the way in which the tube is connected, the amplification will always be less than 17, the amplification factor of the tube.

This is the basis for calculating *voltage amplification*. The tube is regarded as producing an amplification of the input voltage, stated by its amplification factor at the chosen operating point. The actual voltage amplification is then this number, divided between the plate resistance of the tube (at the same operating point) and the plate load or coupling resistance. In the example, the total amplification of 17, theoretically provided by the tube, is divided into 7 (lost in the plate resistance) and 10 (actually delivered across the coupling resistor).

Different Types of Triodes

A great variety of tubes can be made by varying the structure of the grid and the position that it occupies relative to the cathode and plate. If the tube manufacturer winds the grid wires closer together and also puts them closer to the cathode, he will have a tube with both higher plate resistance and higher amplification factor.

Variations in structure change tube parameters



CHARACTERISTICS OF TUBES

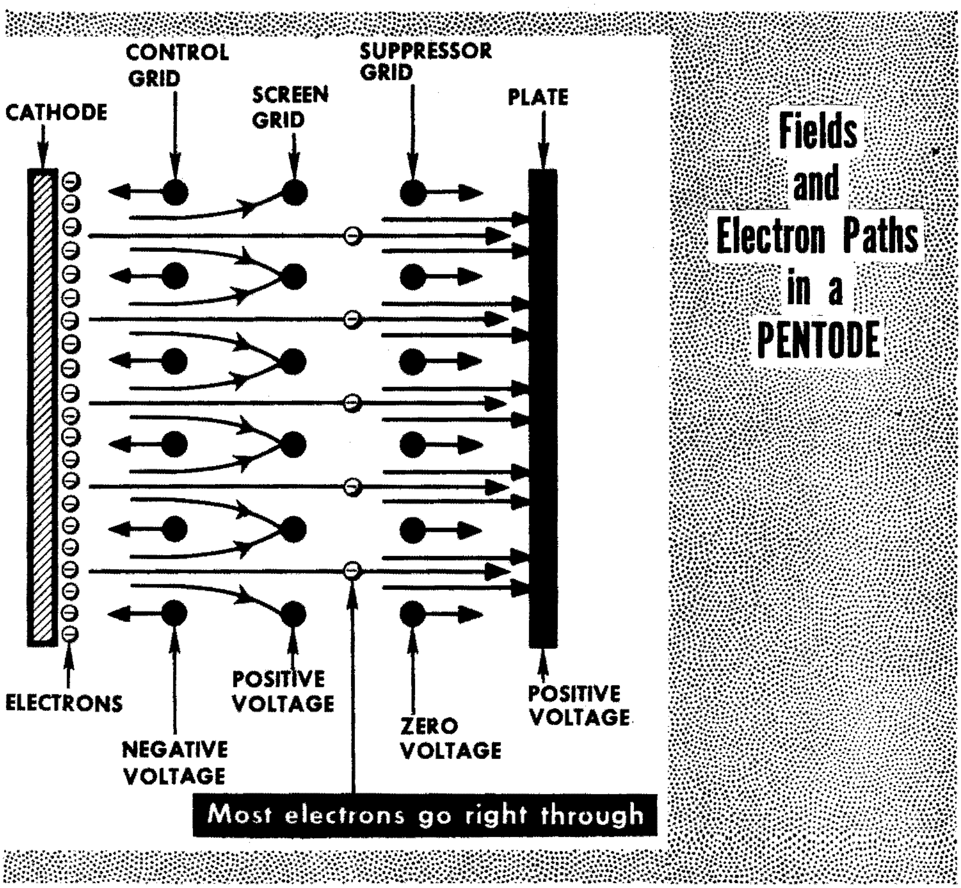
Type	Transconductance (milliamps per volt)	Plate Resistance (ohms)	Amplification Factor
12AT7	5.5	10,900	60
12AV7	2.2	7,700	17
12AV7	8.5	4,800	41
12AX7	1.6	62,500	100
12AY7	1.75	22,800	40
12BH7	3.1	5,300	16.5
ECC81	6.7	10,500	70

One-half of a 12AX7 double triode, for example, has a transconductance of 1.6 milliamps per volt, with a plate resistance of 62,500 ohms, which gives an amplification factor of 100. The transconductance is lower than that in the 12AU7 because the structure of the tube reduces the possible plate current. Tubes with a high transconductance usually have a lower plate resistance and vice versa. This fact sets a limit to the amplification that can be obtained from a single triode. What we need, to get still more amplification from a single stage, is a tube with a high transconductance and also a high plate resistance.

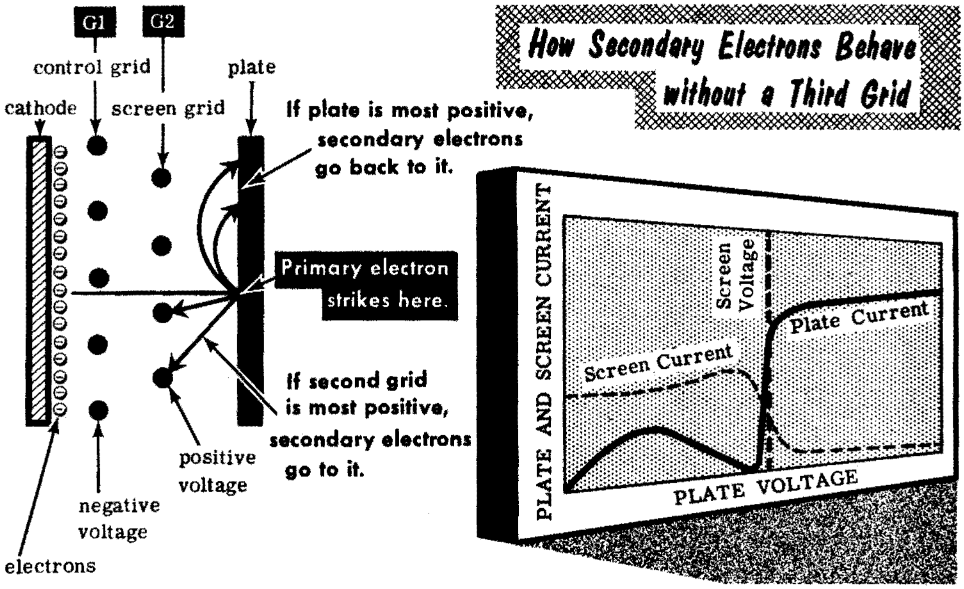
The Pentode

In a triode, the plate current is controlled by the combined electric field at the cathode due to both the grid and the plate. Because of the triode's open grid structure, the plate voltage influences the plate current. If we could decrease this dependence of plate current on the voltage at the plate we could boost the amplification of the tube.

This is what the pentode type tube does. Two extra grids are inserted between the number 1 or control grid and the plate. The first of these, the *screen* grid, is maintained at a constant positive potential and, from the viewpoint of plate current, is responsible for providing the electric field at the cathode. This means the control-grid voltage controls the electron flow from the cathode in conjunction with a constant field derived from the screen grid, not a fluctuating one from the plate. Because the screen grid is an open mesh and not a solid plate, most of the electrons reaching it go through the spaces instead of hitting the wires.

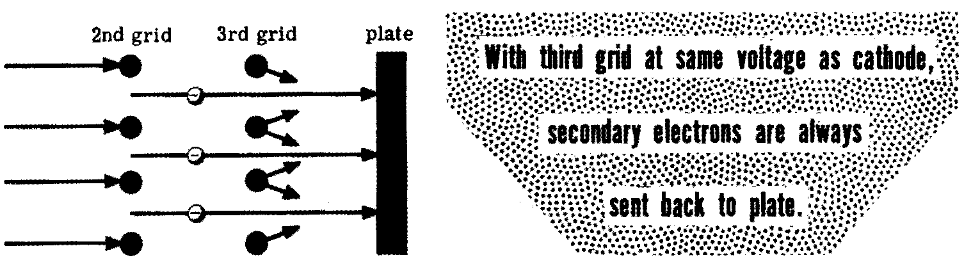


The Pentode (contd.)



The second extra (i.e., third) grid is inserted to stop electrons from bouncing back from the plate to the screen grid when the plate potential is lower than the screen potential. If the second grid has the highest voltage of all and there were no third grid, electrons would hit the plate, cause further electrons to bounce off it, and be attracted by the more positive voltage at the screen grid. This *secondary emission* could even result in negative plate current.

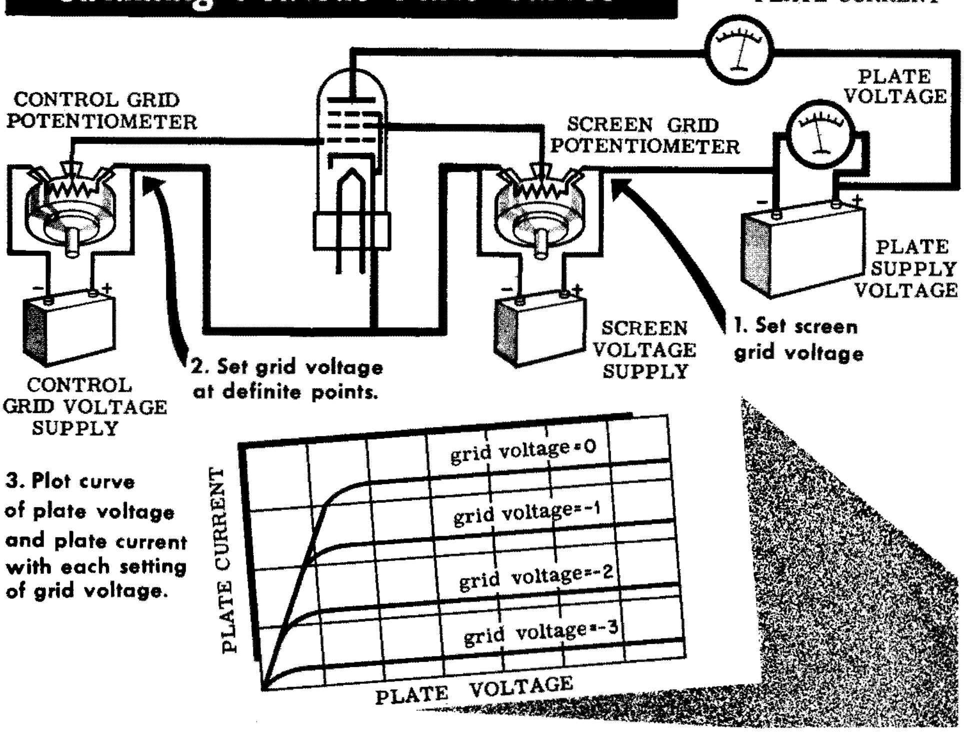
The third *(suppressor)* grid prevents secondary emission by providing another electric field (between the suppressor grid and the plate) that discourages the bouncing off of any electrons. The third grid does not collect any electrons, because its voltage is so much more negative than that of both the screen grid and the plate. It merely serves to insure that all the electrons that come through the spaces in the screen grid arrive at the plate *and stay there.*



The Pentode (contd.)

This results in a tube in which the plate current is almost completely independent of the plate voltage, and is determined by the voltages on the control and screen grids. (Usually the potential on the screen grid is kept constant, and the audio fluctuations are applied to the control grid.) The plate resistance of this type of tube runs to very high values—perhaps several megohms (millions of ohms). A pentode with a transconductance of, say, 2 milliamperes per volt, and a plate resistance of 1 megohm has an amplification factor of $.002 \times 1{,}000{,}000 = 2000$.

Obtaining Pentode Plate Curves

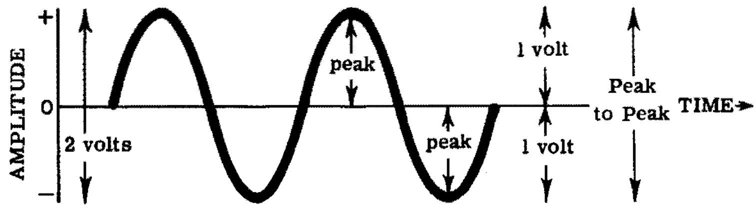


This much amplification cannot be achieved in practice because a very high plate load resistor would be necessary. Supposing that we use a 1-megohm coupling resistor, using the formula we should achieve a practical amplification of 1000. As we shall see presently, even 1 megohm is rather a high value for a coupling resistor, unless we will be content with poor-quality amplification. A resistor of about 220K, however, is quite normal and will give an amplification in the region of 360, using a tube with a transconductance of 2 milliamperes per volt.

The Sine Wave

We have started to talk about alternating (or a-c) voltages, giving them figures in some cases, as we might measure them with a suitable instrument. But there are several ways of measuring alternating voltages and currents. Later on these will cause confusion if we do not get the differences straight now.

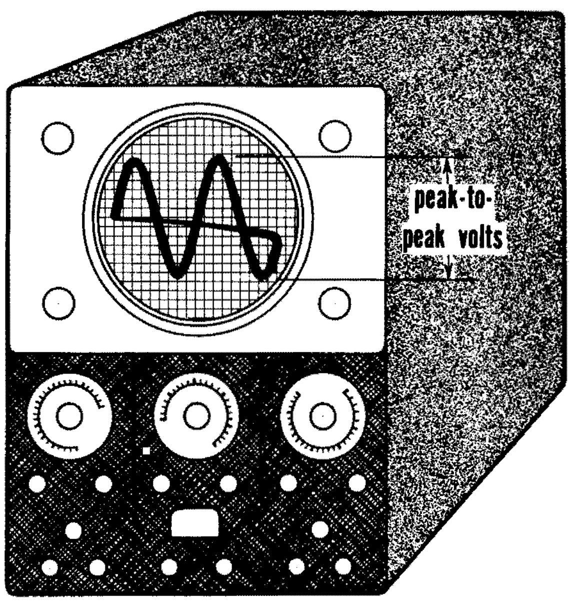
A Sine Wave of 1-Volt Peak



A pure frequency has a waveform called a *sine wave*, and is called a *sinusoidal wave*. If the wave fluctuates between +1 volt and —1 volt from 0 as the starting point, the total change, called the *peak-to-peak* voltage, is 2 volts. The "swing" each way from the central starting point is called the *peak* voltage, in this case 1 volt.

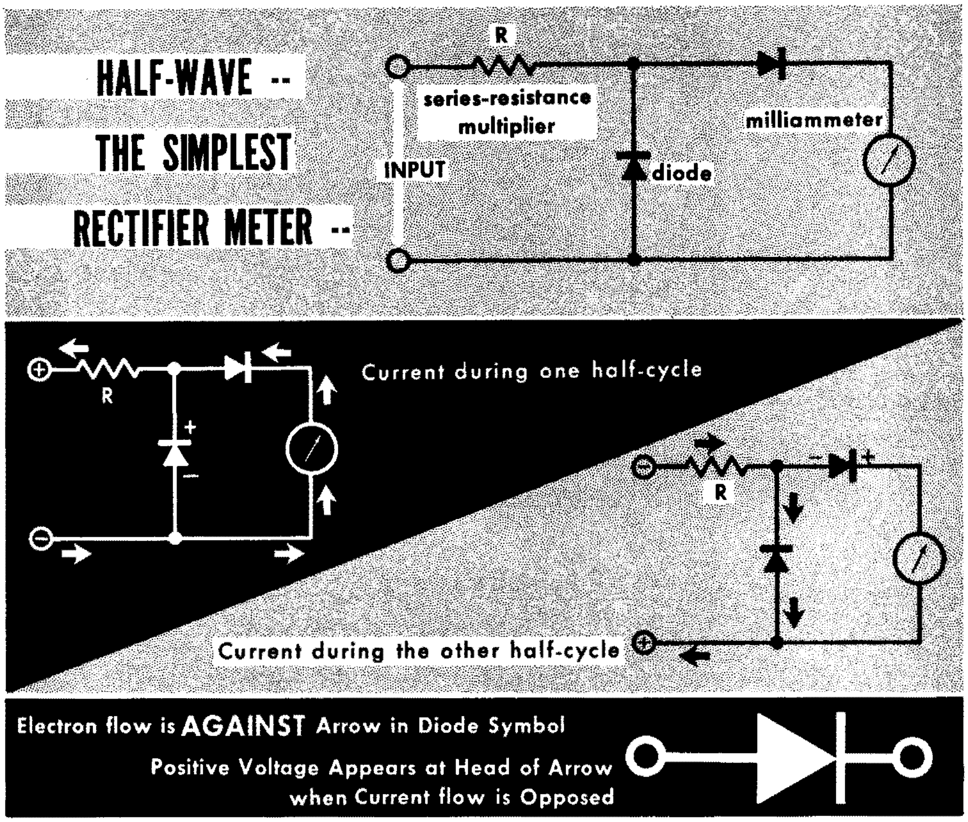
When an oscilloscope is used to observe a waveform, it is fairly easy to measure the peak, or peak-to-peak, waveform with a ruler, or by using a calibrated scale in front of the screen. Most meters however, do not measure either peak, or peak-to-peak,voltages.

Waveforms can be measured by graph transparency on oscilloscope



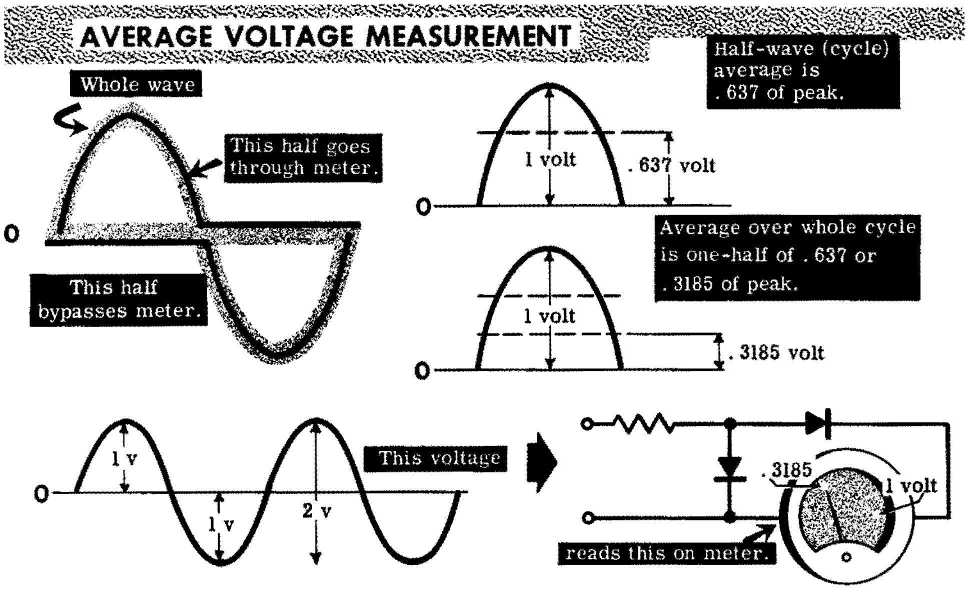
A-C Meter Readings

If an ordinary D'Arsonval moving-coil voltmeter is connected to an alternating voltage, one half-wave will try to make it read in the normal way, while the other half-wave will try to make it read backwards, as if it had been connected the wrong way around. As the pointer will not have time to move back and forth so rapidly, it will stand still, or vibrate slightly, without moving far from the zero marking. Because of this, moving-coil meters must be used with rectifiers to get a reading.



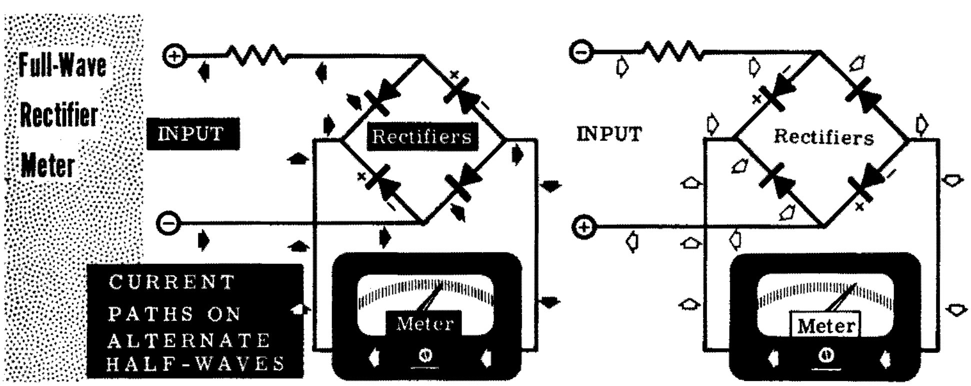
The rectifier is a metal-contact device that acts in the same fashion as a diode—it lets current flow in one direction, but not the other. In the symbol for the rectifier, the arrow points in the direction of current flow (opposite to electron flow). As a result of the action of the rectifiers, when this kind of meter is connected to an alternating voltage, one half-wave is bypassed through one rectifier, while the other half goes through the meter. Of course, the meter pointer will not have time to follow the fluctuations, so it will average out the current that flows through it.

A-C Meter Readings (contd.)



During one half-wave, no current flows through the meter, while during the other half-wave, it follows half a sine wave in form. The *average* of half a sine wave is 0.637 of peak value. However, during half the time, no current goes through the meter; therefore, the average over the whole time will be half of 0.637 or 0.3185. (If an ordinary d-c meter movement is used in this circuit to measure an alternating voltage, a 1-volt peak—2-volt peak-to-peak —voltage will only give a reading of 0.3185 volt.)

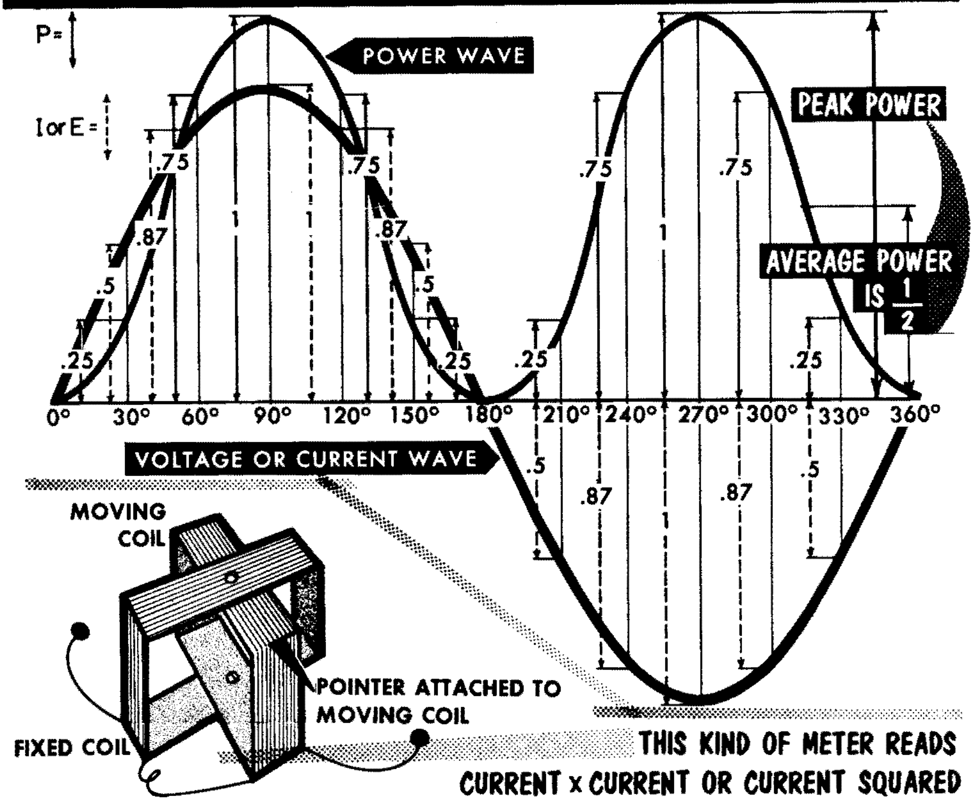
In an arrangement that uses four rectifiers, the meter gets both halves of the wave, and it will read 0.637 of the peak voltage if a regular d-c meter is used.



A-C Meter Readings (contd.)

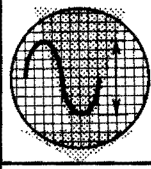
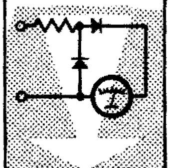
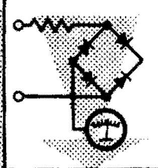
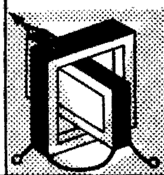
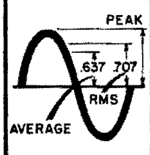
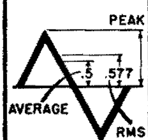
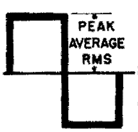
Thus all rectifier-type meters indicate a value that is some kind of average, but the average value is not directly related to practical circuit behavior. In audio work, we must know either the maximum voltages or currents or the power ($E \times I$) associated with them. If only the level changes (impedance or resistance is constant), the power is proportional to voltage or current squared, at any instant. Average power is determined by the average (or *mean*) of the squares.

Relationship between Voltage or Current and Power in Sine Wave



The power-equivalent reading of any a-c waveform could be read on a meter where both the field and the current in the moving coil come from the circuit being measured. This is called a *dynamometer* movement. Torque at every instant is proportional to power. So, the reading will be the average *power* represented by the current or voltage. If it is marked with square-root readings, or the steady current or voltage that would give the same reading, this meter will read true *rms* of any waveform.

DIFFERENT KINDS OF METERS GIVE DIFFERENT READINGS ACCORDING TO THE WAVEFORM

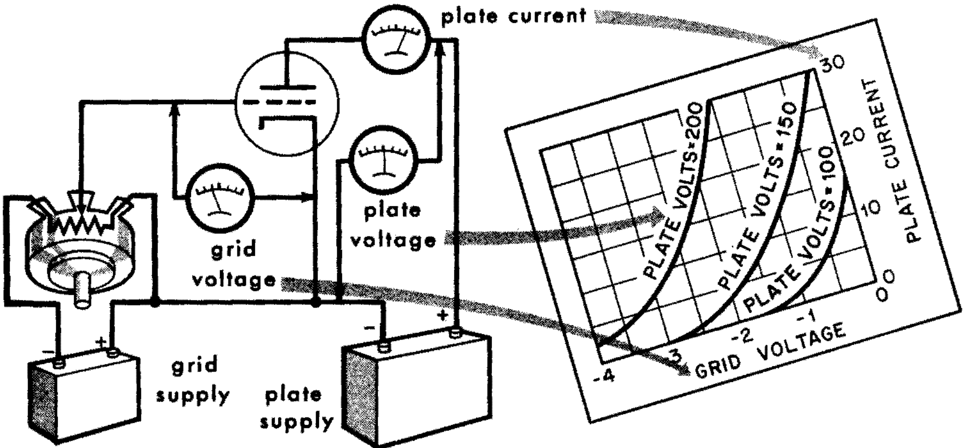
Waveform	When the Actual Value is...	oscilloscope will read 	half-wave meter will read 	full-wave meter will read 	rms meter will read 
Sine 	1 volt peak-to-peak	1	.1593	.3185	.3535
	1 volt peak	2	.3185	.637	.707
	1 volt rms	2.828	.45	.9	1
	1 volt average	3.142	.5	1	1.11
Triangular 	1 volt peak-to-peak	1	.125	.25	.2885
	1 volt peak	2	.25	.5	.577
	1 volt rms	3.464	.4085	.817	1
	1 volt average	4	.5	1	1.224
Square 	1 volt peak-to-peak	1	.25	.5	.5
	1 volt peak	2	.5	1	1
	1 volt rms	2	.5	1	1
	1 volt average	2	.5	1	1

So there are three main values related to waveforms: peak, average, and rms. To these may be added peak-to-peak (twice peak for a symmetrical waveform), and various other possibilities of averaging.

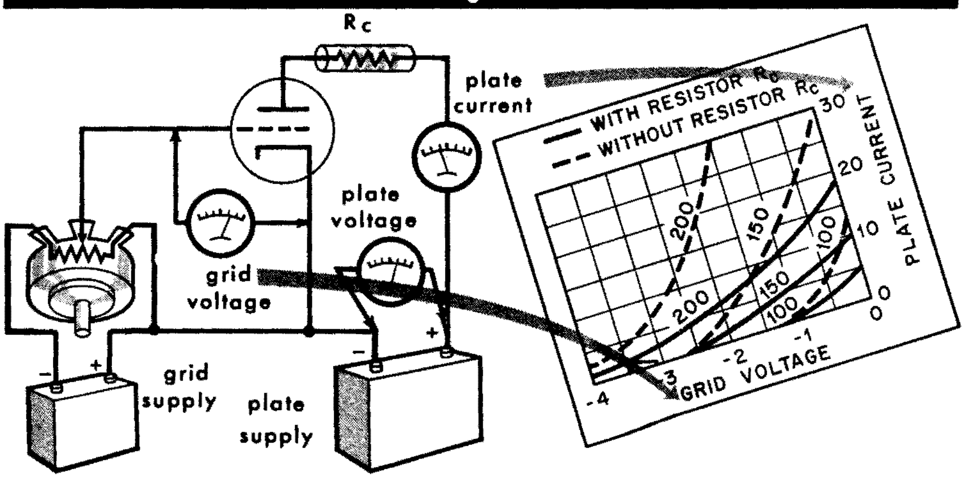
Load-Line Construction

We can find out more about amplification by plotting the curves of plate current and grid voltage with resistance connected in the plate circuit. However, we have to plot a new curve for each value of resistance with which we want to experiment. If, for some reason, we are not satisfied with any of the resistance values that we have already tried, the only thing to do is get the equipment out and plot some more curves.

Obtaining Plate Current versus Grid Voltage Curves for various values of plate voltage, Without Plate Load Resistor

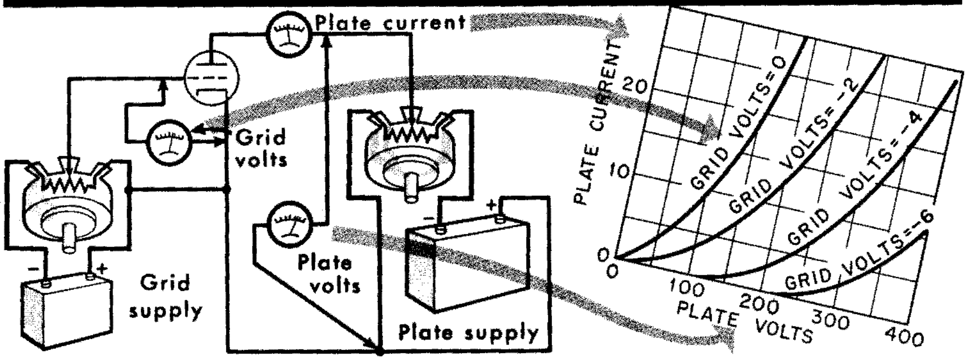


Obtaining Plate Current-Grid Voltage Curves With Plate Coupling Resistor R c in Circuit



Load-Line Construction (contd.)

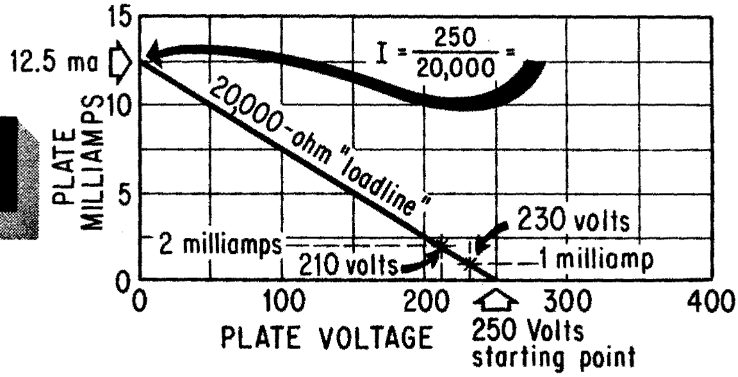
Setup for taking plate voltage-plate current curves, at different grid voltages



Fortunately there is a more direct way of obtaining all these curves. To start with we plot quite a different set of curves. Each of these curves shows all the possible combinations of plate current and voltage that can occur for one fixed potential on the grid. If we can draw, on the same graph, another line we can find all the possible relationships in the plate circuit, corresponding to whatever resistance or other circuit component is connected. Then these curves can be used to find how the circuit—any circuit—will work.

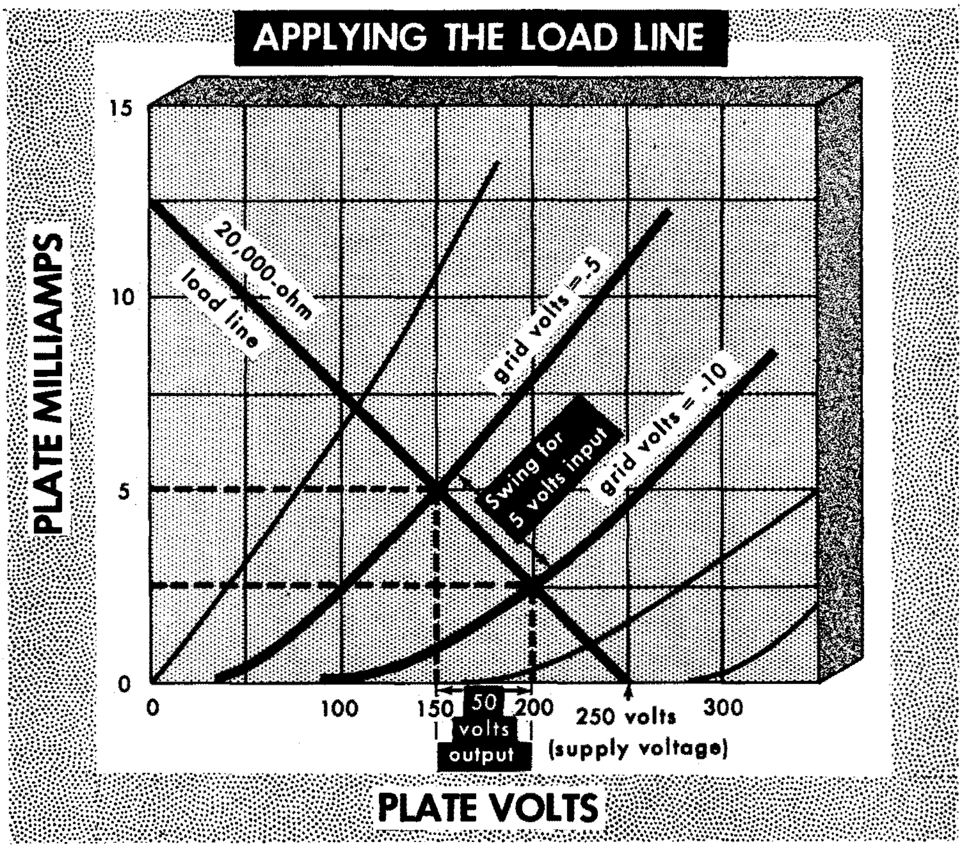
For example, if the supply voltage used is 250 volts, and the resistance used is 20,000 ohms, without any plate current flowing, the plate potential will be the same at both ends of the resistor—250 volts. If 1 milliampere of plate current flows, the drop across the resistor will be 20 volts, leaving 230 volts at the plate. With a plate current of 2 milliamperes, the drop will be 40 volts, leaving 210 volts at the plate, and so on. With 12.5 milliamperes flowing, the whole 250 volts will drop in the resistor, leaving 0 at the plate. These possibilities are shown by drawing a straight line through all these points.

Construction of a Load Line



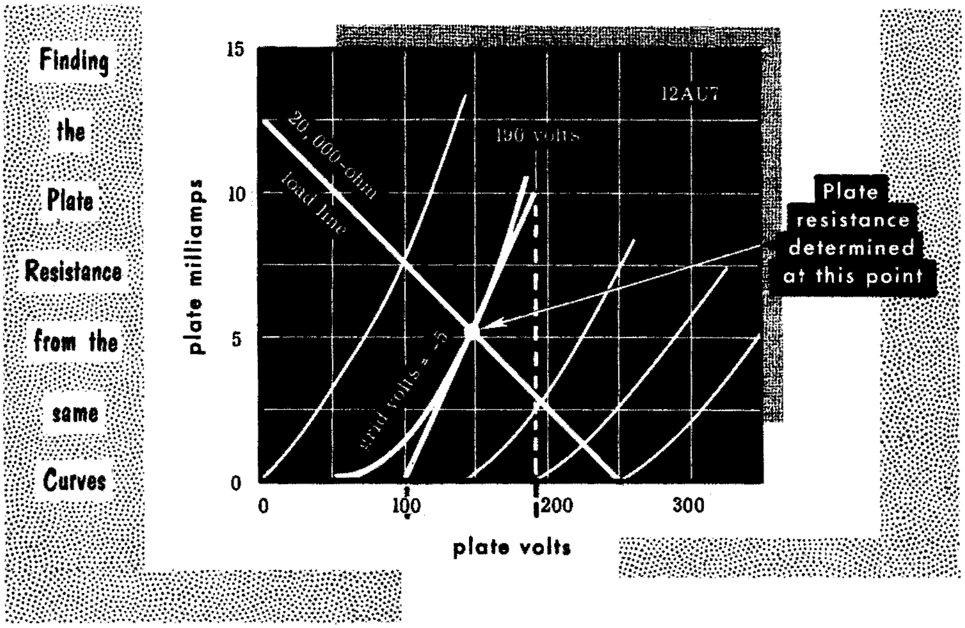
Finding the Plate Resistance

Now suppose the grid potential varies between —5 volts and —10 volts. These curves show that, when the grid potential is —5 volts, the plate current is 5 milliamps and the plate potential is 150 volts. This is where the —5-volt curve, showing possible combinations at this grid voltage, crosses the 20,000-ohm load line, showing the possible combinations with this resistor connected in series with the plate from a 250-volt supply.



When the grid potential is —10 volts, the plate current is 2.5 milliamperes, with a plate potential of 200 volts, where the load line crosses the —10-volt curve. Thus with this resistor, we can see quite easily that there is a swing between 150 and 200 volts on the plate. If a 5-volt input swing gives a 50-volt (200 — 150) swing at the output, the tube is working at a gain of 10, because the output swing is 10 times the input swing. These curves can also be used to find the plate resistance of the tube at different operating conditions (combinations of grid potential plate current and plate potential).

Finding the Plate Resistance (contd.)



By laying a straight-edge (ruler) along the curve at the point where the two lines cross, its slope *at this particular point* can be found. Extending the line drawn as tangent to the curve down to the zero current line and up to the 10-milliampere line, the corresponding voltages may be read off and the plate resistance calculated.

$$\text{PLATE (or A-C) RESISTANCE (in ohms)} = \frac{\text{Change in plate volts}}{\text{Change in plate current (In Amps)}} = \frac{90}{.01} = 9{,}000 \text{ ohms}$$

For example, using the 12AU7 curves and laying the ruler along the curve for —5 volts, where the 20,000-ohm load line crosses it, the line drawn as tangent goes through the zero current line at 100 volts and through the 10-milliamp line at 190 volts. This means, if the slope at the point we chose is extended, that a *change* of 10 milliamps causes a *change* of voltage of 90 volts. This is why the relationship is sometimes called an a-c resistance—because it deals with changes of fluctuations in current. The value of the resistance, in ohms, is given by Ohm's law. In this example, the plate resistance is 90/.01 or 9000 ohms.

QUESTIONS AND PROBLEMS

1. Explain why a diode rectifies current, and how the geometry of a triode affects the way in which the plate and the grid control electron flow.

2. What is transconductance, and why is it so called? State the conditions defining it.

3. What two units are used for measuring transconductance? If a tube has an amplification factor of 100 and a plate resistance of 62,500 ohms, give its transconductance in both units.

4. What is the plate resistance of a tube, and why is it sometimes called *a-c resistance?*

5. What is the amplification factor of a tube? State the conditions defining it.

6. How can the amplification of a tube be increased by its internal design? Discuss the practical limitations in achieving higher amplification from a triode tube.

7. A tube has an amplification factor of 100, a plate resistance of 80,000 ohms, and is used with a load resistor of 100,000 ohms. Calculate the approximate amplification of the stage from the appropriate formula.

8. What steps were taken in the design of tubes to obtain higher amplification than is possible with a triode? Explain the function of each added electrode.

9. Why is the amplification factor of a pentode seldom quoted in tube data for this type of tube? What figure is used instead?

10. What would happen to a moving-coil meter connected in an a-c circuit, without the use of a rectifier?

11. Draw two different types of rectifier meters commonly used for measuring a-c. State the relationship between the reading obtained and the maximum current alternations causing it.

12. Why is it that rms values are the standard used for specifying alternating voltages and currents, although most meters measure average values? How does the relationship change when different waveforms are used?

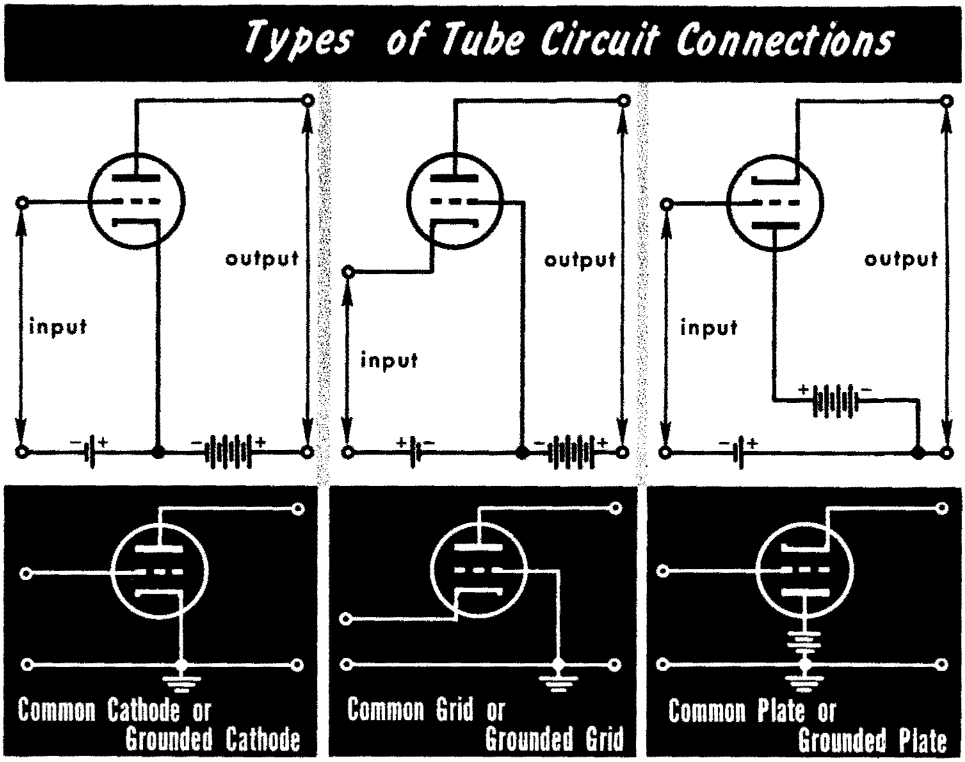
13. What two graphical ways are there of presenting the characteristics of a tube? State the relative merits of each method.

14. What is a load line? Show how it can be applied and discuss the information that may be obtained from it.

15. How can plate resistance be obtained from a tube's characteristic curves? Is plate resistance constant as various tube voltages are changed?

The Grounded-Cathode Amplifier

Thus far we have measured all voltages against the cathode of the tube. This assumes the cathode potential is zero, or in engineering language, the cathode is called the *reference* point, because everything is measured back to it. Because we also measure all potentials against the biggest reference body we can find—the earth—the part of the circuit to which everything is *referred* or measured, is called *ground,* although it may not always be literally connected to ground.



We encounter one more expression for this: because the point to which measurements are all made is *common* to both input and output circuits, this is called a common point. The way of working a tube we have described is called *common cathode,* or *grounded cathode.*

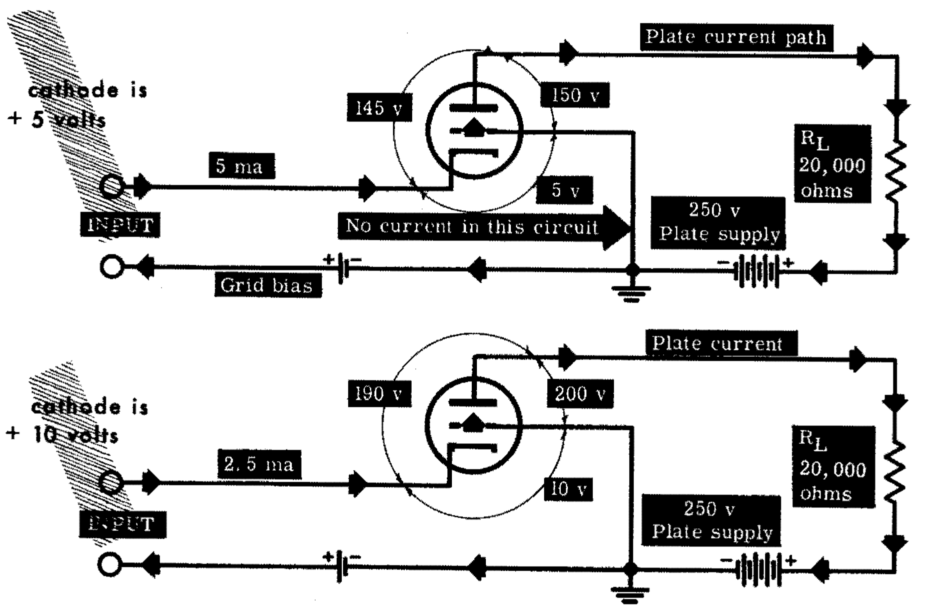
The grounded- or common-cathode circuit of a tube is the easiest to understand because it is the voltage between plate and cathode and between grid and cathode that controls the current between plate and cathode. Any other circuit is more complicated to understand, because we have to find out how it is related to this basic one.

The Grounded-Grid Amplifier

Let us first take the grounded- or common-grid circuit. If the input voltage is applied between cathode and grid, it is the same as placing an opposite voltage between grid and cathode. (Making the grid 5 volts *negative* to the cathode is the same as making the cathode 5 volts *positive* to the grid.) No current flows in the grid circuit because it is negative with respect to the cathode and repels all electron flow. Current does flow in the cathode circuit, and it is the same current that flows in the plate circuit.

With the same resistance in the plate circuit as before and a supply of 250 volts, with a fluctuation between +5 and +10 volts at the cathode, the current will fluctuate, as in the grounded-cathode arrangement, between 5 milliamperes and 2.5 milliamperes, respectively, while the plate-to-cathode voltage fluctuates between (150—5) or 145 volts and (200—10) or 190 volts. Thus the input fluctuates between +5 volts at 5 milliamperes and +10 volts at 2.5 milliamperes, the current in each case opposing the input voltage. This means the change in input current corresponding to a 5-volt change in input voltage is 2.5 milliamperes, which represents an *a-c* resistance for the input circuit of 5/.0025 or 2000 ohms.

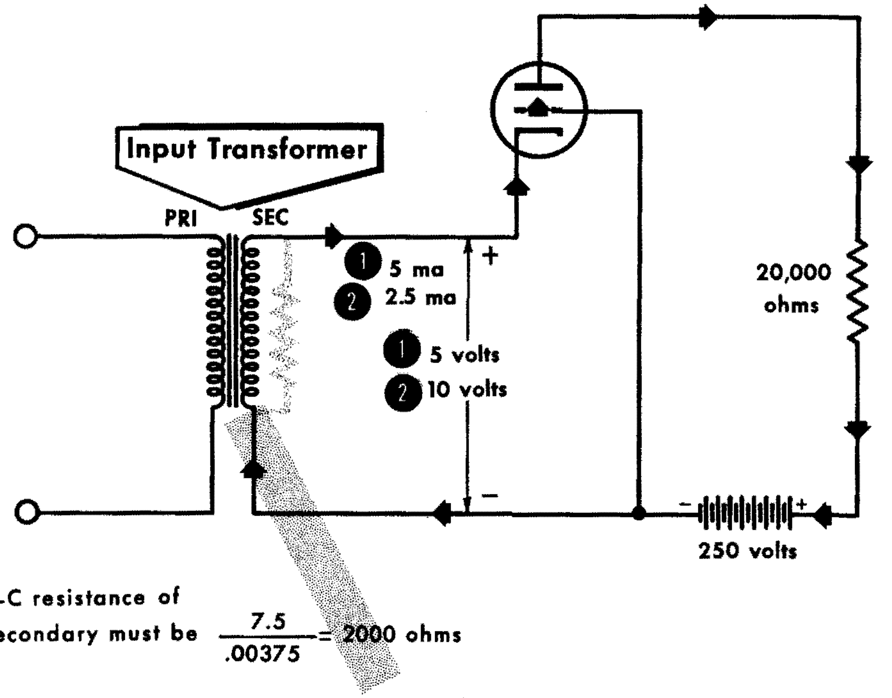
In the grounded-cathode arrangement, there is no current in the input circuit, hence the a-c input resistance is 5/0 or infinity. (Anything divided by zero is infinity.) Here, however, we have what in tube circuits is a low input resistance—in this example, 2000 ohms.



Grounded-grid amplifier

Input to the Grounded-Grid Amplifier

GROUNDED-GRID AMPLIFIER --



Use of an INPUT TRANSFORMER

With the grounded-cathode arrangement, the input circuit only needs to provide the right *voltage*. Current does not matter. In this circuit, however, we have to provide the right current *and* voltage fluctuations.

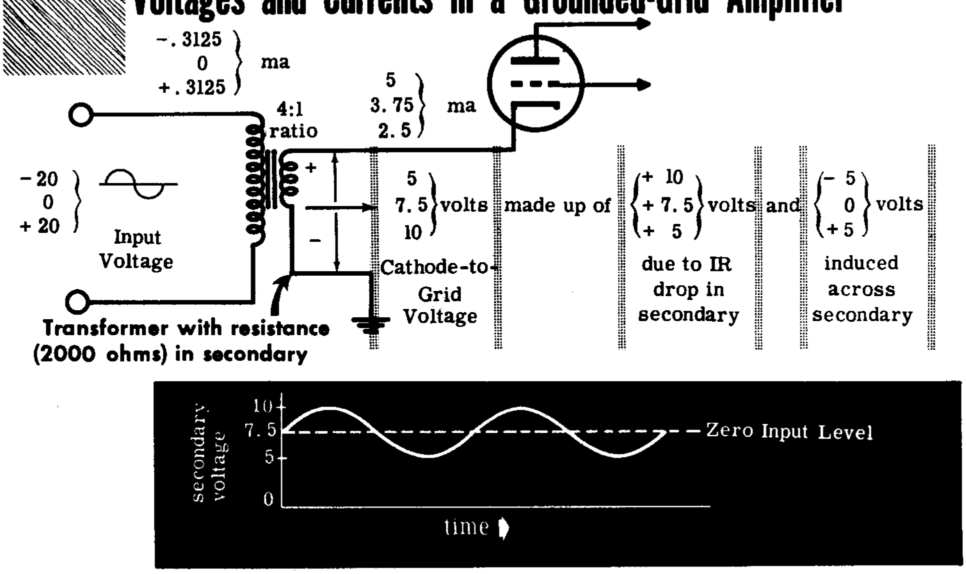
This is achieved by using a transformer. Assuming that the 5- and 10-volt figures are the two extremes of an alternating fluctuation, the average value will be 7.5 volts. Similarly the average current will be halfway between 2.5 and 5 milliamps, or 3.75 milliamps. Hence the d-c resistance in the path from cathode to ground needs to be 7.5/.00375 or 2000 ohms. This can be in the form of the winding resistance of the transformer secondary, or may include a separate resistor.

Input to the Grounded-Grid Amplifier (contd.)

Assuming that the transformer uses a 4:1 step-down ratio, we can work out the conditions of the input circuit. In the secondary of the transformer, there is a current that fluctuates up to 5 milliamperes and down to 2.5 milliamperes. The voltage drop due to this current (when no input is applied) is 7.5 volts. The change in voltage due to change in current will be from 10 volts (at 5 milliamperes) to 5 volts (at 2.5 milliamperes). For proper operation of the tube, however, the cathode-to-grid voltage must be +5 volts at 5 milliamperes and +10 volts at 2.5 milliamperes—just the reverse of what we have. These potentials must be provided by the induced voltage from the transformer.

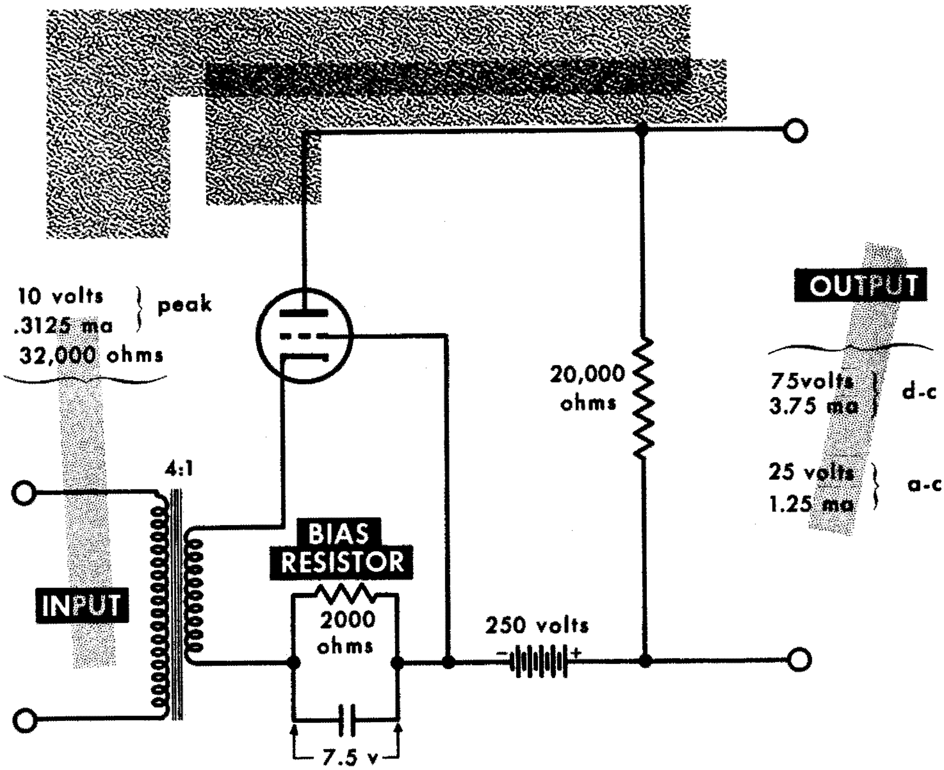
Thus the induced voltage will have to offset the change in voltage drop and also provide the extra voltage for the cathode. It will have to fluctuate 5 volts each way from zero. Of this fluctuation, 2.5 volts in each direction will be taken up by the change in voltage across the secondary resistance due to change in current, and 2.5 volts will change the cathode-to-grid voltage. To produce this induced voltage on the secondary, the primary will need four times as much (or 20 volts) fluctuation each way from zero. It will also have to offset the *change* in magnetization due to the *change* in secondary current. Since this change requires 1.25 milliamperes each way on the secondary, only one-fourth of this (0.3125 milliampere) will be needed in the primary. Thus, the effective primary input must be a 20-volt fluctuation each way, accompanied by a 0.3125-milliampere fluctuation each way.

Voltages and Currents in a Grounded-Grid Amplifier



Input to the Grounded-Grid Amplifier (contd.)

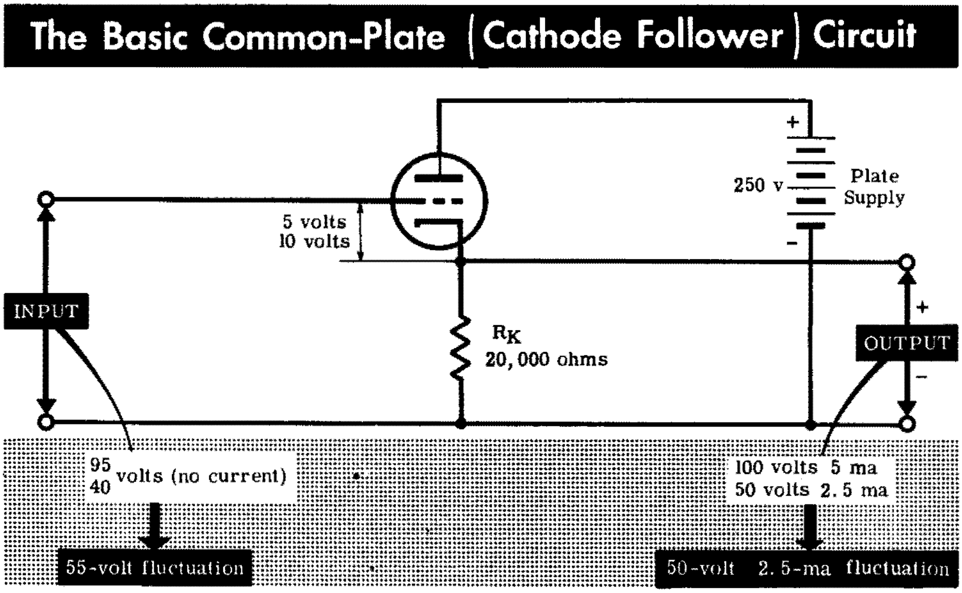
The basic grounded-grid circuit can be improved by using a separate resistor in the cathode connection to provide bias. Its value should be 2000 ohms, so that 3.75 milliamperes provide the correct middle potential of 7.5 volts. A large capacitor is connected across this resistor, so that when the current through the circuit changes, the capacitor will absorb some of it before the voltage can change. If the capacitor is big enough, the voltage will stay almost fixed at 7.5 volts.



Adding a Bias Resistor improves grounded-grid amplifier circuit

This means the secondary has only to provide a voltage fluctuation of 2.5 volts each way and the primary has to receive only a 10-volt fluctuation in each direction. The input resistance thus is 10/.0003125 or 32,000 ohms.

The Cathode Follower



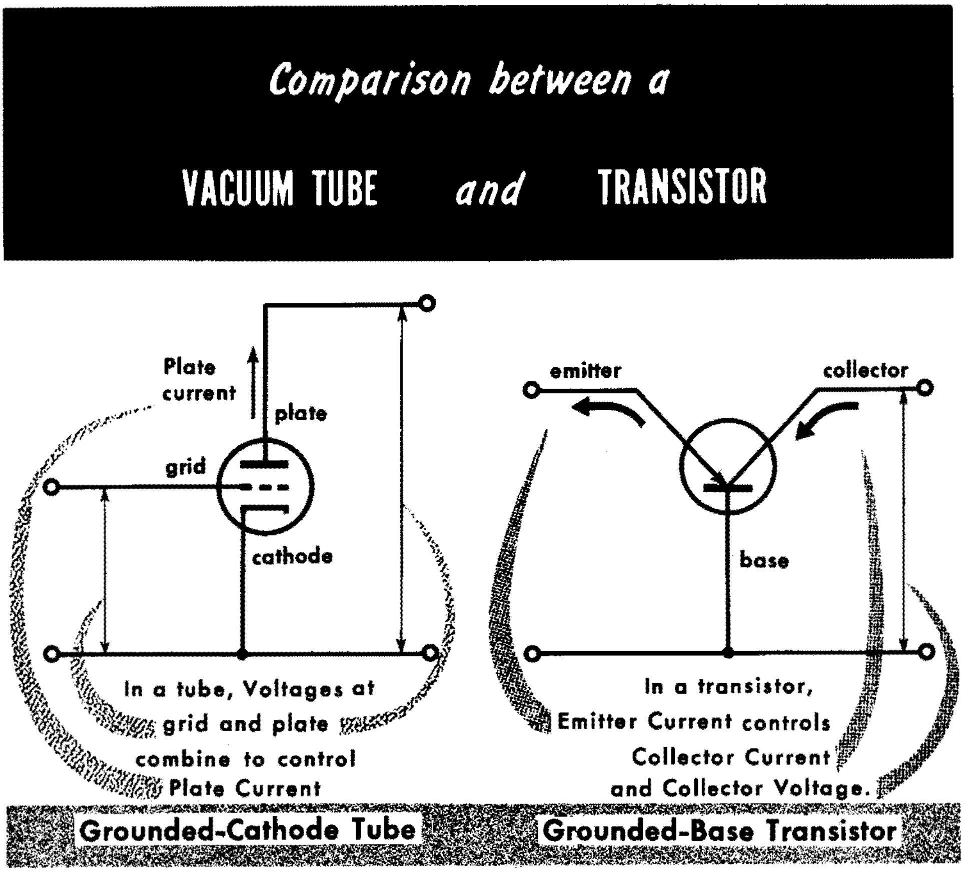
The third way to connect a tube makes the plate common. We do this by connecting the 20,000-ohm resistor in the cathode circuit. Now the change of —5 volts to —10 volts between grid and *cathode* results, as before, in a change in current through the tube between plate and cathode, from 10 milliamperes to 5 milliamperes. Thus the cathode will fluctuate between 100 volts and 50 volts, respectively, due to the different currents flowing in the 20,000 resistor.

Adding these voltages together, we can find what input voltages we have to use to get these output voltages. With —5 volts between grid and *cathode,* the voltage from cathode to ground is 100 volts, so the voltage from grid to *ground* must be (100—5) or 95 volts. With —10 volts from grid to *cathode,* the cathode to ground voltage is 50 volts, hence the grid to *ground* voltage must be (50—10) or 40 volts. Thus, this circuit requires an input fluctuation between 95 and 40 volts (55 volts) to get an output fluctuation of only 50 volts, between 50 and 100. How can this be amplification?

The answer is that the cathode follower does not give *voltage* amplification. Because the grid input does not require any current to cause the output circuit to give a current fluctuation of 5 milliamperes, we can see that this arrangement can be regarded as a *current* amplifier. Because the output voltage at the cathode is almost the same as the input voltage at the grid (although more current fluctuation is available at this voltage), this circuit is called a *cathode-follower,* the idea being that the cathode voltage *follows* the grid voltage.

Transistor Operation

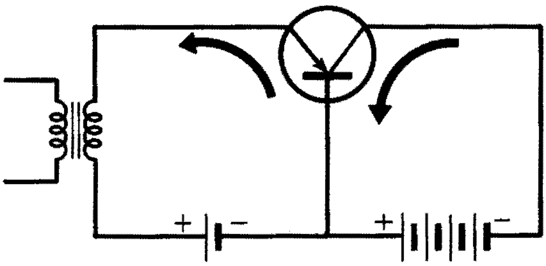
Tubes are not the only devices used for amplification. *Transistors* use an entirely different principle. We could epitomize the operation of a tube by saying that the plate and grid voltages combine to control the plate-to-cathode current. Or alternatively, in a given arrangement, the grid voltage controls the voltage and current in the plate-to-cathode circuit. In a transistor, it is the *current* in the circuit between *emitter* and *base* that controls the voltage and current between *collector* and *base.*



To understand what all this means, we need to know something about a transistor. It is essentially a piece of special alloy, using germanium or silicon as a base, with very small amounts (a few parts in a million, very accurately controlled) of an "impurity" added. Two electrodes are connected to this carefully blended base. They may be point contact "whiskers" or a "grown junction," formed by a process called electrolytic deposition.

Transistor Operation (contd.)

Transistor Operation



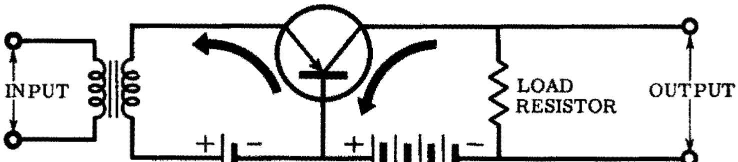
Current in emitter circuit (at low voltage) produces almost identical current flow in the collector circuit (at much higher voltage).

However it is made, the transistor works by having a potential applied to the electrode called the *collector* in the direction *against* the normal flow of current for this kind of device. (Either of the junctions between the electrodes and the base would act as a rectifier allowing current to flow one way but not the other).

Until the *emitter* circuit has current flowing in it, the current in the collector will be very small—almost zero. When current is drawn through the emitter in the direction of easy flow, however, it allows an almost equal flow of current between collector and base, so the total current flowing in the base circuit is quite small compared to the other currents in the circuit. This action can be regarded as the emitter supplying the base with surplus electrons, which the collector can then draw off. The collector current is dependent on the emitter current. It is, of course, also dependent on the collector voltage; if there were no voltage on the collector, no current would flow in the collector circuit.

Amplification is achieved, as in the tube circuit, by connecting a resistor in the collector circuit, so the changes in collector current (which follow the changes in emitter current) produce fluctuations in collector *voltage*. We can, therefore, make an amplifier by having the current in the emitter circuit control the current *and voltage* in the collector circuit.

TRANSISTOR AMPLIFICATION

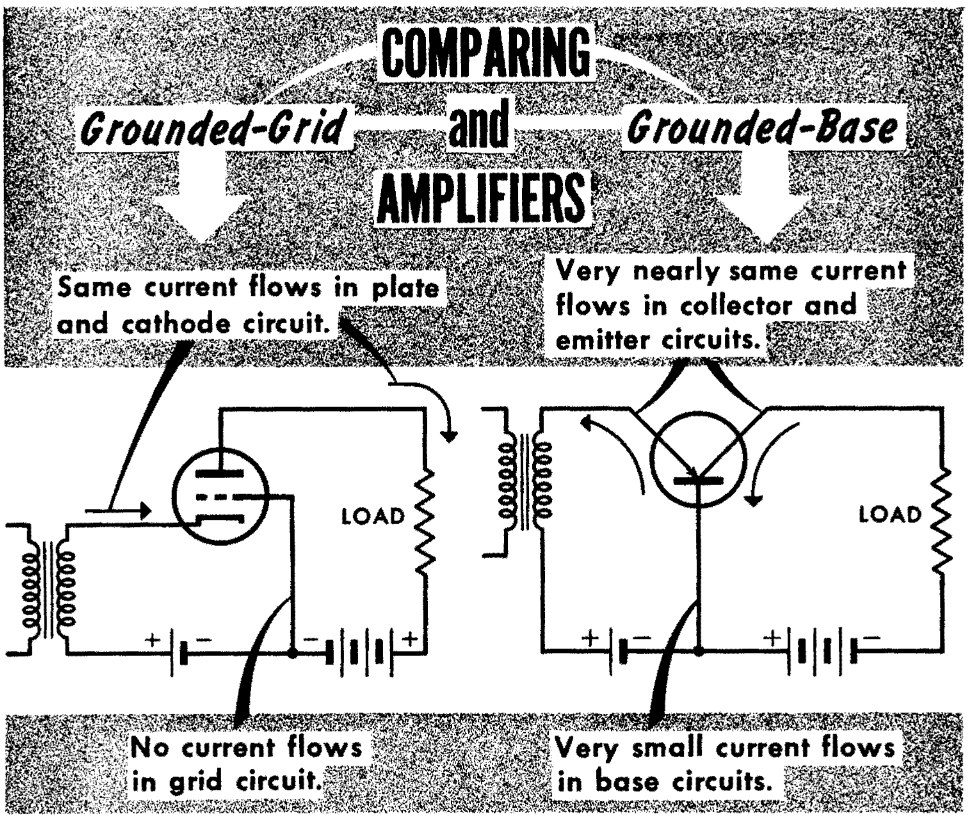


• Current fluctuations in emitter and collector circuit are almost identical
• Emitter voltages are smaller, collector voltages are much larger

The Grounded-Base Amplifier

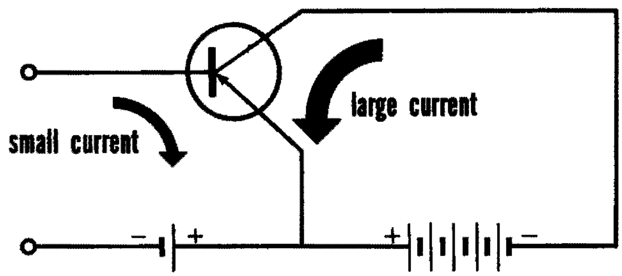
The grounded-base transistor amplifier is very like the grounded-grid vacuum-tube amplifier. In the vacuum-tube amplifier there is no current in the grid connection to ground. In the transistor amplifier, this current is small compared to that in the emitter and collector connections. In addition, only small voltages are needed to cause the necessary current fluctuations in the emitter circuit, whereas (by using a higher voltage and a high resistance in the collector connection) much higher voltages can be obtained at the collector.

This gives us a clue to how the transistor can be used to amplify. If the grounded-grid amplifier is like the grounded-base transistor, we should find that the tube grid is like the base of a transistor, the cathode like the emitter (they even sound similar, because a cathode is used to *emit* electrons), and we must always remember that *voltages* control electron flow in a tube, whereas currents (or electron distribution) control voltage and current effects in a transistor.



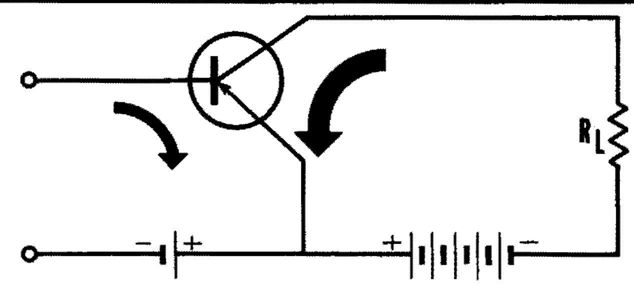
The Grounded-Emitter Amplifier

GROUNDED-EMITTER AMPLIFIER WITH NO LOAD



A small current and voltage change in the base circuit produces a much larger Current Change in the collector circuit.

GROUNDED-EMITTER AMPLIFIER WITH ADDED LOAD RESISTOR R_L



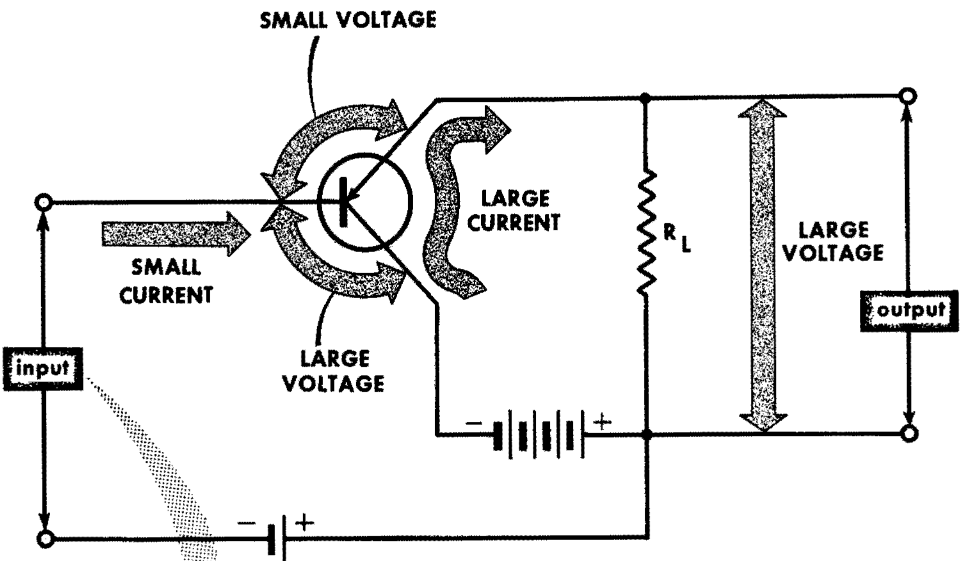
A small voltage and current change in the base circuit produces a much larger Voltage and Current Change in the collector circuit.

By turning the circuit around to what is called *grounded-emitter* connection, the input is applied to the base. Only a small proportion of the emitter current flows through the base connection. With a transistor, the current applied to the base is more important than the voltage. Because of the small size of the base current compared to the collector current, this circuit can be regarded as providing current amplification. In addition, voltage amplification can be obtained by using a resistor in the collector circuit.

The Grounded-Collector Amplifier

The remaining possibility for connection of a transistor—grounded collector—is very like the cathode-follower connection of a vacuum tube. In this arrangement, the voltage between base and emitter is small because it is in the direction of easy current flow. The collector-circuit resistor is connected in the emitter circuit, hence the collector-emitter current fluctuations pass through it, causing an output voltage between the emitter and ground. Because the base is used for the input connection, both the a-c and d-c components of the input current are much smaller than in the output current.

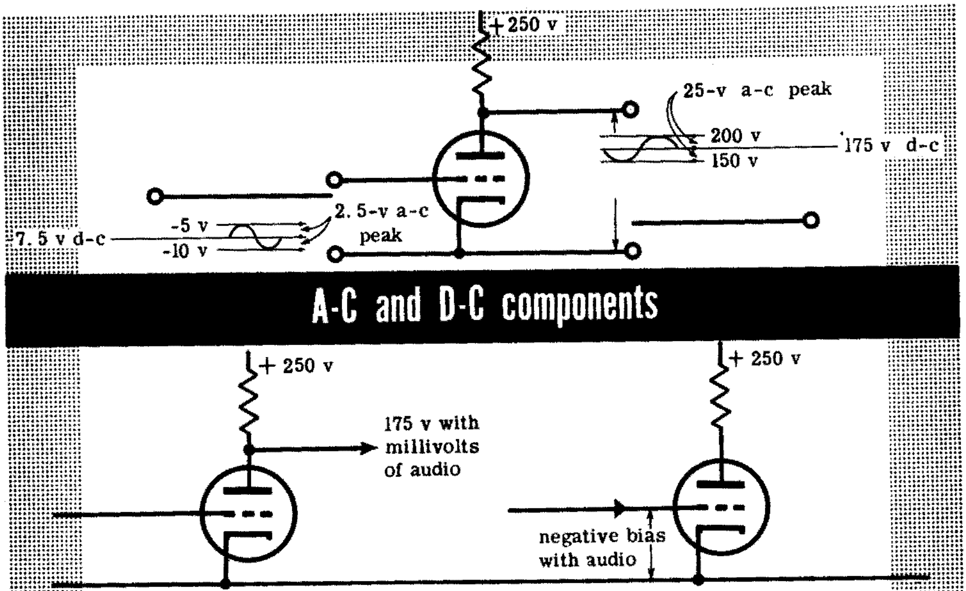
Grounded-Collector Amplifier



A large voltage and small current control almost the same voltage at much larger current

A-C and D-C Components

Thus far we have shown how smaller voltage or current fluctuations can, by using tubes or transistors, cause bigger voltage or current fluctuations. This is the essence of amplification, but one thing more is necessary to be able to make use of amplification.



How do we pass on the millivolts of audio, and give the grid of the next stage the right combination of d-c bias and audio (a-c)?

In the first practical amplifier stage discussed, an input fluctuation of 5 volts at the grid produced an output fluctuation of 50 volts at the plate. We did not pay much attention to the fact that these fluctuations do not conveniently start from zero or the negative bias required by the grid of a following stage. The grid circuit fluctuation could be regarded as being 2.5 volts away from an average (d-c) bias of 7.5 volts. The plate output is a fluctuation of 25 volts each way from an average (d-c) component of 175 volts.

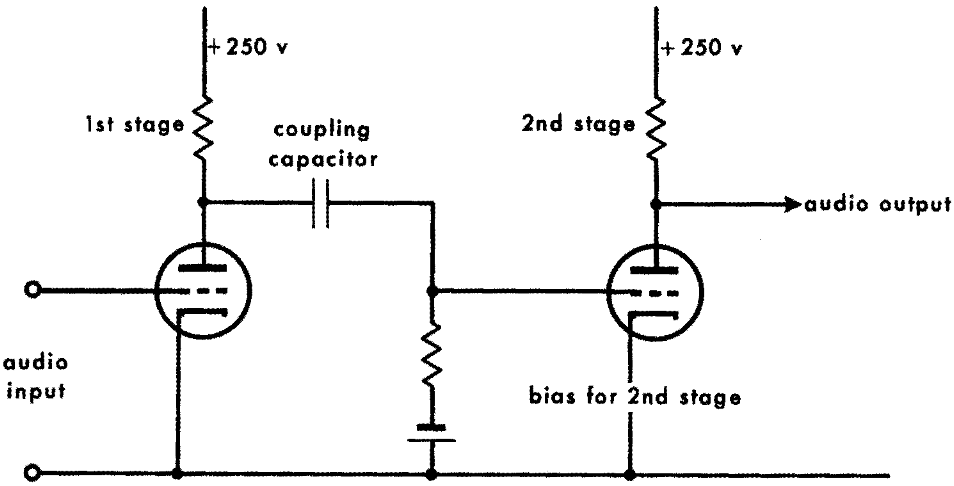
This would, of course, run in proportion. If we start from a microphone, the voltage fluctuations will only be measured in millivolts, or thousandths of a volt. The first stage of amplification would raise this to tens of millivolts—still a rather small signal that would require more amplification to make it useful.

The easiest way to eliminate the d-c components in the output is to use a capacitor. Current does not flow from one plate to the other of a coupling capacitor, but current can flow *to* the plates, producing a charge on them, accompanied by a difference in voltage between them.

Coupling Capacitors

In the absence of fluctuations (audio), the plate or foil of the capacitor connected to the grid resistor takes the grid bias voltage. If there is any charge that would produce a different voltage, it flows away through the resistor, so that the voltage is again equalized. The same thing happens at the foil connected to the plate of the previous tube— it is at the same voltage as the tube's plate.

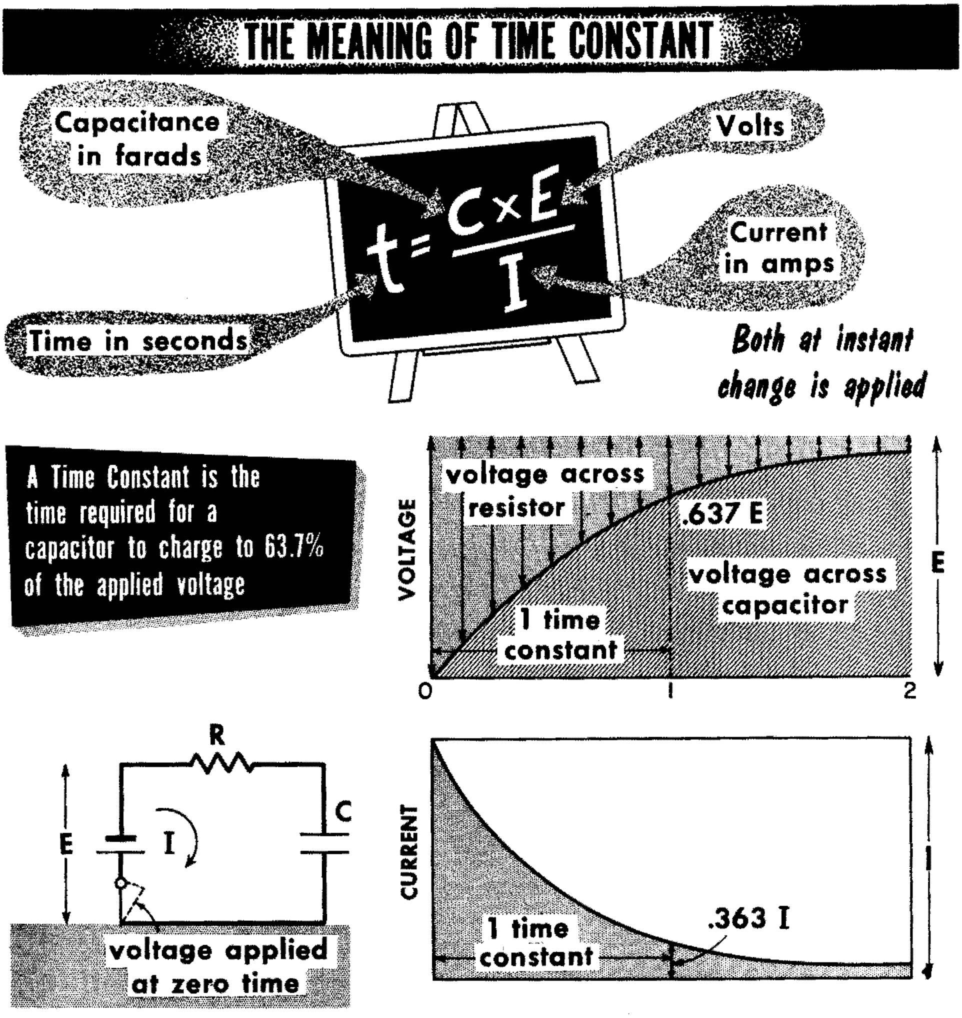
USE OF A COUPLING CAPACITOR



The COUPLING CAPACITOR prevents the d-c voltage from reaching the grid of the next stage while passing on the audio fluctuations

When audio signals come along, the voltage at the plate goes up and down from its steady resting point. Because the grid resistor is so large, the charge on the capacitor does not have time to change, hence the voltage across the capacitor does not change. This means the voltage at the grid fluctuates up and down from its steady voltage in exactly the same way as the previous tube plate. (The capacitor acts as a short-circuit for the fluctuations, while isolating the d-c components at plate and grid.) If the voltage at the previous tube plate changed permanently, the charge on the capacitor would also change, so that the potential on the grid side would be the same as ground potential, while that at the plate side assumed the new plate potential. This process (as we have shown) takes a time dependent on the size (value) of the capacitor.

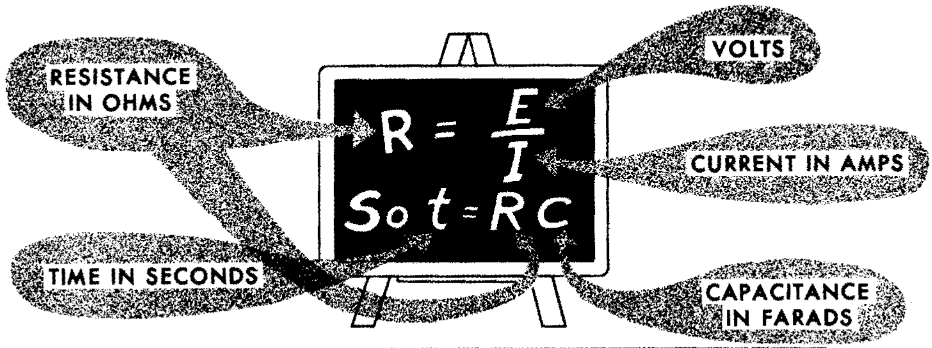
Time Constants



To understand the way in which a coupling capacitor operates, we need to know about time constants. When the voltage between the terminals of a resistance-capacitance combination (such as the coupling capacitor and grid resistor) changes, the charge across the capacitor does not change immediately. Therefore the voltage across the capacitor is initially the same as before the change. Current immediately *starts* to flow in the resistor, because all the change in voltage appears across it. This will determine the initial current according to $I=E/R$. If a current of this size continued to flow, it would take a certain time to change the charge on the capacitor, equalizing the voltage across it. In seconds, the time would be given by $t = C \times E/I$.

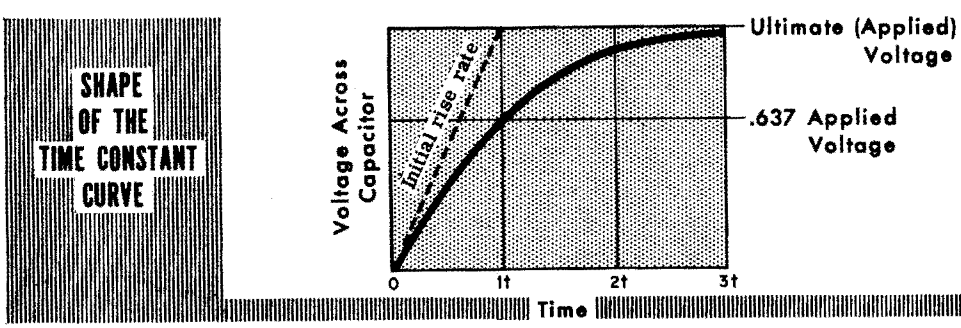
Time Constants (contd.)

Because the current is also determined by the voltage change, E, that started it, the time can be written as t = RC. This now eliminates both E and I from the formula, which shows that the time would be the same whatever the voltage change involved. (A larger voltage change would produce a larger current, hence the time for the charge to change would be the same.)



The time constant is numerically equal to the resistance times capacitance

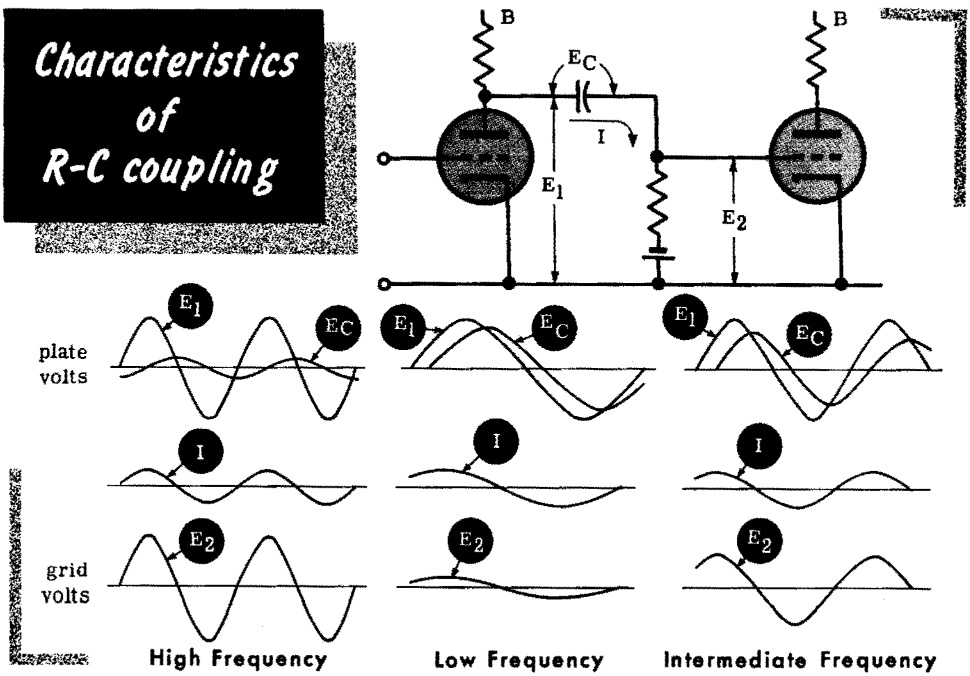
However, the current does not remain constant until the voltage equalizes, and then stops. The flowing of current causes a rise in the voltage across the capacitor and a corresponding fall in the voltage across the resistor. This, in turn, causes the current in the resistor to decrease. The current thus drops off gradually before the change in voltage is complete. In fact, in the time it would take to make the whole change if the starting current were maintained, the change actually reaches only 0.637 of its complete change. In theory it never does quite reach the complete change, because the current keeps falling off indefinitely, and so does the voltage difference.



R-C Coupling

At higher frequencies, the charge on the capacitor hardly changes at all during an audio-frequency fluctuation. At low frequencies, however, there is time for the charge to change, which changes the voltage across the capacitor. If the frequency is low enough, the voltage across the capacitor changes as much as the plate potential, and the potential of the grid of the second tube hardly changes at all. At intermediate frequencies, the voltage at the second-stage grid fluctuates by an intermediate amount.

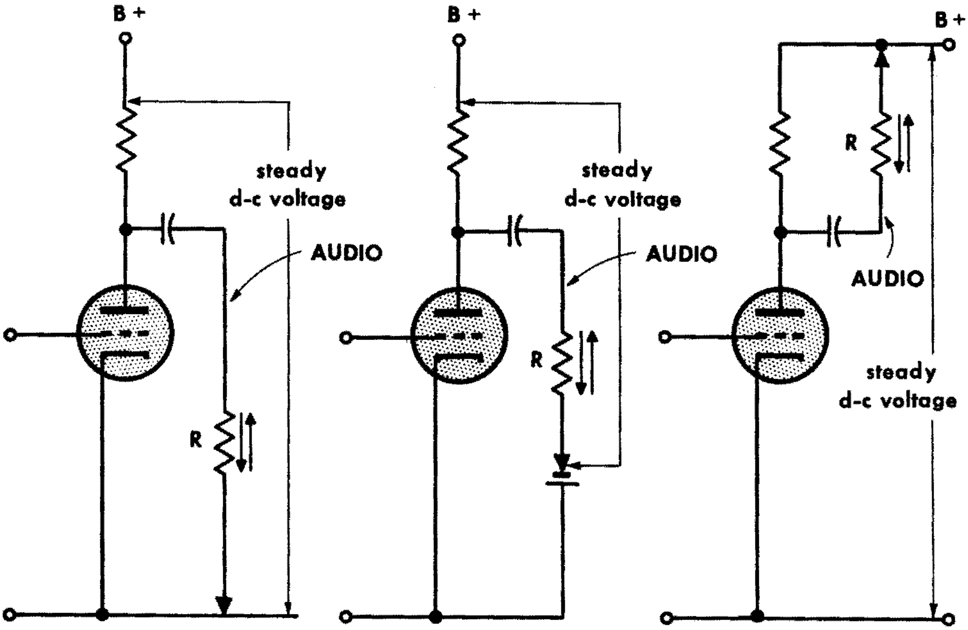
At the higher frequencies, the current through the grid resistor, due to the audio fluctuations, is the same as if the resistor were connected directly to the plate, without the steady d-c voltage difference being there. Where does this audio current come from? The plate circuit of the tube has to supply it. When the plate fluctuates negative, due to momentarily greater current through the coupling resistor, the grid of the following stage goes negative as the result of current flow through the grid resistor from grid to ground, adding to the momentary plate current. We use the term *coupling resistor* for the component that feeds B plus to the plate. Some call it the *plate resistor,* which must be carefully distinguished from plate resistance. Many call it the *load* resistor, which can be misleading. As we have just seen, current fluctuations related to stage amplification divide between this resistor and the grid resistor coupled to it by the coupling capacitor. So, at most frequencies, the *load* for the tube's plate is these two resistors effectively in parallel. We will call this the *load resistance.*



R-C Coupling (contd.)

Because this increase in plate current is controlled by the voltage applied to the grid, this extra current means that the rise in current *through the coupling resistor* will not be so great as before the capacitor and grid resistor were connected. The effect is the same as if the grid resistor were connected in parallel with the load resistor.

Taking Audio Fluctuations from the Plate Circuit



No matter where the other end of resistor R is connected, it takes the same current and voltage fluctuations from the plate

Whether the resistor to which the output side of the capacitor is connected goes to ground or to B+ will only make a difference to the charge on the capacitor and the steady voltage across it. The *fluctuations* across the capacitor and the audio currents will be the same either way because B+ is always a fixed voltage difference from the ground.

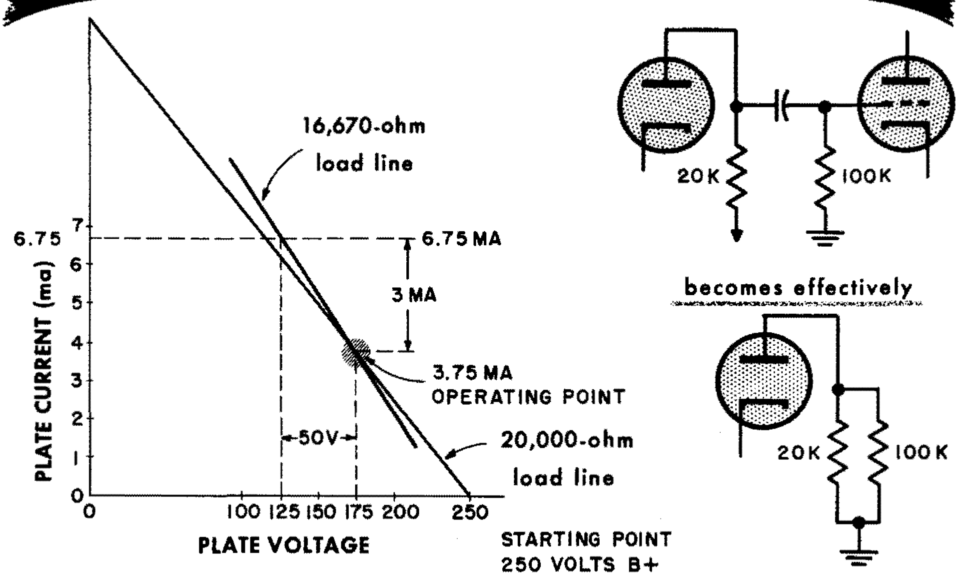
Constructing The New Load Line

We can apply these facts to the load line. The actual load resistor connected between B+ and the plate controls the steady operating point, according to the steady grid bias voltage. This operating point is found by using the load line corresponding to the coupling resistor, starting from the B+ voltage used.

Audio fluctuations cause the plate current and voltage to fluctuate in a manner that can be indicated by drawing a load line through the operating point, at a slope representing the combined resistances of the coupling resistor and grid resistor in parallel.

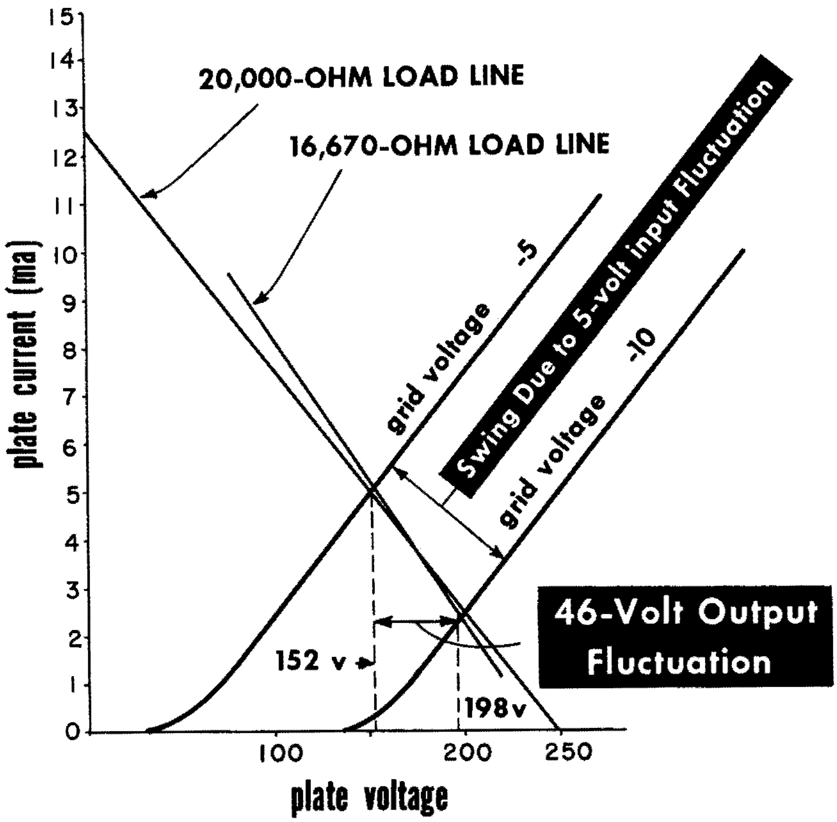
In the example we used before, the plate coupling resistor was 20,000 ohms. If we use a grid resistor of 100,000 ohms, the *effective* resistance of these two working in parallel for (audio only) is (20,000 $\times$ 100,000)/(20,000 + 100,000) or 16,670 ohms. Drawing a line representing this resistance, through the 175-volt/3.75-milliampere operating point, can be achieved as follows: three milliamps through 16,670 ohms will produce a voltage drop across it of 50 volts. Hence 6.75 milliamps on this load line will correspond with 175—50 or 125 volts.

Constructing the Dynamic Load Line



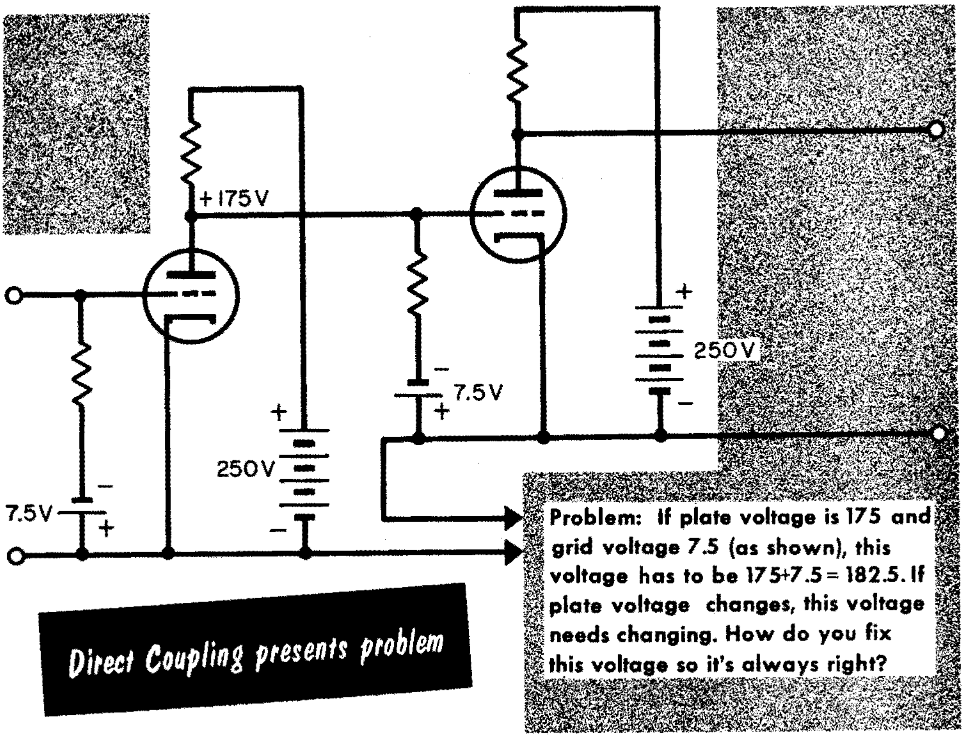
Constructing The New Load Line (contd.)

THE COMPLETE DYNAMIC LOAD LINE



Joining these points and extending the line we can find the new values of plate voltage corresponding to fluctuation of grid voltage between —5 and —10. The new voltages, instead of 150 and 200, as on the 20,000-ohm load line, are now 152 and 198 volts. Thus a 5-volt input fluctuation yields a 46-volt output fluctuation. The gain is 46/5 or 9.2, instead of the 10 obtained before.

Direct Coupling



An alternative to resistance-capacitance coupling is direct coupling. This system requires several separate supply voltages, however, and there is a problem in getting them all set—and maintained—at their right values. For example, if the working plate voltage of the first tube is 175 volts and the next stage needs 7.5 volts bias, the second-stage cathode must be exactly $175 + 7.5 = 182.5$ volts positive from the first-stage cathode. This voltage must be maintained in addition to the *voltage supplies* needed to provide plate current.

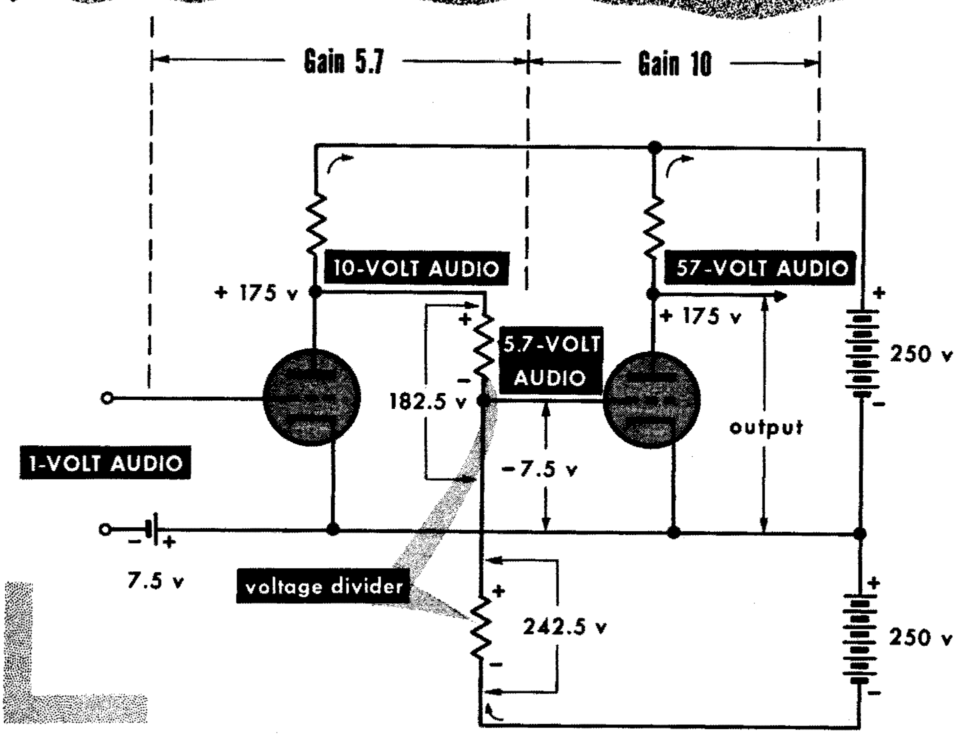
If the plate voltage should happen to be 170 volts and the second cathode is still positive 182.5 volts, its bias will become 12.5 volts, which is too much. When more stages are added to this system, a very small error in voltage at the input end can result in later stages being biased completely out of operation.

The problem then is that a number of supplies are needed, and they have to be very accurately controlled to the right value—which may vary with room temperature or a number of other things.

Such designs have been developed for computor applications, which are not audio and therefore will not be discussed here.

Direct Coupling (contd.)

A TWO-STAGE AMPLIFIER USING RESISTANCE COUPLING



Another system that uses resistance coupling has just two supplies, one positive and one negative. To bring the voltage at one plate down to a suitable level for the following grid, two resistors are used as a voltage divider. Of course these resistors will divide the available fluctuation at the plate as well as dropping the steady voltage.

Suppose the gain of the previous stage is 10 (as calculated at first), and the plate potential is 175 volts. If the negative supply is 250 volts, like the positive supply, the resistors that give the correct steady voltage for the following grid must give —7.5 volts at their junction, or be in the ratio of 182.5 : 242.5. The fluctuations will be divided by $(182.5 + 242.5)/242.5 = 1.75$. The gain from the grid of the first stage is now 10/1.75 or 5.7, being first multiplied by 10 in the tube, then divided by 1.75 in the resistance coupling. Resistance-capacitance coupling gives better results than direct coupling. Its main disadvantage is that it ceases to be effective at some low frequencies, fixed by the time constant of the resistor-capacitor combination.

QUESTIONS AND PROBLEMS

1. Why is grounded-cathode operation referred to in this book as the basic method of connecting a tube?

2. How does a grounded-grid amplifier differ from a grounded-cathode amplifier? Does the load resistance in the plate circuit affect the performance of either circuit?

3. Why is it advantageous to bypass the bias resistor in a grounded-grid amplifier? Illustrate with numerical example.

4. What is a cathode follower, and why is it so called? Explain how the cathode follower provides a form of amplification.

5. A certain tube, operated at a grid bias of -1 volt, with a plate load resistor of 220,000 ohms, has a gain of 60. Calculate (a) its input resistance, operated grounded grid, and (b) its voltage gain, operated as a cathode follower.

6. Compare the essential control properties of a tube and transistor. What would you regard as the basic mode of operating each of them?

7. How can a transistor be operated to give characteristics very similar to a tube; with what essential difference?

8. Why is some form of coupling circuit necessary in an amplifier using two or more stages? What is the most common form of coupling?

9. How would you explain the concept of the *time constant?* Why is it not the full time taken for a capacitor to reach its final charge?

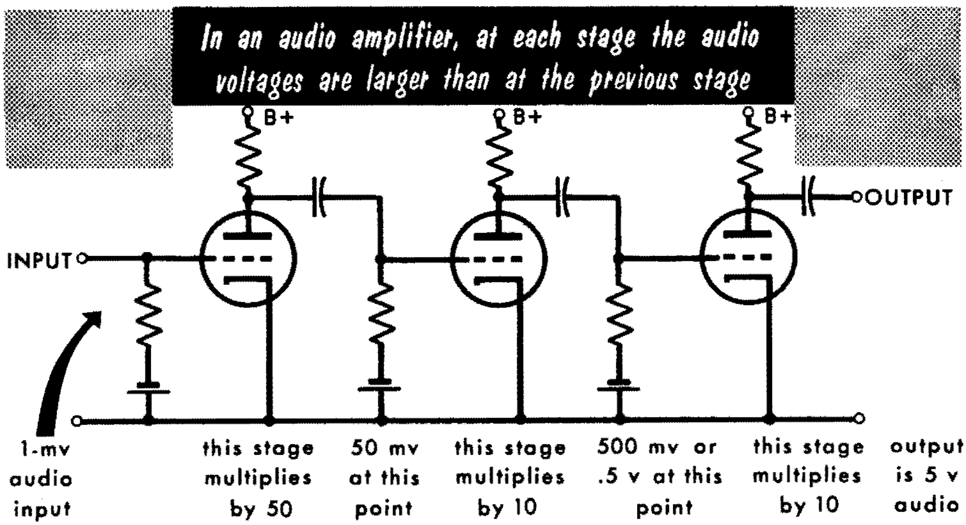
10. What properties of amplifier circuits in general are responsible for (a) low-frequency response, (b) high-frequency response?

11. How does an a-c load line differ from a d-c load line? A tube with a plate resistance of 60,000 ohms using a load resistor of 120,000 ohms gives a gain of 54. Calculate (a) its amplification factor; (b) its gain when coupled by a capacitor to an additional resistance of 200,000 ohms.

12. What is direct coupling? Describe two possible ways of using it.

Problems of Amplification

To make up an amplifier we have to see that each tube, or stage of amplification, can handle the audio fluctuations expected of it. We saw earlier that there is a limit to the maximum voltages an amplifier can handle. This is particularly important towards the later (output) stages of an amplifier. (At each stage the audio voltages are larger than at the previous stage, so the biggest audio voltages are encountered near the output end.)



There is also a limit to how *small* a voltage can be amplified. This is because of *noise*. In any circuit, even when there is not supposed to be any signal (audio), the natural electronic agitation going on in all matter causes a basic random fluctuation, called noise. According to physical theory, every molecule of matter is in a stage of agitation, the amount of which depends on its temperature. As we do not operate electronic equipment in a temperature of absolute zero (—273°C) we always have this agitation to contend with.

In every amplifier circuit, various kinds of noise set a lower limit to the audio voltages that can be successfully amplified. If the noise produces, say 2 microvolts (2 millionths of a volt), then audio voltages lower than this will get "lost" in the noise. In fact a good margin *above* the noise should be available for quality reproduction.

If, at a certain stage in an amplifier, the noise level is 2 microvolts and the maximum level before the amplifier distorts the waveform is 2 volts, the *dynamic range* or the ratio between maximum and minimum levels that can be handled is 1,000,000 to 1, or 120 decibels. (1,000,000 to 1 in voltage or current ratio is equivalent to 1,000,000,000,000 to 1 *power* ratio.)

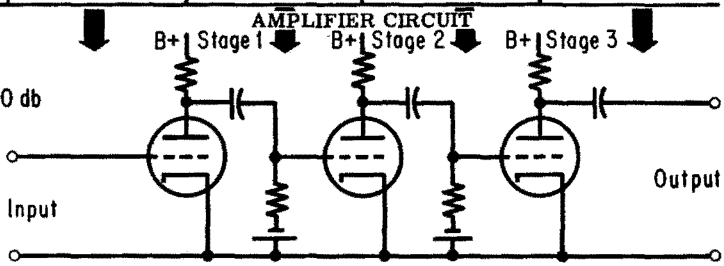
Noise and Dynamic Range (contd.)

MAXIMUM AND MINIMUM LEVELS OF DISTORTION AND NOISE

Noise level	1µv	5µv	10µv	50µv
Distortion level	300 mv	2 v	5 v	50 v
Dynamic range of stage	110 db	112 db	114 db	120 db
Lowest audio	1µv	50µv	500µv	5 mv
Highest audio	10 mv	500 mv	5 v	50 v
Margin over noise	0	20 db	34 db	40 db
Margin below distortion	30 db	12 db	0	0

AMPLIFIER CIRCUIT

Dynamic range of whole amplifier = 80 db



Each stage of an amplifier will have its own maximum and minimum levels, fixed by distortion and noise, respectively. Each stage will have different limits. Also, the audio level corresponding to maximum and minimum through the whole amplifier will be different. For example, a three-stage amplifier may have successive voltage amplifications of 50, 10, and 10, making a total amplification of 5000. At each place from input of the first stage to output of the last stage, there will be a point at which distortion would start and a level of noise *for that circuit*. In short, each stage will have its own dynamic range, which may vary from stage to stage, but will be more than 100 db at each point.

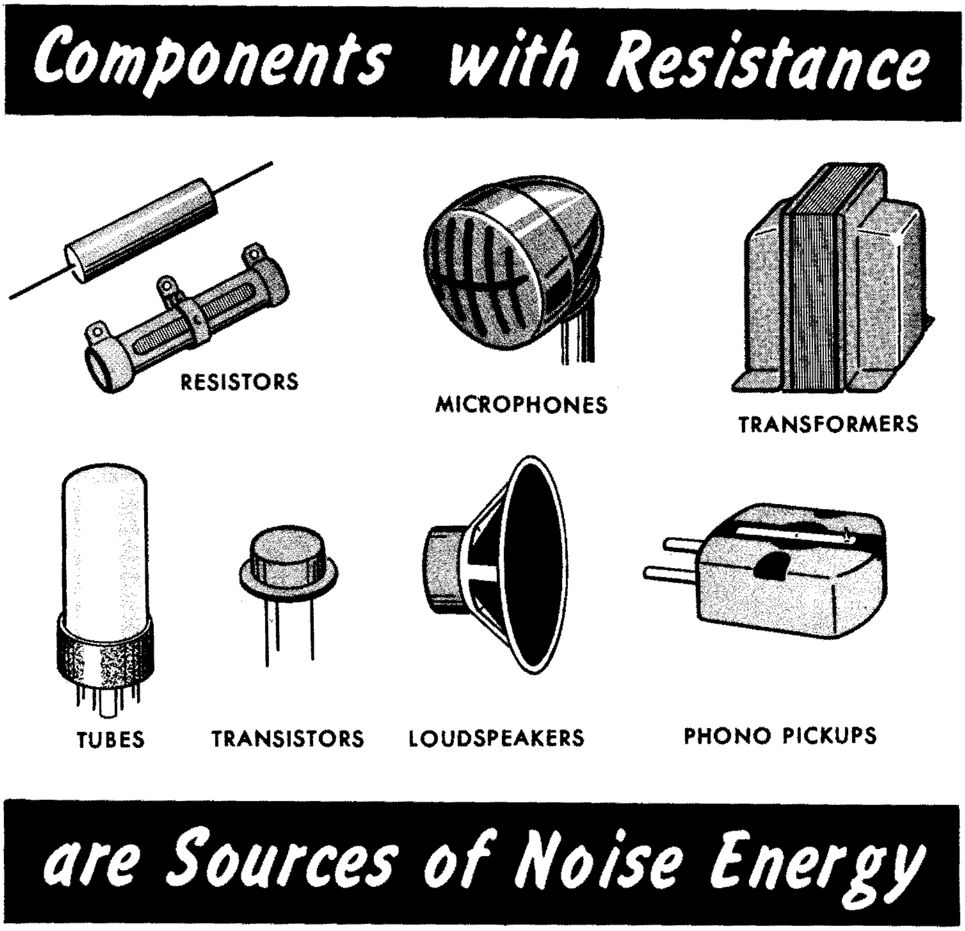
But if we start at the lowest level, the noise level, of the first stage and amplify up (say it is 1 microvolt), the output will be a minimum of 5 millivolts (mv), although the noise level for that stage is only 50 microvolts (µv). Due to the amplification, each stage progressively gets a bigger margin above the noise level, for lowest audio amplified.

Now if we start at the maximum level for distortion at the output—say, 50 volts—and work backward, we find this needs an input of 10 mv, although the first stage will accept 300 mv before distortion. As we work back from the output, we find a bigger margin between the maximum audio signal in the amplifier as a whole and the maximum for that stage.

So, although each stage by itself has a dynamic range of more than 100 db (the lowest is 110), the whole amplifier is limited by the distortion of the output stage and the noise of the input stage. The ratio between lowest and highest audio in the amplifier as a whole is 10,000:1, or 80 db. In a very high gain amplifier, an overall dynamic range of 60 db would be good.

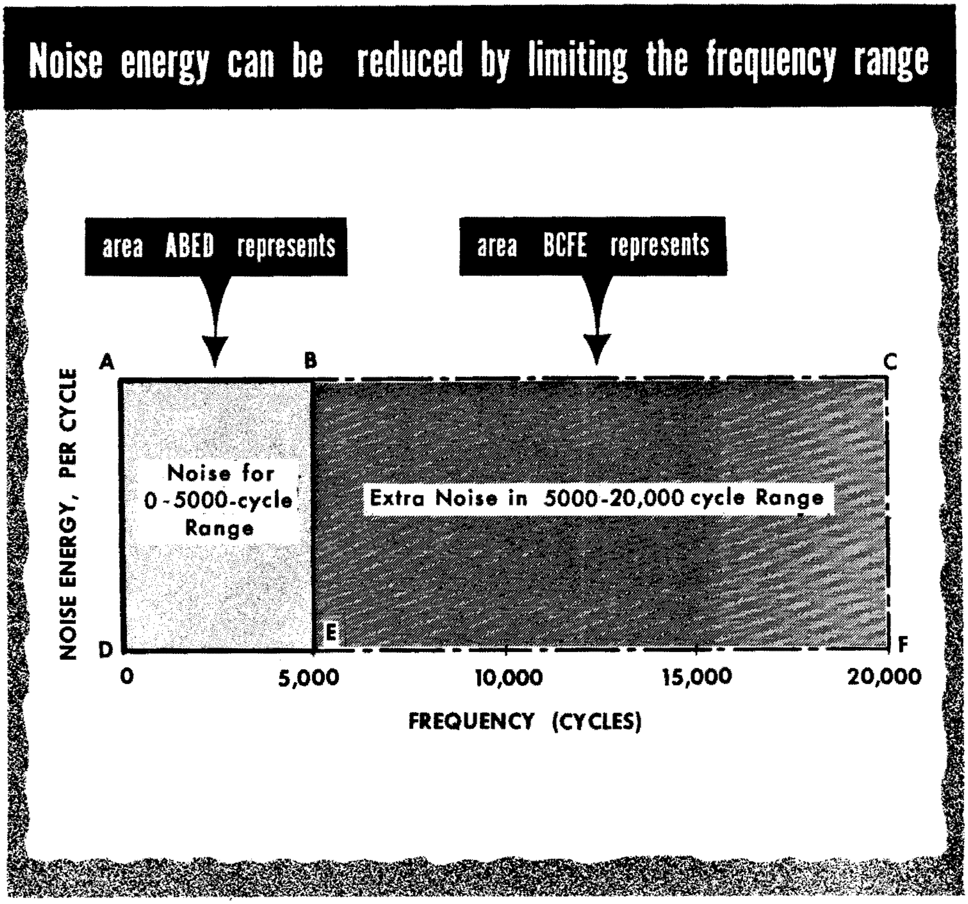
Sources of Noise

Anything that is in an electrical circuit can be a source of noise. In most components the effective noise is in some way related to their *resistance.* A resistor, regardless of its dimensions, produces a certain noise energy, dependent on its temperature. Any resistance produces noise in this way—the resistance of transformer windings, of the coils in microphones and pickups, and so on.



As most audio equipment is operated at room temperature and the noise voltages cannot be measured very precisely anyway, we need only deal with an average figure. The energy output depends on the overall range of frequencies involved. If we use measurements that include frequencies from 0 to 20,000 cycles, the energy will be twice that for frequencies from 0 to 10,000 cycles.

Noise and Bandwidth



At an average room temperature, the noise energy of any resistor is .0165 micromicromicrowatt, or .000 000 000 000 000 000 0165 watt, per cycle. If we take the bandwidth as 10,000 cycles, the noise energy will be 165 micromicromicrowatts. Once we know the value of the resistance, we can calculate the noise *voltage* it will produce, by the usual formula, $E = \sqrt{P \times R}$. For example, for a 1-megohm resistor (1,000,000 ohms), the noise voltage for a bandwidth of 10,000 cycles is:

$$\sqrt{.0165\ \mu\mu\mu\ \text{watt} \times 10{,}000 \times 1{,}000{,}000\ \text{ohms}} = \sqrt{165\ \mu\mu\ \text{volts}} = 12.9\ \mu\ \text{volts}.$$

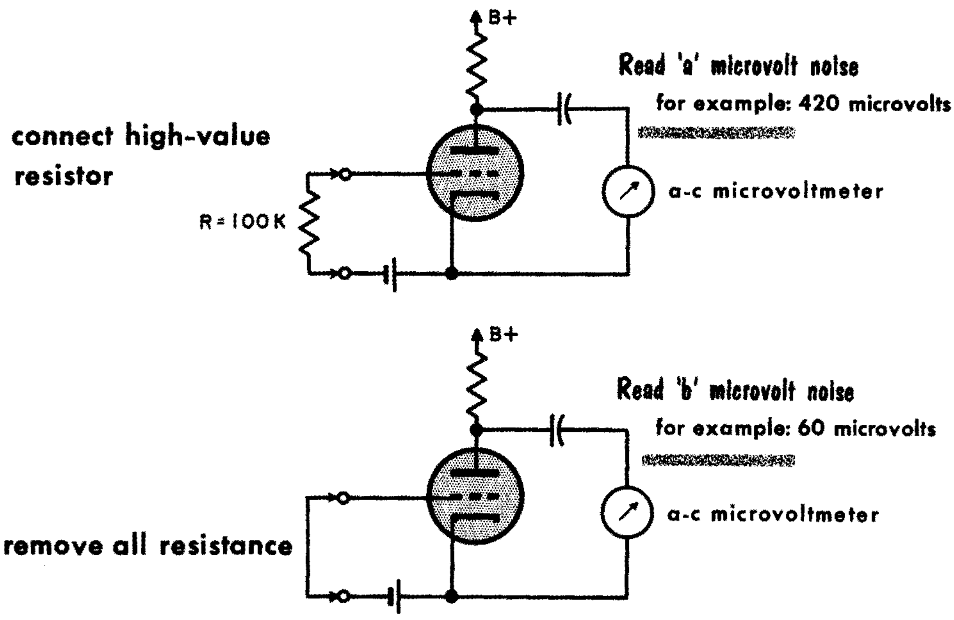
This suggests one convenient way to reduce effective noise when that is a severe problem. By reducing the frequency range at the high end, say from 20,000 cycles to 5,000 cycles, we lose 2 octaves of audio (and the top one doesn't have much in it anyway) and we reduce the noise in the ratio of 4 to 1, or 6 db. This technique is useful for communications work.

Noise in Tubes and Transistors

Both tubes and transistors generate noise components. For a tube this can be measured and calculated as an effective resistance value connected between grid and cathode. Changing the amplified bandwidth will modify the noise voltage in exactly the same way as it does with an ordinary resistor. The value of resistance calculated from measurements shown serves to give the effective noise the tube will produce, however the circuit happens to be arranged.

In transistors, the calculations for noise source are not as simple as in tubes. Connection between emitter and base and between collector and base each are associated with a resistance value inside the transistor. Each of these effective resistances causes a noise component, in the way just discussed, although neither of them is a constant resistance value. The input and output resistance are dependent on each other and dependent on the temperature of the transistor at its critical junction point. This temperature, in turn, is dependent on the combined currents, which produce a small amount of heat. These fluctuations also cause noise, which is predominant at low frequencies.

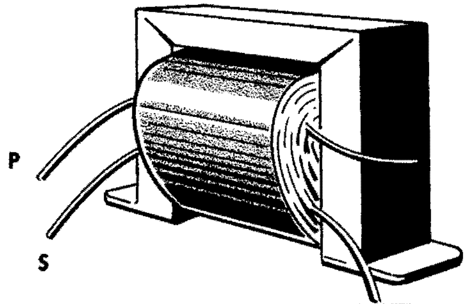
MEASURING A TUBE'S EQUIVALENT NOISE RESISTANCE



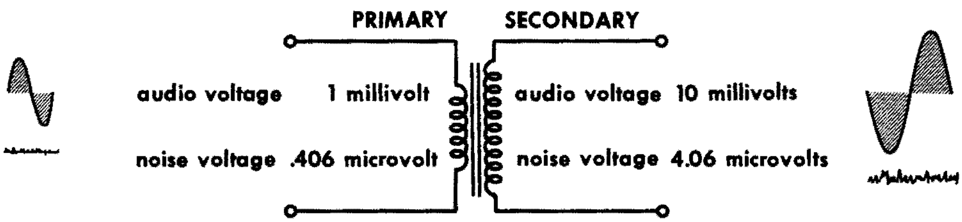
$$\text{EQUIVALENT NOISE RESISTANCE} = R \times \left(\frac{b}{a}\right)^2 = 100{,}000 \times \left(\frac{60}{420}\right)^2 = 2040 \text{ OHMS}$$

The Effect of a Step-Up Transformer

The Audio Transformer steps up the *Noise Voltage* as well as the *Audio Voltage*



typical audio transformer (step-up)



STEP-UP AUDIO TRANSFORMER

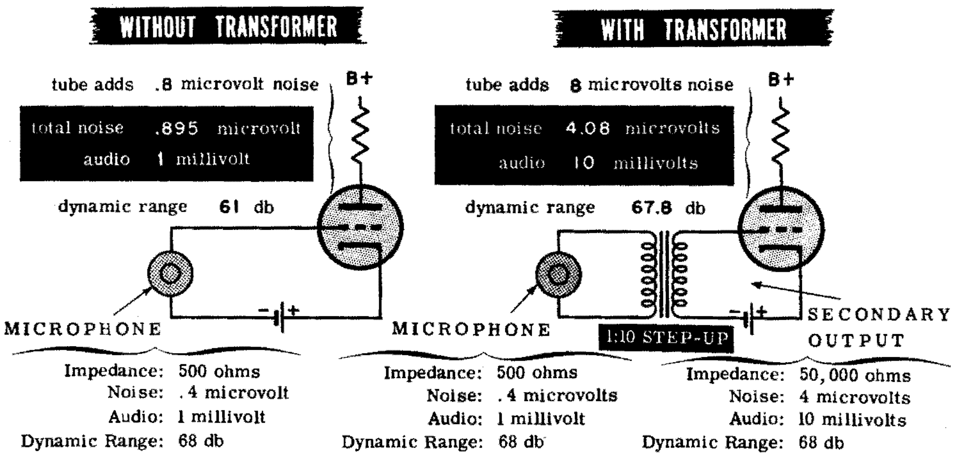
When a transformer is used to step up the audio voltages from a microphone or pickup, it steps up the noise voltage that comes with it. Suppose the audio voltage is 1 millivolt, and the resistance of the microphone or pickup is 500 ohms. In a bandwidth of 20,000 cycles (which we expect for high fidelity), the noise voltage from 500 ohms is

$$\sqrt{.0165 \text{ micromicromicrowatt} \times 20{,}000 \times 500 \text{ ohms}}$$
$$= \sqrt{.165 \times \text{micromicrovolts}} = 0.406 \text{ microvolt}$$

If we use a step-up transformer to raise the 1 millivolt to 10 millivolts, the voltage is stepped up 10 times, the current down by 10:1, and total power transferred remains the same. This means the effective resistance (voltage divided by current) is multiplied 100 times. The 0.406-microvolt noise signal will be stepped up with the 1 millivolt audio to become 4.06 microvolts of noise. In this statement, we have not considered any possible noise the transformer may add of its own, due to the resistance of its windings and core magnetization effects.

Input Transformers

An input transformer is used to improve the signal-to-noise ratio in certain circumstances. If the noise resistance of a tube is 2000 ohms and the impedance or resistance of a microphone is only 500 ohms, the predominant noise when the microphone is connected in directly will be due to the tube. In fact, the tube will produce twice the noise voltage, for a given bandwidth, that the microphone does.



Input Transformer improves signal-to-noise ratio

With the microphone producing 0.4 microvolt of noise, and the tube 0.8 microvolt of noise, the two add together by the process of "the square root of the squares"; 0.4 squared is 0.16 and 0.8 squared is 0.64. Adding these figures gives 0.8, the square root of which is 0.895 microvolt with an audio voltage of, say, 1 millivolt.

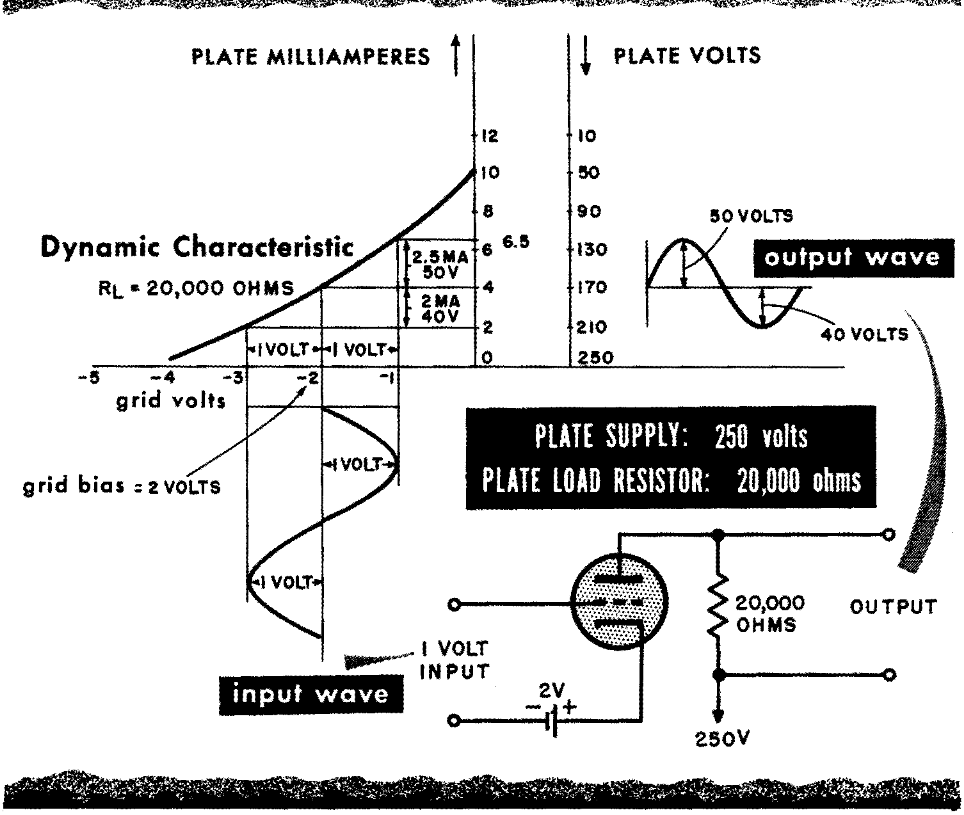
By stepping up the audio and microphone noise with a 10:1 input transformer, the microphone gives 4 microvolts of noise on the transformer secondary. The tube still gives 0.8 microvolt of noise. Now the total noise voltage is $\sqrt{4^2 + .8^2} = \sqrt{16 + .64} = 4.08$ microvolts, which is very little more than the microphone alone. The audio voltage, however, has been stepped up to 10 millivolts in place of the original 1 millivolt. Thus signal-to-noise ratio has been improved by better than 2:1. Beyond this point, however, using a transformer with greater step-up will not improve matters, because it will step up the audio and the noise in the same ratio. In addition, making a transformer with a big step-up gives problems in getting uniform response over a wide range of frequencies.

Nonlinear Amplification

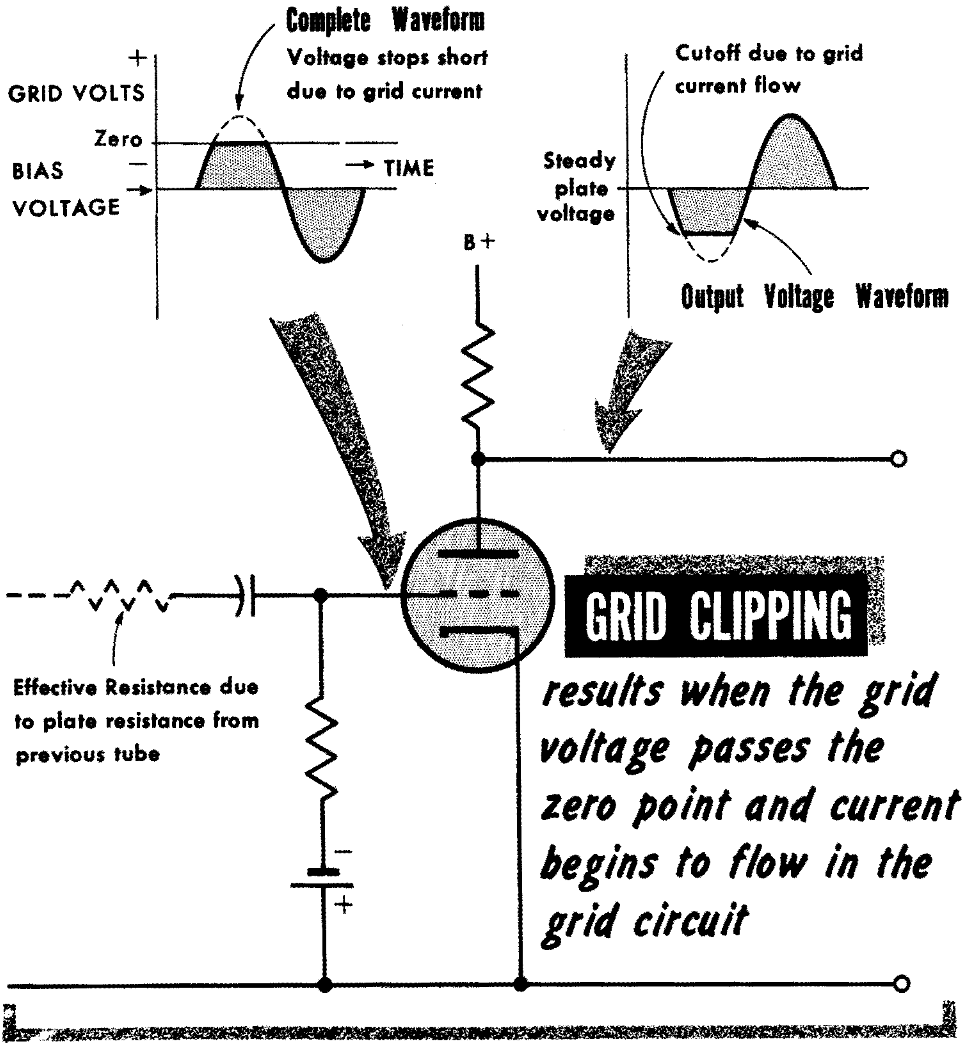
Grid voltage/plate current *curves* are always that—curved. Because of this curvature, the same change in input voltage will not always cause the same change in output current and voltage. For example, a negative change of 1 volt may produce a 2-milliampere change in plate current, accompanied by a 40-volt change in plate potential (using a plate load resistor of 20,000 ohms); a positive change of 1 volt may produce a 2.5-milliampere change in plate current, accompanied by a 50-volt change in plate potential.

The positive half of a wave will thus get amplified more than the negative half. At the input both positive and negative halves of the wave measure 1 volt, but at the output the negative half is 50 volts, and the positive half is 40 volts. This distortion of the wave is due to the curvature or nonlinearity of the tube characteristic.

Nonlinear amplification causes distorted output



Clipping

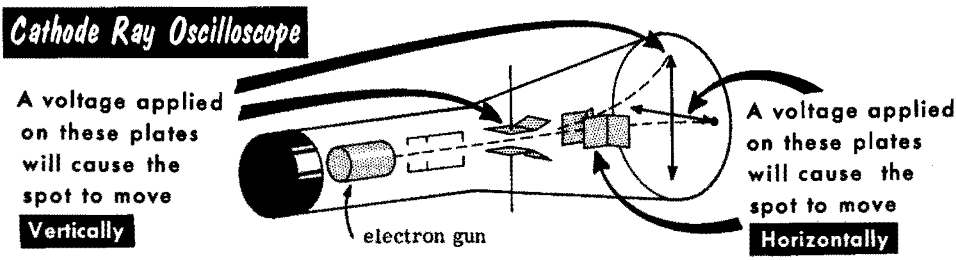


Another kind of distortion occurs because the input fluctuations abruptly cease to produce a corresponding output fluctuation. The most common cause of this is known as *grid clipping*. Normal amplification takes advantage of the fact that, when the grid has a negative voltage, no grid current flows. When the grid passes the zero point between negative and positive, however, grid current *will* start to flow. Thus, when the grid is negative, no current flows in the grid circuit, and the applied signal is undistorted. When the grid reaches the zero point, however, it starts to draw current, and "short-circuits" positive-going voltage beyond that point.

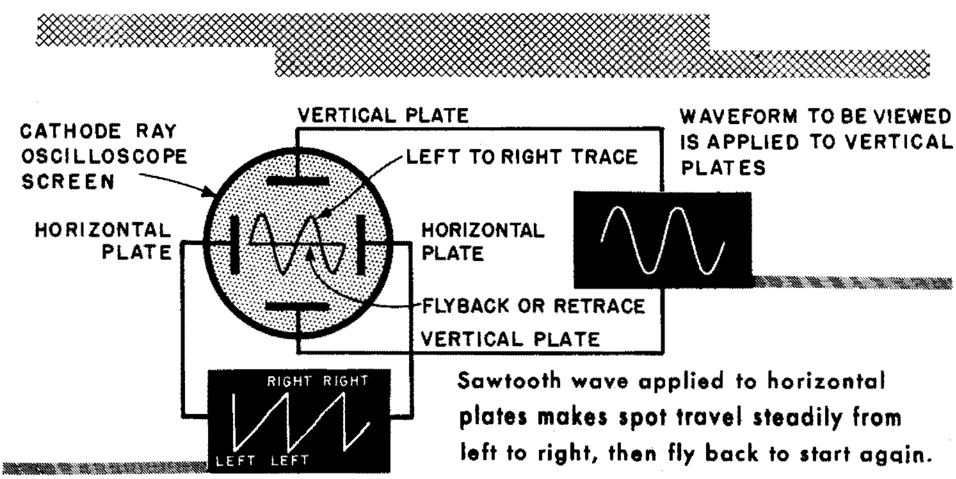
Showing Waveforms with the Oscilloscope

Waveforms can actually be examined by means of the instrument known as an oscilloscope. It uses a special tube in which a beam or "pencil" of electrons is focused to a point on a fluorescent screen that glows with the impact of the electrons. Two pairs of deflecting plates bend the beam in accordance with the voltages applied to them.

Applying different voltages to the plates at the sides of the beam will move the spot sideways, and voltages applied to the upper and lower plates deflect the beam up or down. If different fluctuating voltages are applied to both pairs of plates, the spot will trace a pattern on the screen representing the combined effect of the two voltage fluctuations.

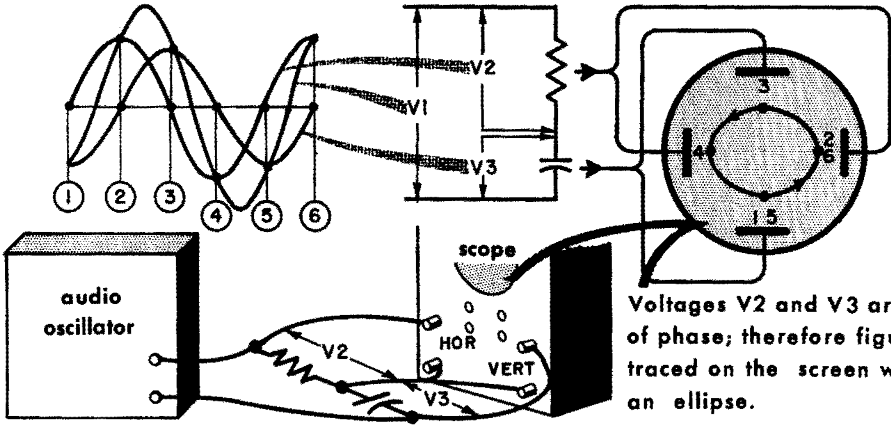


If the fluctuation applied to the side plates follows a "sawtooth" waveform, the spot will move steadily from left to right across the screen, and then rapidly return to its starting point. By using this waveform as a "time-base" in this way, the voltage applied to the vertical plates traces its own waveform. In this way we can see the curvature distortion or clipping just discussed.



Sawtooth wave applied to horizontal plates makes spot travel steadily from left to right, then fly back to start again.

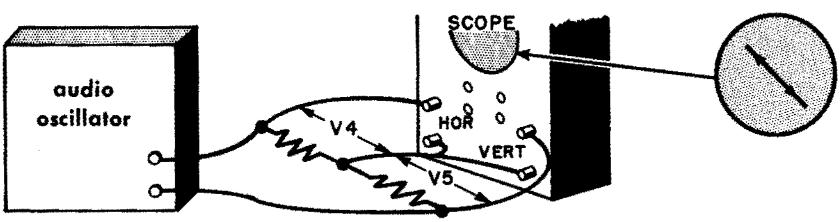
Showing Simple Patterns with the Oscilloscope



Voltages V2 and V3 are out of phase; therefore figure traced on the screen will be an ellipse.

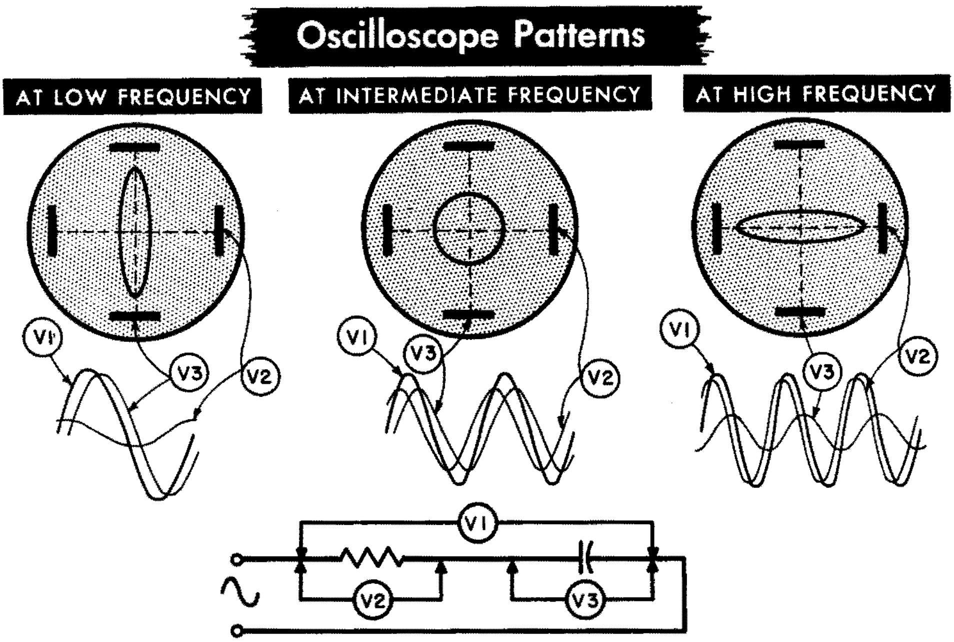
When a resistor and capacitor are connected in series, the same current must flow through both. The voltage across the resistor is always proportional to the current through it. The voltage across the capacitor, however, depends on the instantaneous charge on it, which is constantly changed by the fluctuating current. Of course the current does not flow *through* the capacitor, although that expression is often used. What really happens is that current reaching the capacitor changes the charge on it and, with it, the voltage across the capacitor. This voltage changes in proportion to the instantaneous current reaching it. A bigger current will change the voltage more rapidly.

When a sinusoidal voltage fluctuation is applied to the resistor-capacitor combination, there will be two sine waves, one across the resistor and one across the capacitor. The waves will have a time difference, such that the steepest slope on one coincides with the peak on the other. By applying the voltage from the resistor to one pair of oscilloscope plates, and that across the capacitor to the others, the spot will be made to travel up and down and from side to side, but because the two movements are not synchronized, the spot traces an ellipse. If both pairs of plates are connected across resistors, so the voltages always vary in proportion, the spot will travel along a sloping line, because the up-and-down and the sideways movements always occur at the same time and in proportion.

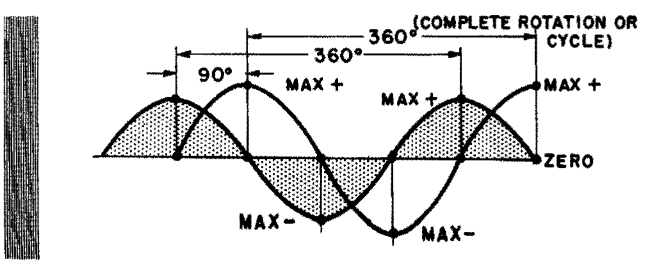


Showing the Effect of Frequency

Because the *fluctuation* of charge on a capacitor is something that takes time, the division of audio voltage drops between a capacitor and resistor in series will differ according to the applied frequency.

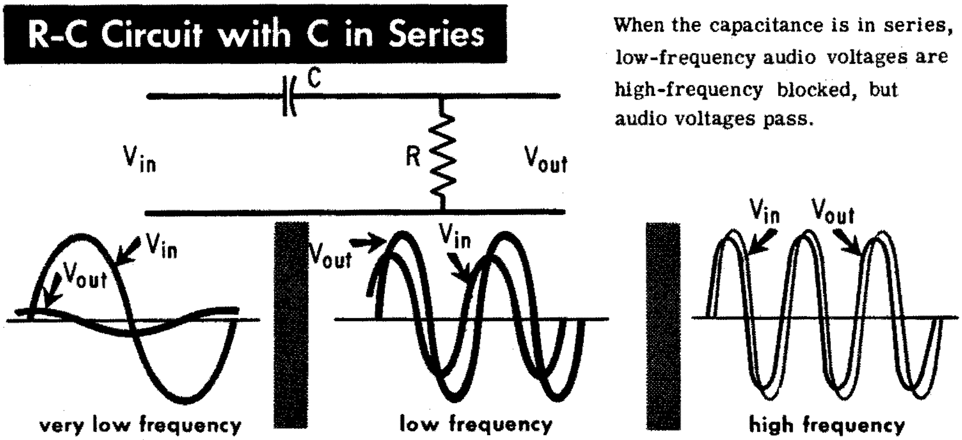


At a high frequency, a charge on the capacitor does not have time to change very much, and most of the voltage fluctuation appears across the resistor. At a low frequency, the charge on the capacitor has plenty of time to change, and can do so with a comparatively small current flowing through the resistor. Consequently, most of the voltage fluctuation appears across the capacitor. At an intermediate frequency the two voltages are about equal. Both of them are sine waves and they are "spaced" apart in time by a phase difference of 90°.



VOLTAGES V_2 and V_3 are always displaced in time by a "phase" of 90°

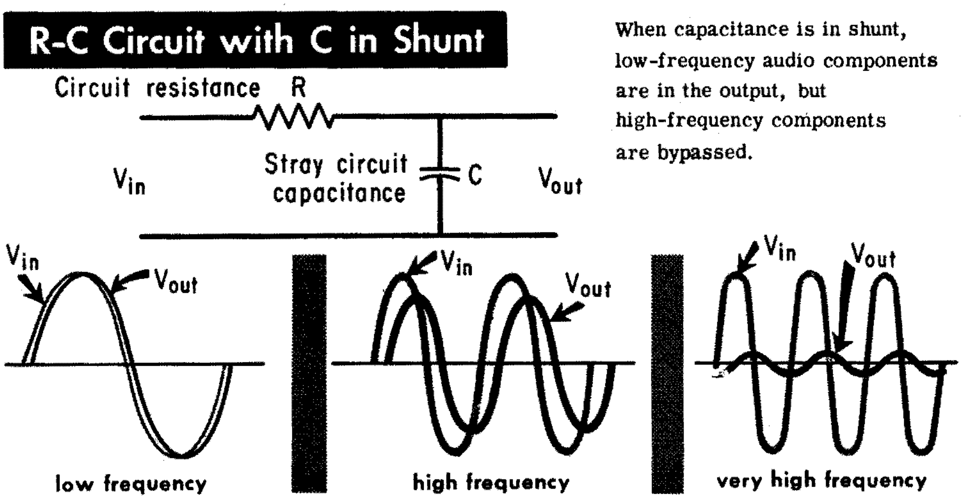
Arrangements of R and C



When the capacitance is in series, low-frequency audio voltages are high-frequency blocked, but audio voltages pass.

The effect of this frequency discrimination of resistance and capacitance in the same circuit depends on the way the components are used. If the capacitance is in series with the audio signal, it will tend to stop the low frequencies, blocking d-c altogether and losing most of the voltages at the extremely low frequencies. At high frequencies, all of the audio voltage will be transferred through the capacitor. The coupling capacitor between stages of amplification used is in this position of capacitance, which means that it produces a loss of low-frequency response.

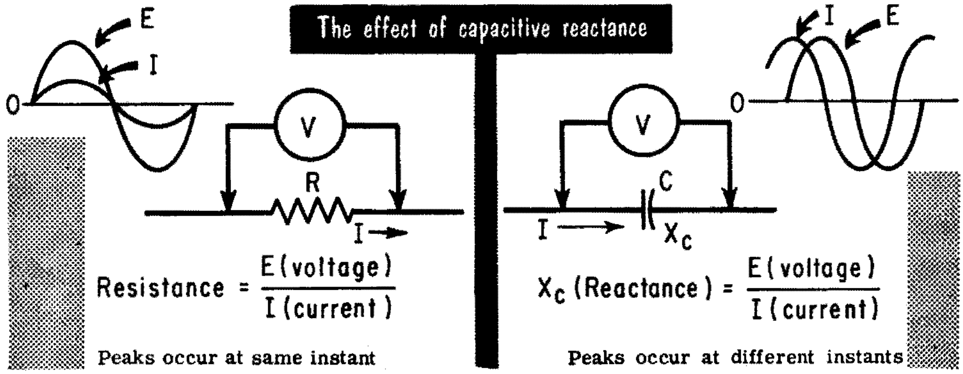
On the other hand, if the capacitance is in shunt, or so that the audio signal at the output of the arrangement appears across the capacitance, then its presence will mean that high frequencies almost disappear.



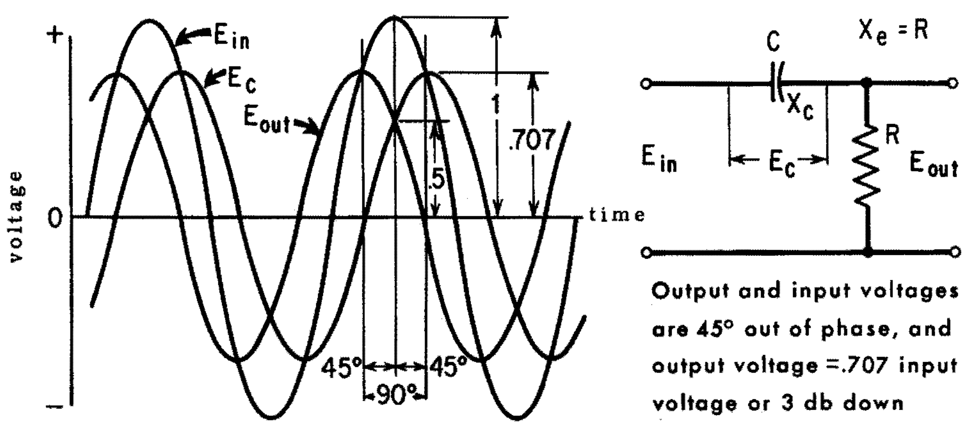
When capacitance is in shunt, low-frequency audio components are in the output, but high-frequency components are bypassed.

Capacitive Reactance

The ratio between the voltage across and the current through a capacitor at a specified frequency is called its *reactance*. This is valued in *ohms* like a resistance. But where voltage and current in a resistance are simultaneous, there is a time (or *phase)* difference (of 90°) in a reactance.



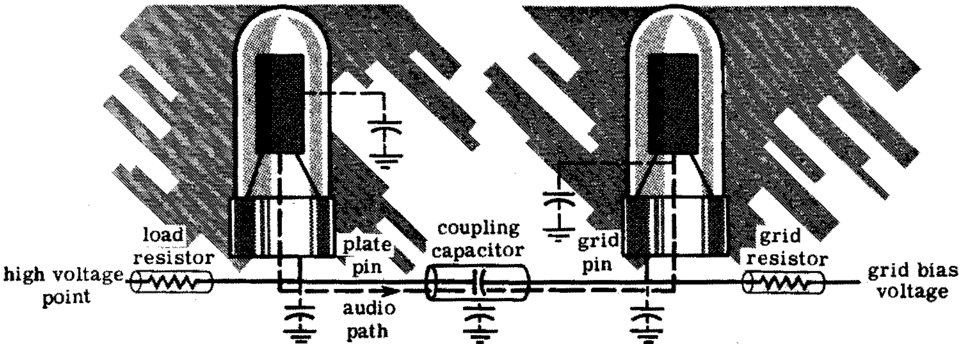
When the reactance of a capacitor is equal to the circuit resistance, the voltage across it is equal to that across the circuit resistance. At this frequency the phase between the input and output will be 45°, and the voltage output will be 0.707 of the voltage input or 3 db down.



Output and input voltages are 45° out of phase, and output voltage =.707 input voltage or 3 db down

Notice that making the two voltages (E_c and E_{out}) equal does not mean that either of them is equal to *half* the input voltage (E_{in}), because there is a 90° phase angle between the voltage across the capacitance and that across the resistance. For this reason, the two voltage drops cannot be added arithmetically.

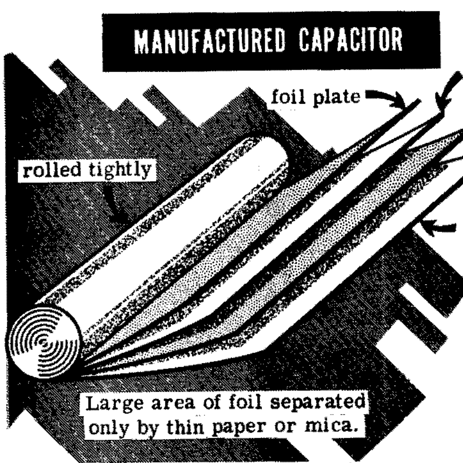
Distributed Capacitance



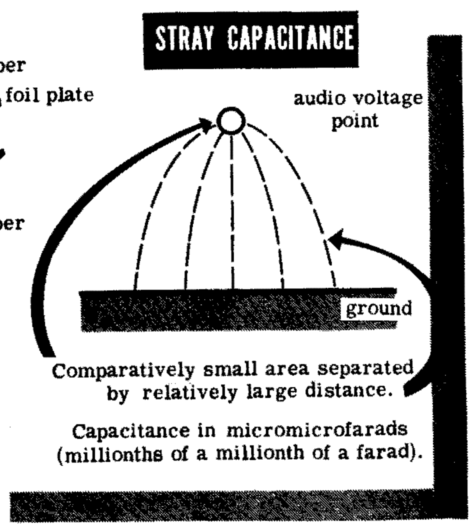
Stray capacitance that loses high frequencies is distributed everywhere in the audio circuit

Loss of high-frequency response occurs because of capacitance we cannot avoid, although it is not in the form of an actual capacitor. A capacitor is a component in which natural capacitance is exaggerated by putting large areas of foil in close proximity to one another. This produces capacitance values that are measured in microfarads. However, the presence of electrical circuits (conductors, components, chassis) in the same space produces a capacitance between different points, although it is measured only in micromicrofarads.

It is this capacitance that causes loss of high frequencies. At the frequency where the reactance of the total capacitance between the audio circuit and ground is equal to the effective circuit resistance, there will be a phase shift of 45° and 3-db loss in audio level.



Capacitance in microfarads (millionths of a farad).

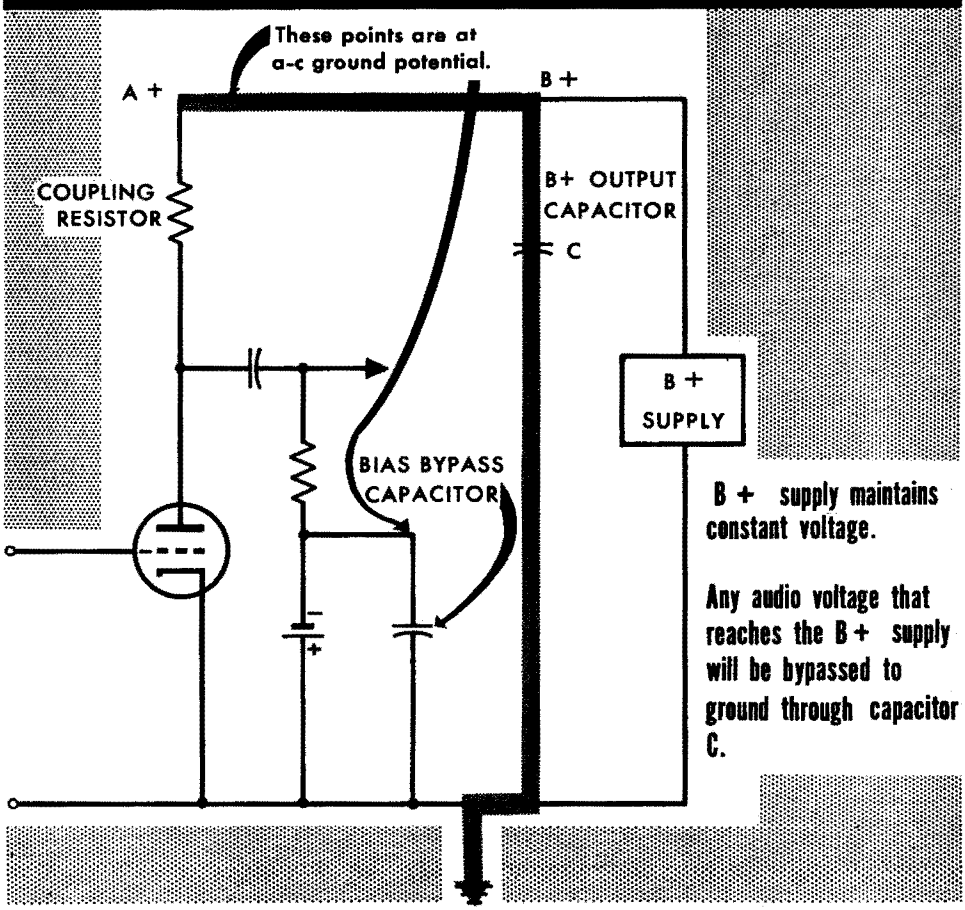


Capacitance in micromicrofarads (millionths of a millionth of a farad).

Improving Low-Frequency Response

Now that we know the action of the basic components, we can see what will happen by changing their values. First, consider the effect of changing the plate coupling resistor. In the case of a triode tube, the plate resistance of the tube is between plate and ground, while the plate coupling resistor is between plate and B+. The B+ supply will invariably have a very large capacitor coupling the high voltage to ground, so it is impossible for the B+ voltage to vary relative to ground. Any audio currents that reach the B+ point will be bypassed to ground through the large output capacitance in the B+ supply.

POINTS OF A-C GROUND OR ZERO POTENTIAL

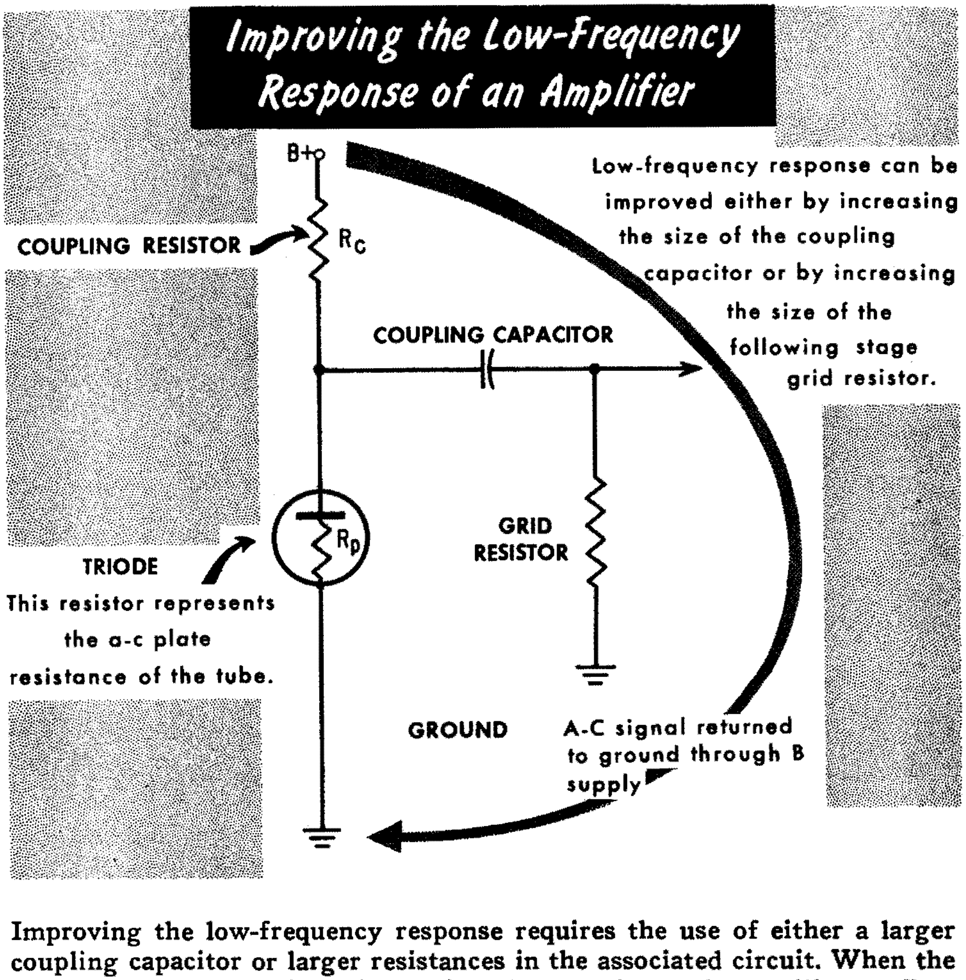


B + supply maintains constant voltage.

Any audio voltage that reaches the B + supply will be bypassed to ground through capacitor C.

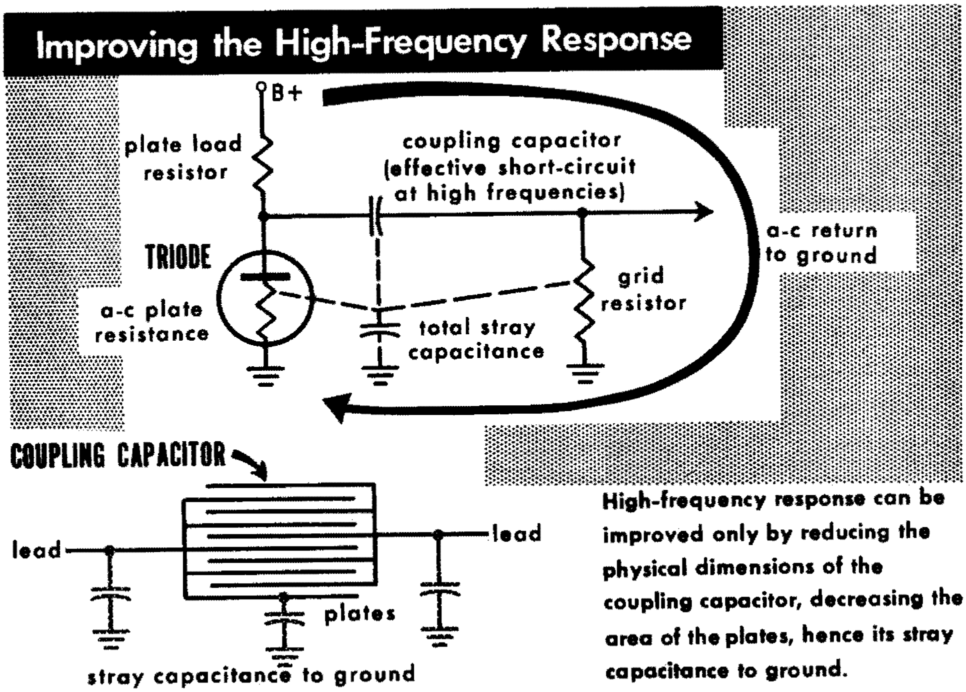
Improving Low-Frequency Response (contd.)

In the case of a triode tube, the coupling resistor is usually larger than the plate resistance of the tube because the a-c resistance of a triode is much lower than its d-c resistance. The plate coupling resistor will be about equal to the d-c resistance of the tube, so that about half the B+ voltage drops across the coupling resistor and half across the tube.



Improving the low-frequency response requires the use of either a larger coupling capacitor or larger resistances in the associated circuit. When the preceding tube is a triode, increasing the coupling resistor will not effectively increase the resistance for audio voltages from the plate to ground because this resistance is limited by the plate resistance of the tube. The only possibility of considerably increasing the resistance associated with the coupling capacitor is to increase the following stage grid resistor. Alternatively, the coupling capacitor can be increased in value.

Improving High-Frequency Response (contd.)



High-frequency response can be improved only by reducing the physical dimensions of the coupling capacitor, decreasing the area of the plates, hence its stray capacitance to ground.

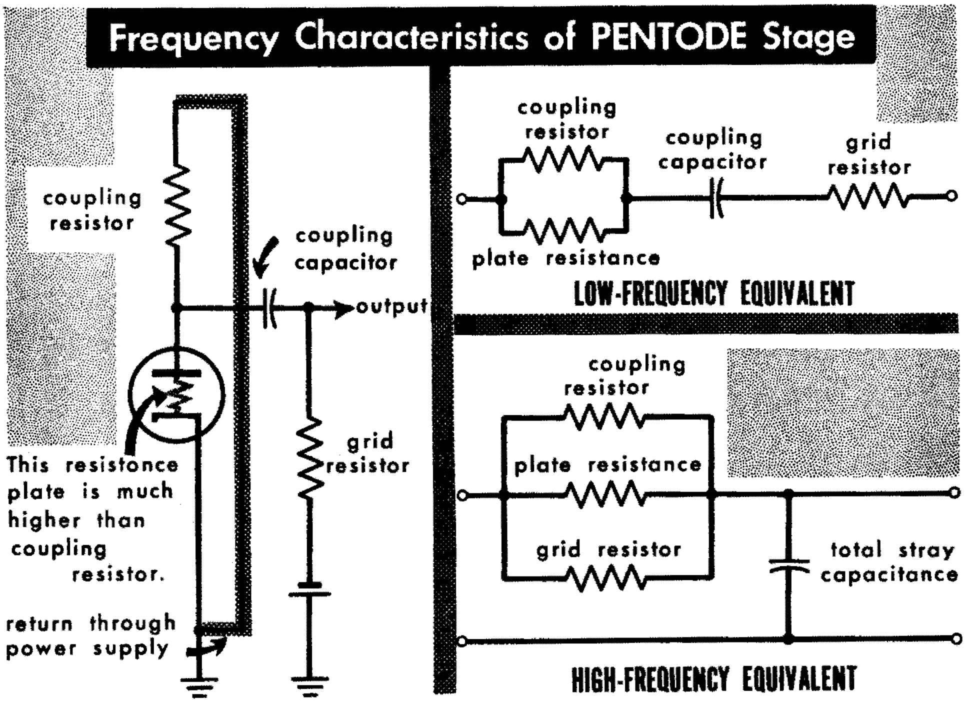
For high-frequency signals, the coupling capacitor is effectively a short circuit—no audio voltages appear across it, and both sides of the capacitor are at the same audio voltage at any instant. This means that we now have three effective resistance paths to ground from the plate of the tube: through the plate resistance of the tube, through the coupling resistor, and through the following stage's grid resistor.

Improving high-frequency response can be achieved by reducing the small capacitances to ground or reducing the total circuit resistance to ground. In the case of a triode amplifier, the lowest individual resistance (which principally controls the effective parallel resistance of the combination) is the plate resistance of the tube, which is fixed by the tube type. Consequently the best way to approach getting better high-frequency response is to see how we can reduce the capacitance.

One way to reduce capacitance is to use a smaller coupling capacitor (smaller in actual dimensions, rather than smaller in value). The coupling capacitor, like all other parts of the audio wiring, will add to the capacitance to ground, in addition to providing the requisite capacitance between its plates. The only way to reduce its capacitance to ground is to use a capacitor of smaller dimensions, so that it does not have such a large surface to provide capacitance to ground.

Improving the Frequency Response of a Pentode Stage

Using a pentode tube for the stage before the coupling alters the situation. Here the plate resistance is usually much larger than the load resistance, because the a-c resistance of a pentode is much greater than its d-c resistance. This means that the load resistance used with a pentode exercises the principal control on the plate-to-ground resistance. Using a larger coupling resistor, as well as increasing gain (which it naturally does with a pentode), will improve the low-frequency response, because it increases the resistance in series with the coupling capacitor. At the same time, the increased load resistance will reduce the high-frequency response because it makes the total resistance to ground higher.



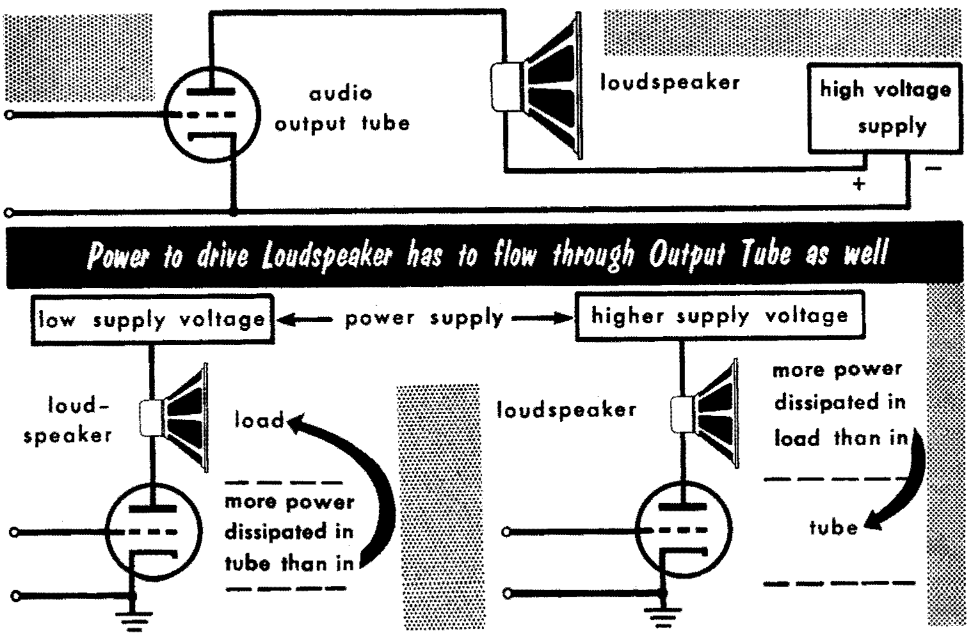
Thus whichever type of tube we use, there are optimum (or best) combinations of circuit values to get the kind of response we want: a response that is good enough, but not unnecessarily good. In each case, struggling too hard to get a good low-frequency response will give us problems at the high-frequency end, and vice versa. Further than this, if we struggle to get a good frequency response at both ends we shall get less gain from the amplifier. The values that we use will be such that the first tube cannot develop as much gain as it would if we were content with a frequency response that was not quite so uniform.

Power-Limiting Factors

The discussion thus far has been about getting audio *voltage* from tubes. To be able to hear sound from the loudspeaker, however, we need *power.* The loudspeaker needs *watts* to drive it, not just *volts.* For every watt audio of output that we get, there must be some power dissipated in the output tube, because the audio current in the plate circuit of the tube that goes to the voice-coil circuit also has to go through the tube.

This means that some power must be used up in the tube, which, in turn, makes the tube get hot. Thus there is a power limit on the tube before it will start to destroy itself. This limit is determined by the size of the internal components of the tube and the arrangements made by the tube designer to dissipate the heat. The manufacturer always specifies the maximum power dissipation of a tube intended for power-output use.

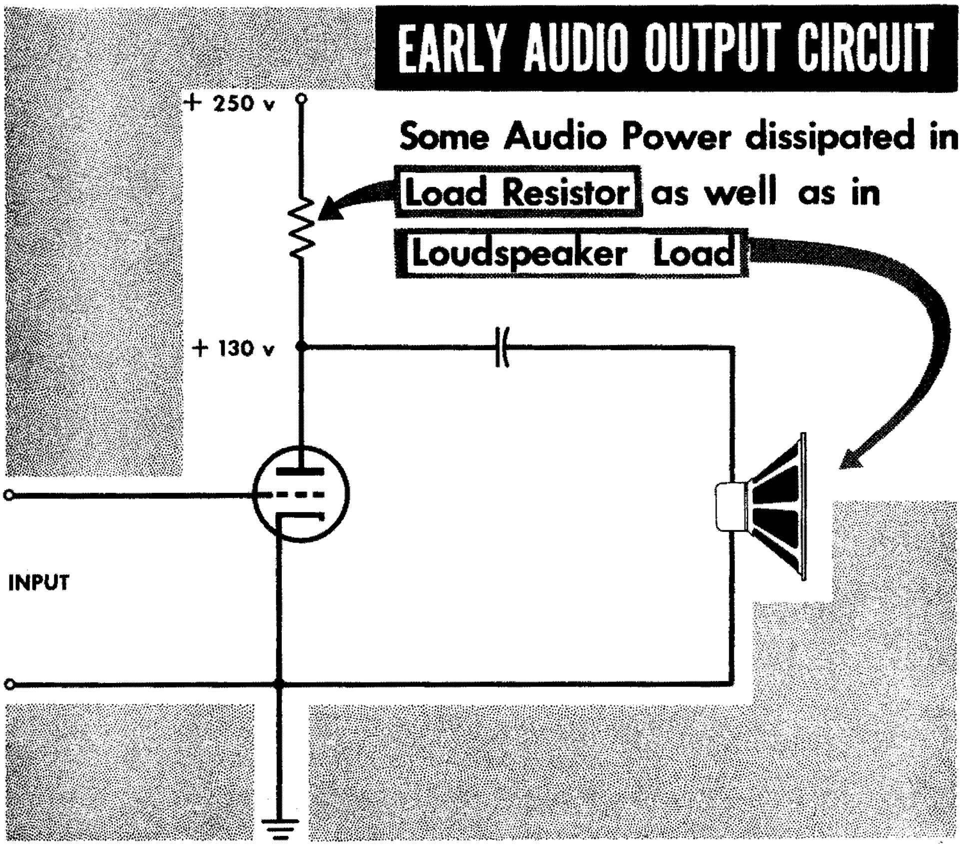
The tube is always in series with the load or impedance into which the power is fed because the tube is basically a *current* controlling device. This means the same current fluctuations flow in the tube as in the output circuit. By using a higher supply voltage, it is possible to drop a smaller *proportion* of the output power in the tube and get a larger proportion into the output circuit. It is thus advantageous to use a higher supply voltage because it gives better efficiency. This unfortunately brings us against another limitation. For any particular tube, there is a limit to the working voltage that may be used before some kind of breakdown might occur. This limit also is usually specified by the tube manufacturer.



Power to drive Loudspeaker has to flow through Output Tube as well

Improving Power Output

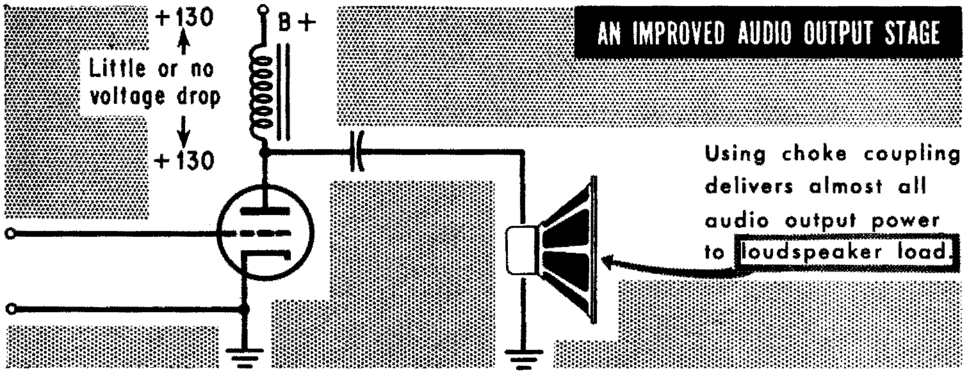
Let us start with the simplest tube circuit and see the various steps that are taken toward improving the available power output. Suppose we use a large version of the ordinary triode tube used for voltage amplification that will handle more power, using a plate coupling resistor and capacitor to get the power into the load.



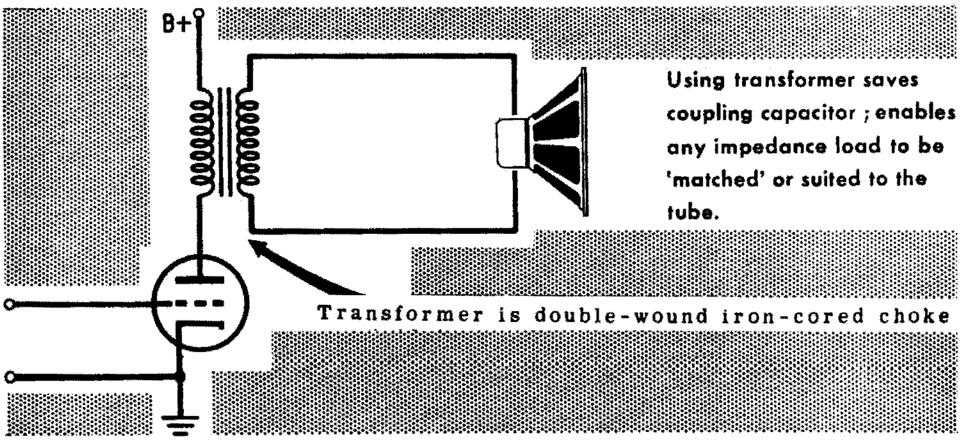
We shall also have to use a high-impedance load suitable to the tube. (This was the way some early radio sets were designed.) Suppose that B+ is 250 volts and that we drop 120 volts in the load resistor. This means the plate potential will be 130 volts. (It will fluctuate, due to the audio, above and below 130 volts, and this fluctuation will be passed through the coupling capacitor to the load.) The audio fluctuation also appears across the plate coupling resistor, hence it will get not only the power necessary to feed the tube with plate current, but also some of the audio power due to the current fluctuations flowing through it.

Transformer Coupling

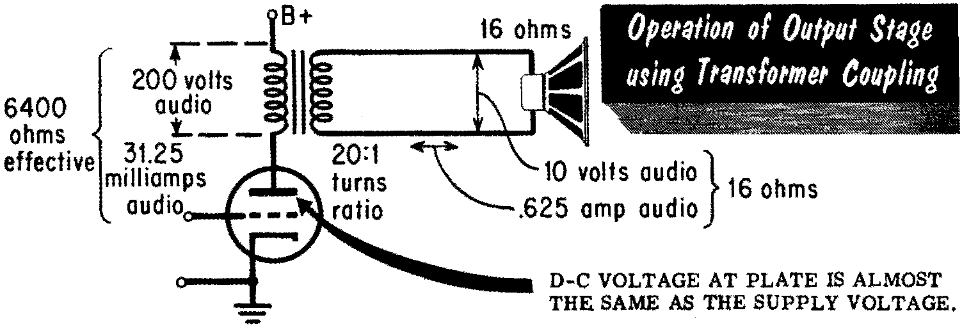
The first step toward improving the efficiency of a power-output stage is to eliminate this loss in the plate coupling resistor by using choke coupling. (This avoids the d-c drop in the plate load component.) We still get the audio fluctuations across the choke due to its *inductance*, but the fluctuations are too rapid to allow the current in the choke to fluctuate. The current fluctuations produced by the tube are all delivered through the coupling capacitor to the load. In this way we can use a B+ supply of 130 volts and still get the same total audio power from the tube as before, using a 250-volt supply.



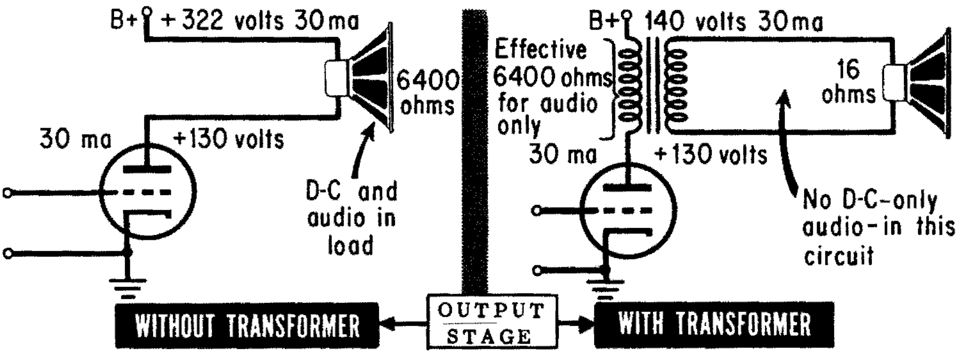
We can save a component here by eliminating the coupling capacitor. We do this by putting two windings on the choke, converting it into a transformer. The winding connected between the plate of the tube and B+ has a large number of turns, whereas the other winding has a much smaller number of turns. This means the audio current fluctuations are stepped up, while the voltage fluctuations are stepped down to suit a lower impedance load; the tube, however, operates as if it had the requisite high-value load resistance connected between the plate and B+.



Transformer Coupling (contd.)

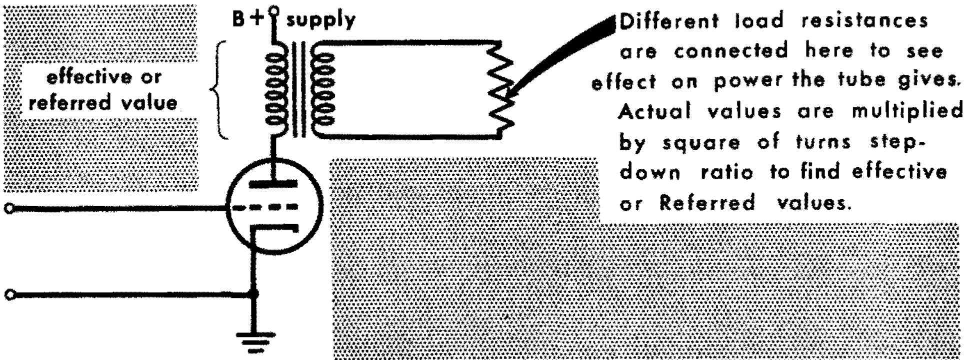


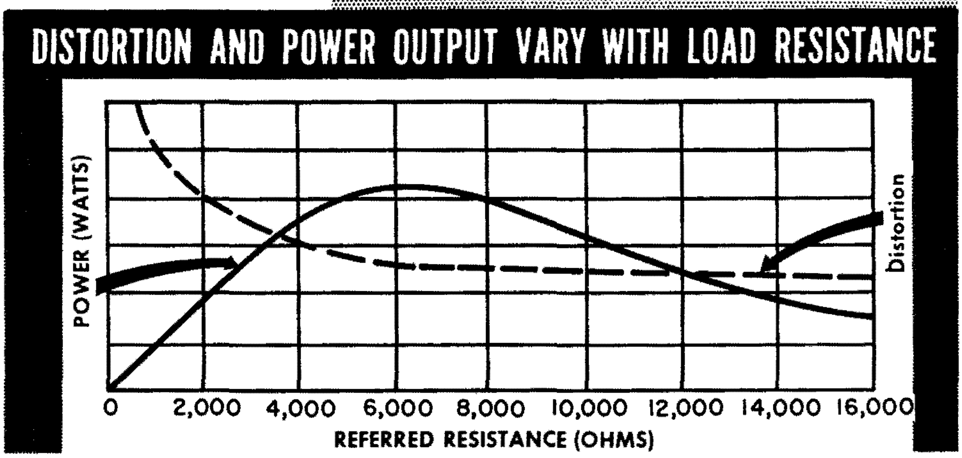
Suppose that the load resistance connected to the secondary is 16 ohms and the transformer step down ratio is 20:1. This means the voltage fluctuations will be stepped down by 20:1 and the corresponding current fluctuation stepped *up* by 20:1. The impedance will be transformed by a ratio of 400:1, producing an effective resistance at the primary of (400 $\times$ 16) or 6400 ohms. This is called the load resistance referred to the primary or referred resistance for short. The d-c voltage drop between B+ and the plate will be quite small—only about 10 volts.



Suppose that the operating point chosen for the tube is 30 milliamperes at 130 volts. A current of 30 milliamperes through 6400 ohms produces 192 volts. The transformer-coupled circuit thus works as if the B+ supply were 130 + 192 or 322 volts, with a 6400-ohm resistor connected between plate and B+ for a load. However, the 30 millamperes are not flowing through the load resistor actually in the circuit (16 ohms), but only through the primary resistance of the transformer, which may produce a drop of about 10 volts and thus require a supply of only 140 volts at 30 milliamperes to get 130 volts at the plate, instead of 322 volts at 30 milliamperes. Furthermore, all of the power developed by the tube is matched to the load, which can be any desired impedance, if the transformer ratio is correctly chosen.

The Effect of Changing the Load





With choke or transformer coupling, we can try the effect of applying different load values to the same tube. If we use referred resistance values slightly higher than the plate resistance of the tube, the waveform of the output follows the input waveform quite closely and there is little distortion. Continuing the increase in load resistance, however, the voltage fluctuation fairly quickly reaches a maximum and doubling the resistance value does not result in a proportionate increase in the voltage developed. For this reason, the power obtained drops off almost in inverse proportion to the increase in load value.

Making the referred load resistance the same as the plate resistance of the tube gives about the maximum power output obtainable, but causes considerable distortion in the output, due to the curvature of tube characteristics. If a high-quality power output is required, the load resistance must be at least two or three times as great as the plate resistance of the tube. Taking the load resistance value even lower than the plate resistance of the tube results in increased distortion and reduced power output.

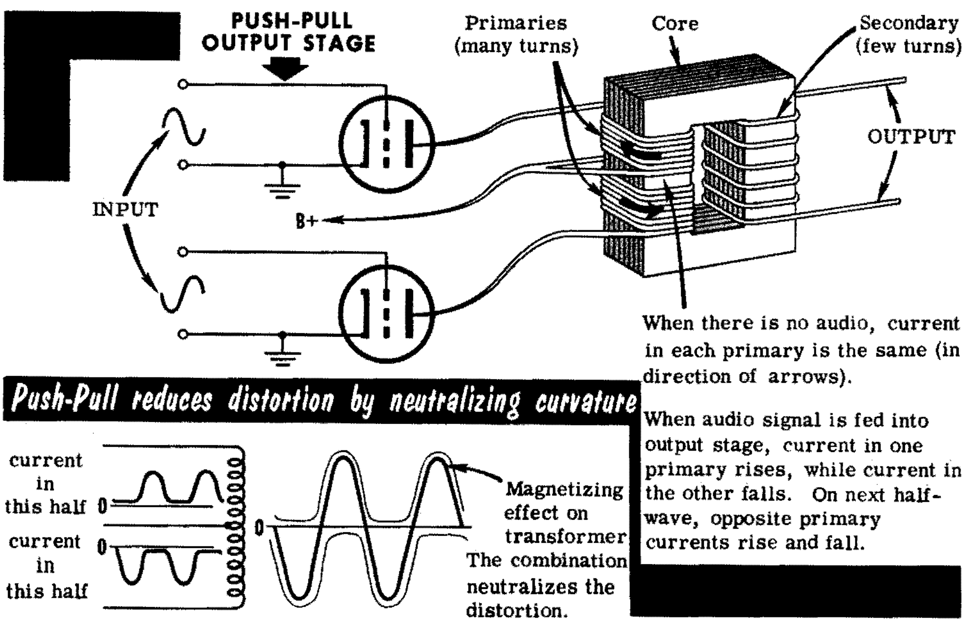
QUESTIONS AND PROBLEMS

1. What limits the dynamic range that an amplifier can handle?
2. What is the difference between the dynamic range of a single stage of amplification and that of a complete amplifier? How would you expect dynamic range to vary with (a) the total amplifier gain, (b) its frequency response?
3. Why does most noise in audio systems manifest itself as a hissing sound from the loudspeaker?
4. Why does a communications receiver use a very limited bandwidth, while a high-fidelity system should have a wide frequency response?
5. How can an input transformer help to increase dynamic range, and under what conditions will it do so?
6. What is the predominant cause of distortion in amplifiers (a) at low levels, (b) at higher levels?
7. What is a time base, and for what is it particularly useful in audio?
8. How does the time difference between input and output sine waves occur, due to the effects of coupling capacitors and how can this be shown on an oscilloscope?
9. What is capacitive reactance? How does reactance differ from resistance?
10. When the reactance in a circuit rises to a point equal to the resistance, the voltage does not drop to half its original value. Why? Explain the significance of the *3-db point.*
11. Explain how circuit resistance values in the coupling between amplifying stages can affect frequency response, and show what are the important quantities for low- and high-frequency response, both for triode and pentode stages.
12. What is the essential difference between the uses of tubes for voltage amplification and for power amplification?
13. Why is resistance-capacitance coupling, so often used for voltage amplifiers, not economic for power stages?
14. Explain the advantages of a transformer for power-stage couping.
15. How does a pentode improve the possibilities of performance (a) of a voltage amplifier, (b) of a power-output stage?
16. What is the difference between a pentode and a beam power tube? Can either of these tubes be connected in such a way as to work as a triode?

Push-Pull

The next step in improving the power output capacity of an amplifier stage is to use two tubes in a connection known as *push-pull.* This arrangement uses transformer coupling, but there are two primaries (the primary winding has two halves), through which the current flows in opposite directions. B+ is connected to the center point of the primary, with the plate of one of the tubes connected to each end. The current, therefore, flows from each plate outward through an equal number of turns to the center point. This means that the total magnetizing effect on the core of the transformer is neutralized as far as the d-c is concerned. (The transformer core only has to carry the magnetization due to the audio fluctuation.) This simplifies the design and cost of the transformer, but the big advantage is in tube operation.

With a single tube, matching the output load to the tube plate resistance results in a poor output waveform, which takes the form of a rounding at the bottom and sharpening at the top. When the tubes are worked in push-pull, the current flows in opposite directions around the transformer core and, consequently, what is the top of the current waveform in the upper part of the winding becomes the bottom of the current waveform in the lower half of the winding. Thus both halves of the current waveform have a sharpened portion added to a rounded portion, and the effect averages out, producing a much better waveform for the load value used. To achieve this, we must provide the correct audio voltages at the grids of the tubes. We shall consider this problem presently.



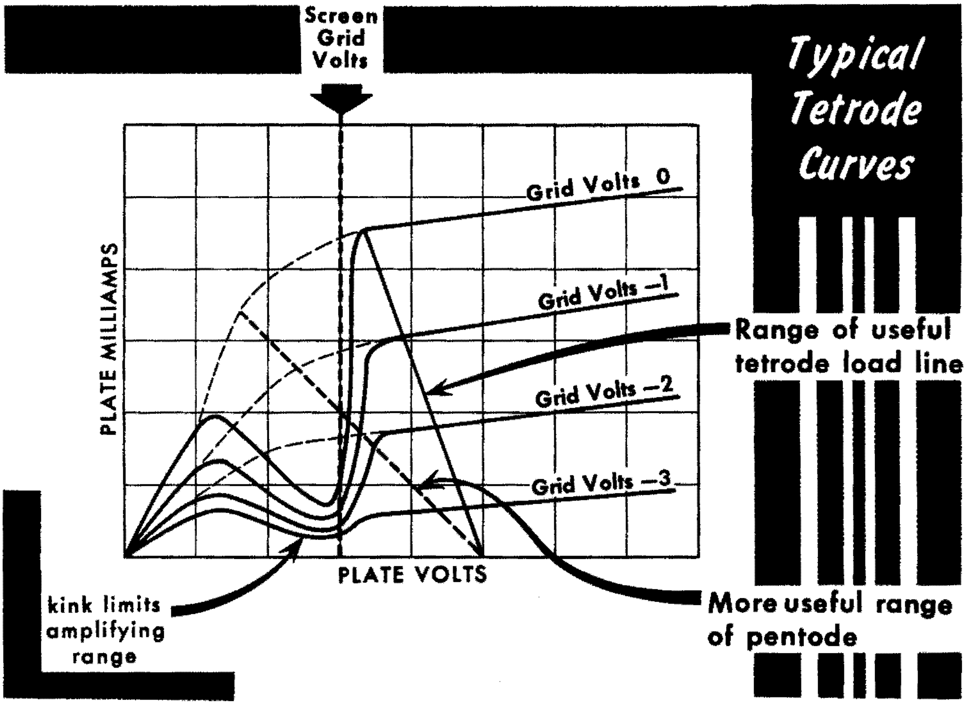
The combination neutralizes the distortion.

When there is no audio, current in each primary is the same (in direction of arrows).

When audio signal is fed into output stage, current in one primary rises, while current in the other falls. On next half-wave, opposite primary currents rise and fall.

The Development of The Pentode

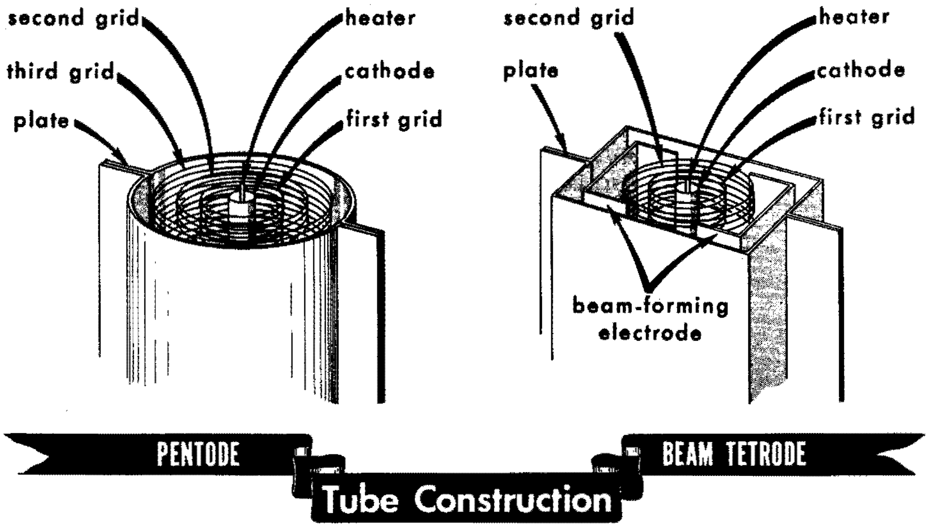
The next circuit improvement comes from the design of the tube itself. We already mentioned the use of a pentode for voltage amplification. Tube development, however, was originally carried out to improve power output. The effect of plate voltage swing in reducing plate current swing in a triode tube, cuts down its power-handling capacity by an even greater amount than it does its voltage-handling capacity.



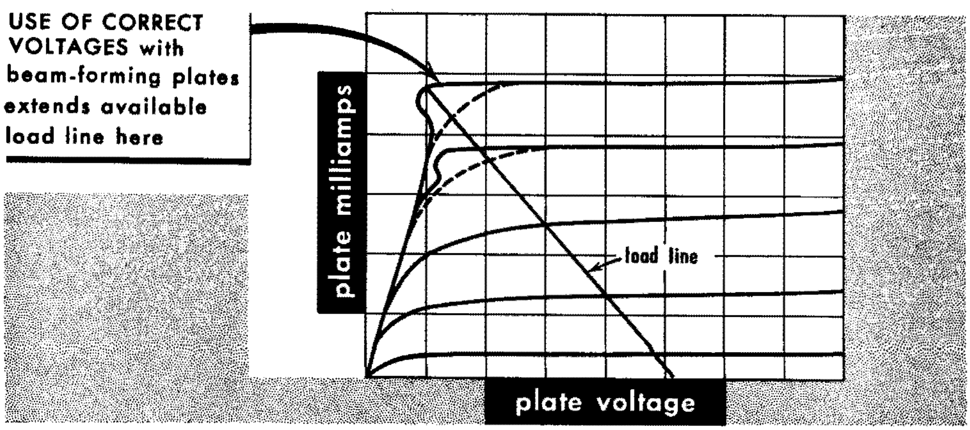
The first development from the triode was the tetrode, made by inserting only the second (screen) grid. This was found to result in an improvement in power output, although it was limited by secondary emission. When the plate potential is lower than that on the screen grid, electrons bounced off by the impact of arriving electrons at the plate will be attracted to the screen, resulting in plate current curves that give a curious kink and limit the range over which the tube can be successfully used for normal amplifying purposes.

The pentode overcomes this difficulty by using the third (suppressor) grid which prevents any electrons bounced off the plate from traveling back to the screen or second grid under any circumstances. All such electrons are repelled by the suppressor grid and sent back to the plate.

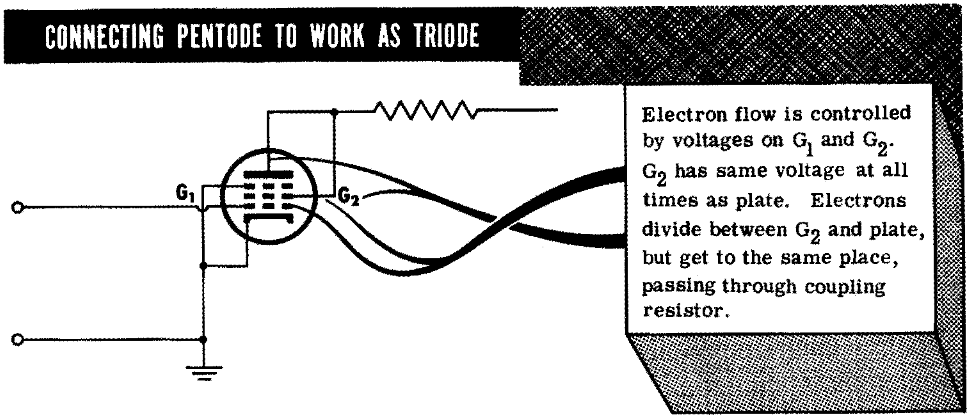
The Beam-Power Tetrode



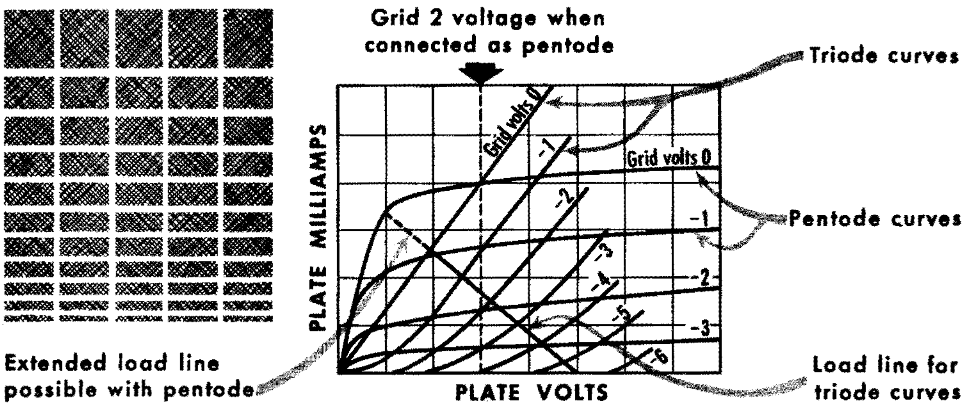
Another method of preventing secondary emission uses a specially shaped electrode between the screen grid and the plate. It is called a *beam-forming* electrode and consists of two pieces of metal placed in a different direction from the grid wires of the normal grid. (The normal grid wires are wound around the cathode at a constant radius; beam-forming electrodes run parallel to the cathode in specially selected positions.) It is connected to the cathode in just the same way as the suppressor grid of the pentode, and its effect on performance is very similar to that given by the addition of a suppressor grid. It succeeds in squeezing the last ounce of efficiency out of the tube, by extending the voltage and current swing to the maximum possible degree.



Triode Connection of Pentodes

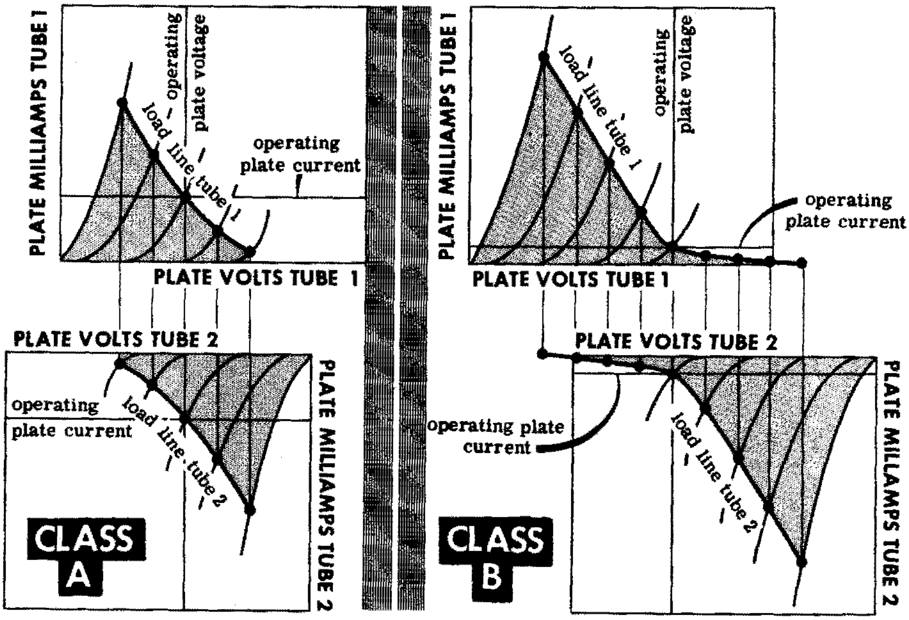


Any pentode-type tube can be made to work as a triode-type tube. This is done by connecting the second grid directly to the plate, so that both swing together at the same voltage. Because the screen grid, in combination with the control grid, is principally responsible for controlling the plate current, the presence of the suppressor grid between the screen grid and plate does not materially alter the tube's performance from that of a triode.



Using the best possible load resistance with triode-connected tubes, the voltage fluctuation between B+ and plate reaches little more than half the B+ supply voltage. Changing the method of connection to pentode alters the curve so that the zero grid-voltage curve is pushed out into a "knee." This extends, very considerably, both the voltage and current fluctuation available in the plate circuit. This, in turn, triples or quadruples the power that any pair of tubes will give.

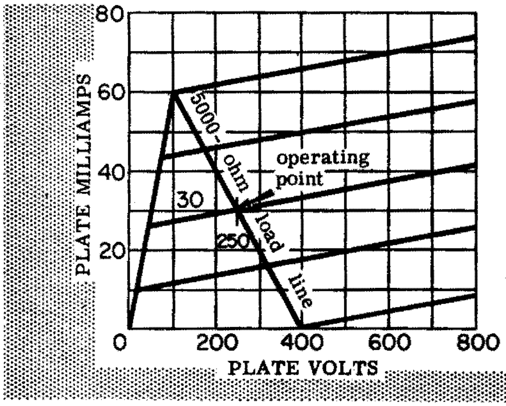
The Lowering of Distortion



Using two tubes in push-pull helps the waveform problems, so that the distortion produced by one tube cancels that produced by the other. This can be understood better if we think of each tube as having a *curved* load line. The input voltages to the grids are equal but 180° out of phase. The plate voltages likewise are out of phase because of the coupling between the two primary windings of the output transformer. So the changes in plate current must adjust between the tubes to allow this condition, while the two of them supply the *total* current fluctuation to the load at all points. The ratio between total voltage and current fluctuation of both tubes is set by the load resistance matched to the secondary of the transformer, but each tube feeds a load resistance whose value is constantly changing, as represented by the curves.

This effect can be extended further to increase the efficiency of the output stage. Normally the steady plate current is about half the maximum plate current (which occurs when the grid voltage fluctuation goes from the operating point up to zero). The current fluctuation in the tube at maximum power level swings between almost zero current and twice the steady current. This sets a considerable limitation on the power-handling capacity of the tube because the steady component is such a large proportion of the maximum current the tubes take. Using a greater negative bias on the grids of the tubes makes the audio fluctuations carry the current from almost zero up to a maximum in one direction and cuts the tube off so that no current flows in the other direction. This makes possible a considerable increase in efficiency and available power output.

Power Output from a Class-A Stage

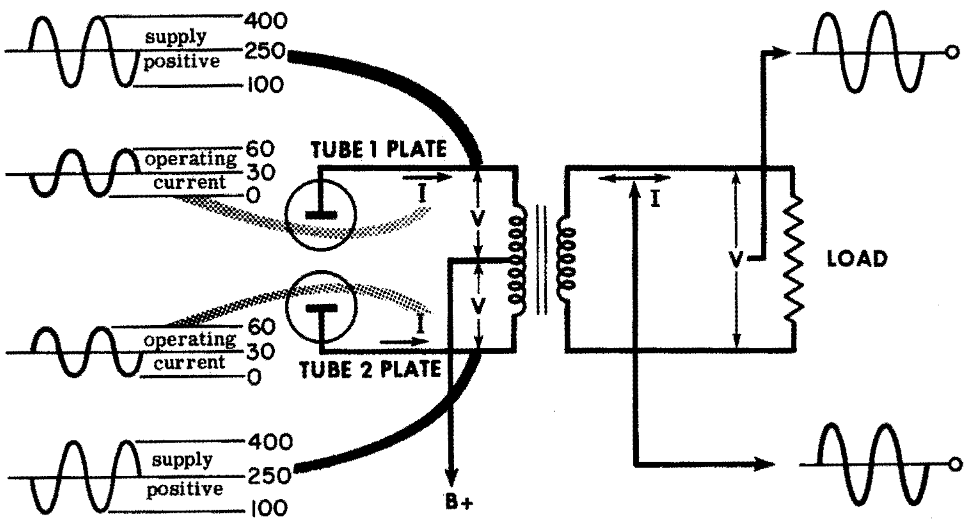


Calculations for Class A push-pull

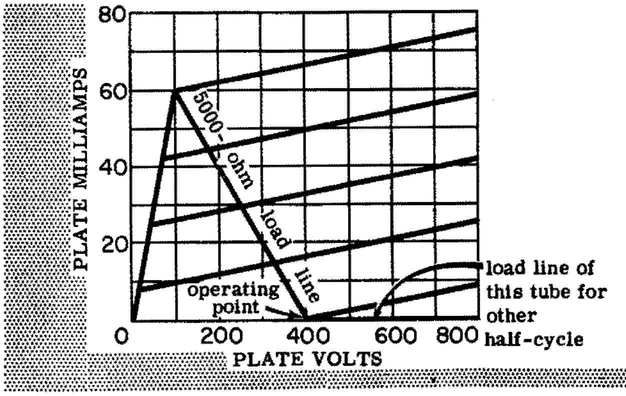
Load line is for each tube , but both work at once.

The characteristics for the other tube (not shown) would be inverted underneath.

As an example, suppose that in ordinary push-pull (known as *class A*, which means that both tubes are conducting current all the time), the operating point for each tube is 250 volts at 30 milliamperes and that the load value presented to each tube is 5000 ohms with pentode operation. Disregarding the curves to make the calculation simpler (if approximate), the audio fluctuation should carry the plate between 100 volts at 60 milliamperes and 400 volts at zero milliamperes. This represents a peak fluctuation from each tube of 150 volts and 30 milliamps in each direction, which is a peak power of (150 × .03) or 4.5 watts per tube or 9 watts for the two tubes in push-pull. The average power, using a sine wave to drive the output, will be just half of this figure or 4.5 watts for the two tubes.



Class-B Operation

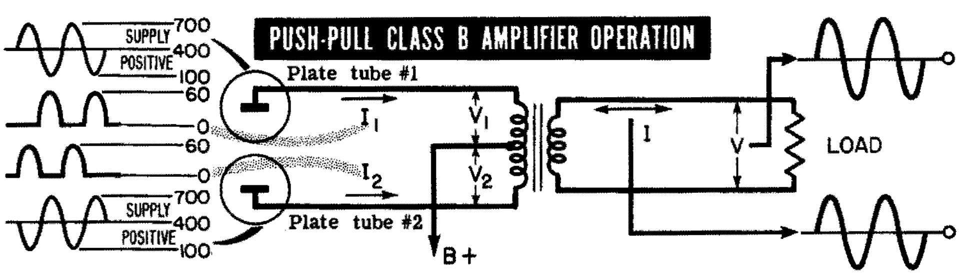


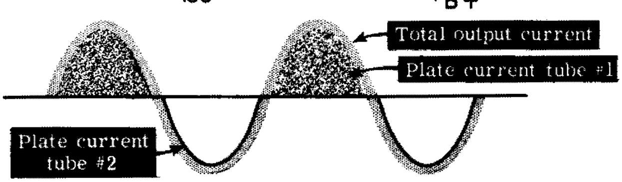
Calculations for Class B push-pull

Load line is for each tube, but only one works at a time on alternate halves of waves.

If we wish to use an extreme economy measure, known as *class-B* operation, we bias each tube back approximately to zero current. (Actually it does not go quite to zero current because of the curvature, but it goes to where zero current would be if the tube characteristics were all straight.) Using the same load line, the operating point for each tube would (in theory) be 400 volts at zero milliamperes. This means the plate potential voltage will swing from 100 volts to 700 volts. Plate current will swing, during one half-cycle, from zero to 60 milliamperes and back, while in the other half-cycle, no current flows in that tube.

The permissible maximum voltage on the plate is considerably increased by this method of operation (from 400 volts to 700 volts). There is less danger of breakdown between the plate and some other electrode when no plate current is flowing. There is, however, a maximum rated voltage even under this condition, which sometimes restricts the amount by which this method of operation can improve efficiency.



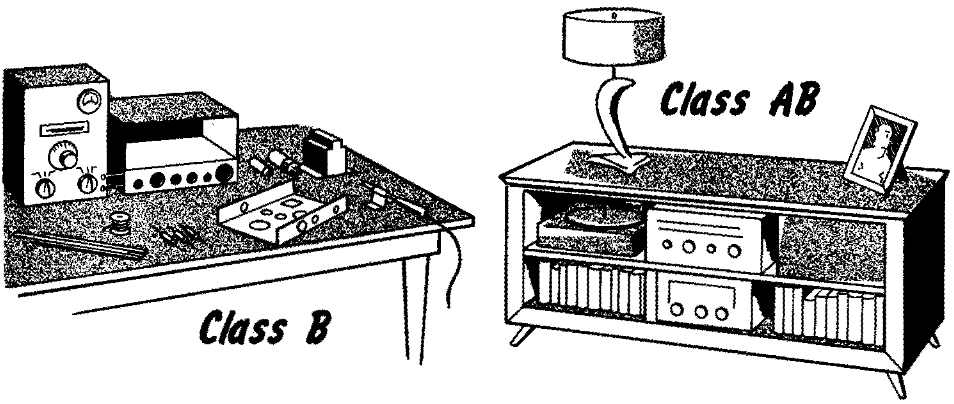


Current Waveforms for push-pull class B amplifier operation

Class-B Operation (contd.)

In this case, the peak voltage and current swing are now 300 volts and 60 milliamperes in each direction, with one tube providing one half-wave and the other tube providing the other half-wave. This represents a peak power of (300 $\times$ .06) or 18 watts for the complete arrangement, or an average power of 9 watts—just double that of class-A push-pull operation.

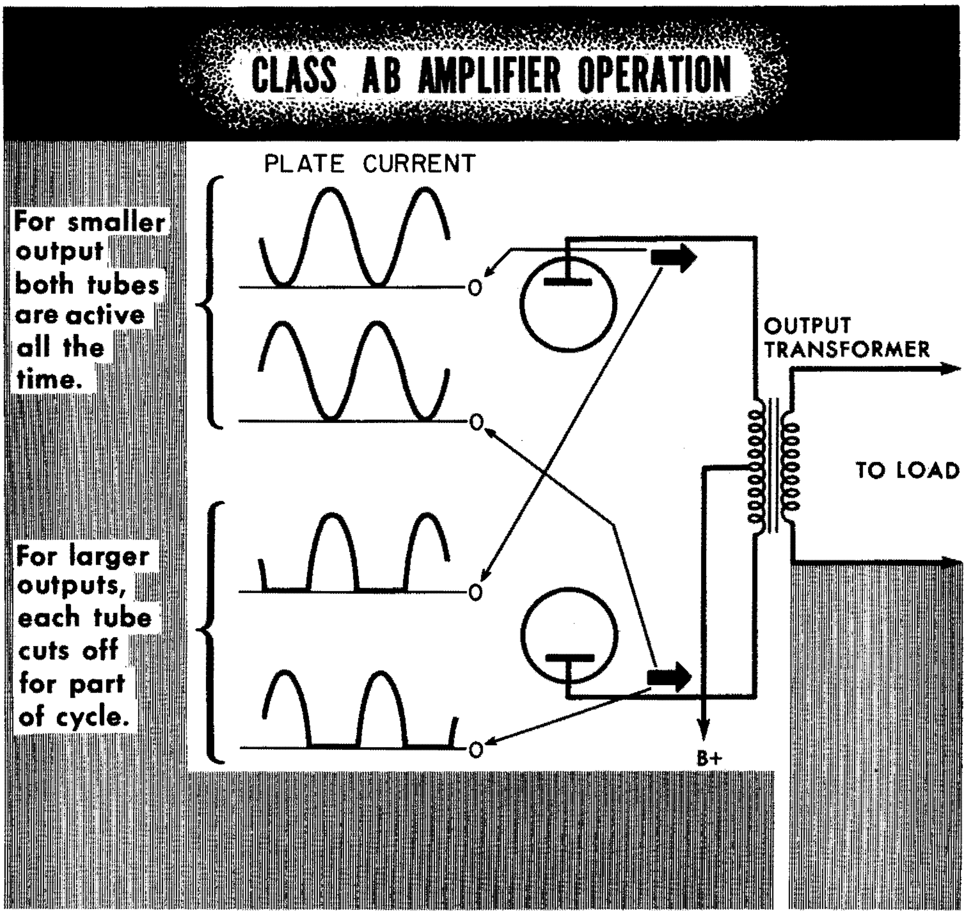
CLASS B OPERATION is suitable only where continuous skilled supervision is available



This method is not too often used because it involves very careful adjustment of the operating point. Tubes vary in individual characteristics due to production differences, with the result that the zero plate current point (the point where the grid bias cuts off plate current) is apt to differ noticeably between individual tubes. This variation requires a critical adjustment in the amplifier to make sure that both tubes are at the correct point for producing undistorted output. Not only is this adjustment critical when replacing tubes, it can also involve continuous attention if the plate voltage changes at all due to line voltage supply fluctuation or due to the fact that tubes sometimes change their characteristics as they get older.

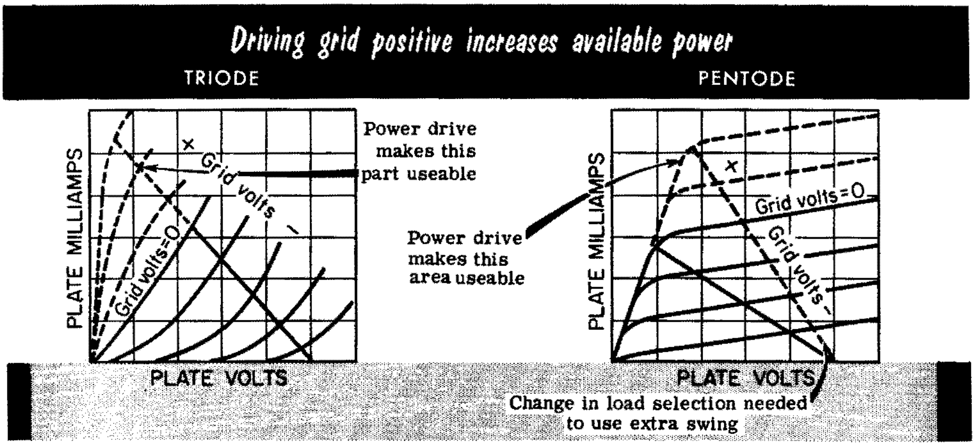
Class-AB Operation

Class-B operation is only really suitable for amplifiers where constant skilled attention is available to insure the conditions are correct. Most practical amplifiers are used under circumstances where they are installed and forgotten. A class-B amplifier would likely give considerable distortion if this were done.



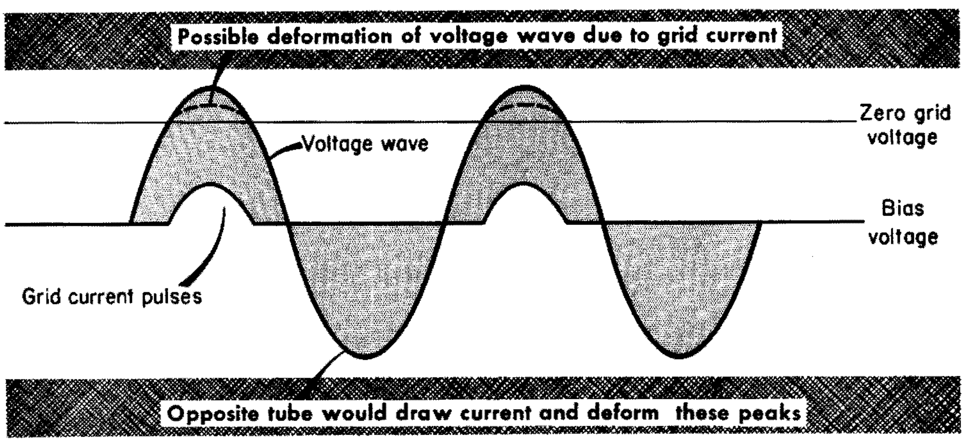
For this reason a popular method (called class-AB) consists of splitting the difference. For small outputs, up to about half maximum power, both tubes are conducting current all of the time. When full output is needed, each tube goes beyond the zero plate current for part of the cycle, but for not a complete half-waveform. This design results in an amplifier that is less critical in operation and can be used without critical selection of output tubes and adjustment of bias.

Power Drive



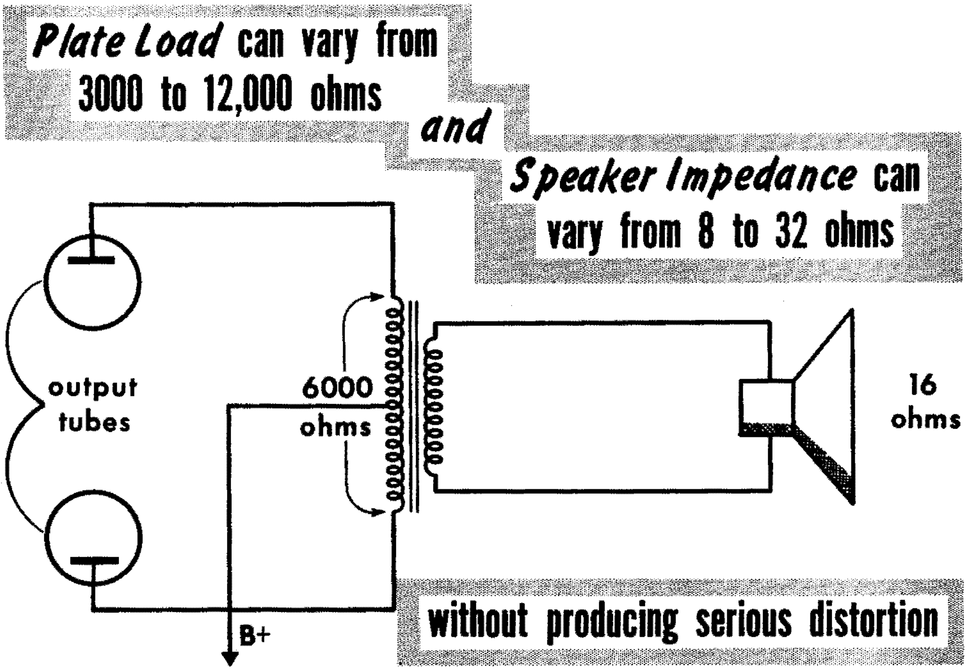
One more thing can be done to increase the power available from a given pair of tubes. If the control grid is made positive, the plate current will rise still farther, whether a tube is operated as a triode or pentode. In the case of pentode operation, the load value must be changed, otherwise additional grid swing into the positive cannot extend the *length* of the load line. This would still result in clipping that no amount of grid current could remedy.

Applying positive drive on the grid means that the grid itself will draw current when it becomes positive with respect to the cathode. (Some of the electrons will be attracted to the grid itself, instead of passing through it to the other electrodes.) This means that the grid is no longer just a voltage-operated electrode in the tube. We now have to make a more complicated design of the *drive stage,* as it is called, to provide the necessary grid current (when it is drawn) accompanying the correct voltage waveform, to give the right output.



Amplifier Efficiency and Distortion

We have a variety of circuits that can be used to get more power from tubes. Increasing the efficiency is in itself an advantage for many purposes, but it brings with it some disadvantages. Probably the best kind of output for general purpose use, if the inefficiency can be tolerated, is the push-pull triode-operated arrangement. It gives good clean amplification, whether the correct load value (according to design) is used, or a value that may differ from it by a ratio of 2:1 or more.



If, for example, the plate load of the two tubes in push-pull should be 6000 ohms, they will not show serious distortion if an actual plate load is used ranging from 3000 to 12,000 ohms, or even more. In terms of the secondary side of an output transformer designed to feed a 16-ohm loudspeaker, it will not produce serious distortion even if it is connected to an 8-ohm or 32-ohm loudspeaker.

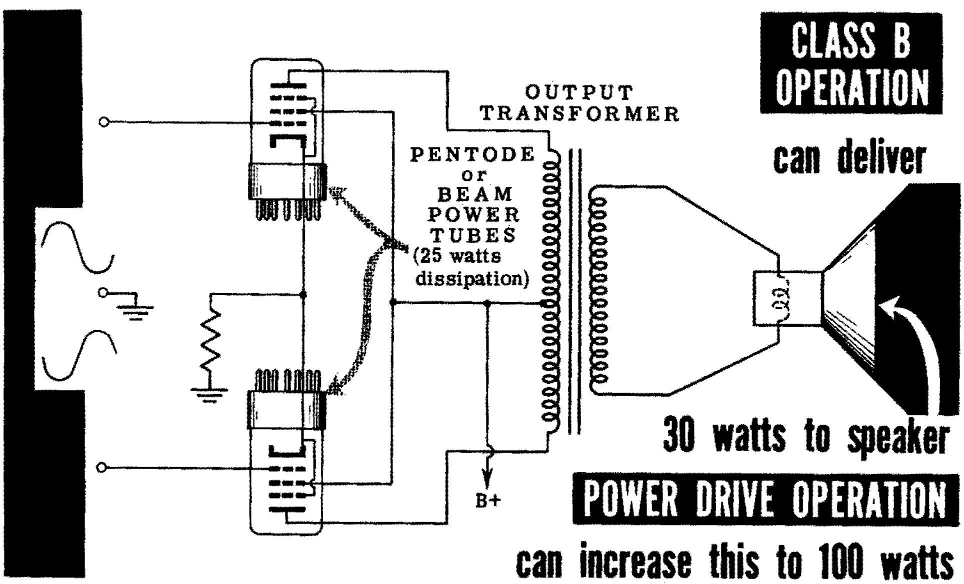
Of course, using an incorrect load in this way will reduce the available output proportionally, but it does not in itself introduce distortion. The maximum rated output will only be given into the correct load but use of a different load does not seriously decrease the maximum output and, provided the amplifier is not run into overload conditions, it will not cause distortion.

Amplifier Efficiency and Distortion (contd.)

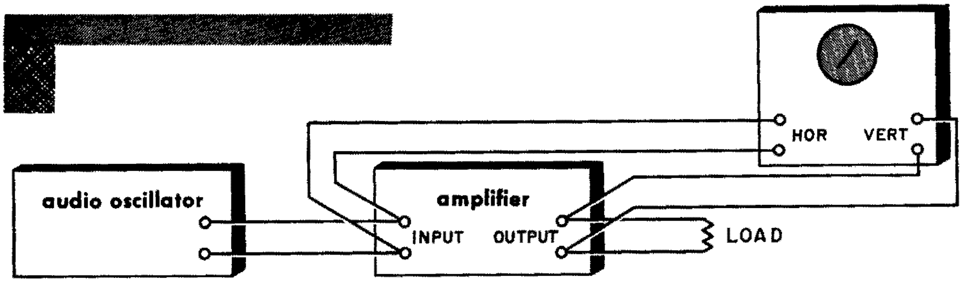
This kind of output requires two 25-watt type tubes, which are relatively large, and must dissipate 50 watts of heat in themselves to give a maximum audio output in the region of 6 or 7 watts. If the operating condition changes to class B, the output will be raised to 10 or 12 watts. With this change, the operation condition is more critical. The amplifier needs careful adjustment, but it does not usually introduce serious distortion by using the wrong load value, provided this is *higher* than the rated value. For example, if the rated secondary impedance of the output transformer is 16 ohms, this circuit will operate successfully with a loudspeaker rated at 32 ohms, but not with one rated at 8 ohms.

The use of pentode- or beam-tetrode-type tubes will considerably boost the available output. A pair of tubes dissipating 25 watts of heat each (a total of 50 watts) can give an audio output of as much as 30 watts. Using this kind of circuit and working into power drive (which means the grids of the output tubes are driven into the positive region at parts of the audio waveform), it is possible to push the power output of two 25-watt triodes working in class-B push-pull up to 50 watts.

Using pentodes this way can give as much as 100 watts from tubes that only produce 50 watts heat between them. The maximum audio output is thus twice as much as the heat generated in the tubes themselves. This is very useful for high-powered systems, but involves critical adjustment, more expensive supply circuits, and careful attention to see that the load value connected to the amplifier is correct. Otherwise, the power output quickly diminishes.



Amplifier Efficiency and Distortion (contd.)

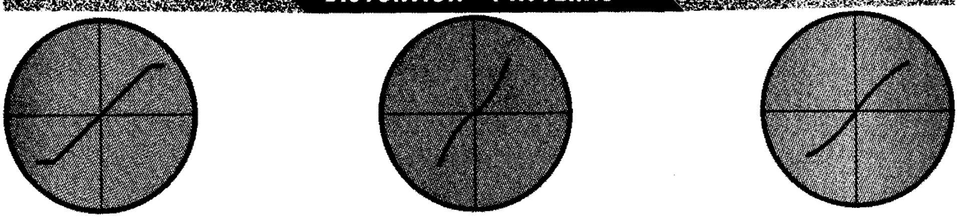


SCOPE TRACE IS A STRAIGHT SLOPING LINE IF OUTPUT IS ALWAYS A PROPORTIONATELY AMPLIFIED VERSION OF THE INPUT

What kind of thing happens when amplifiers are not correctly operated? We can see this using the oscilloscope to look at the waveforms. Possibly the most interesting way to do so from the viewpoint of what happens in the amplifier is to connect the input sine wave to the horizontal deflection plates and the output waveform to the vertical deflection plates, instead of using the sawtooth time base. This gives us what should be a straight sloping line on the screen.

If the line curves or squares off in any way, this shows distortion, because it means the output waveform is not following the input waveform in correct proportion. If an attempt is made to get too much power from an amplifier, this means that the grids of the output tubes will be driven positive (or in the case of output circuits designed to take positive grid drive, they will be driven *more* positive than they should be). The waveform flattens because an increase in fluctuation at the input is not accompanied by a corresponding increase in fluctuation at the output. If a load of an incorrect value is used with tubes that require careful selection of the load, the waveform will change its shape, getting either sharper or flatter towards the tips, or waving about in different ways.

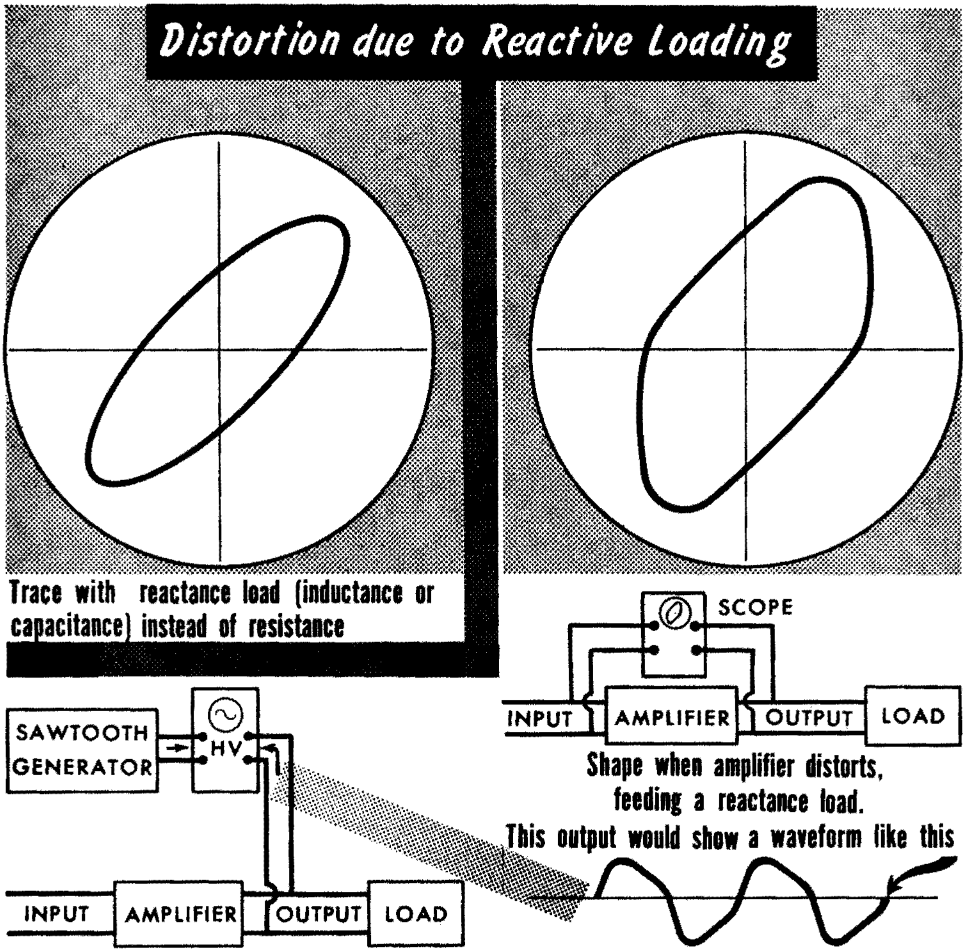
DISTORTION PATTERNS



This trace indicates clipping.

Two curved traces indicating wrong amplifier loading. Both are distorted sine waves.

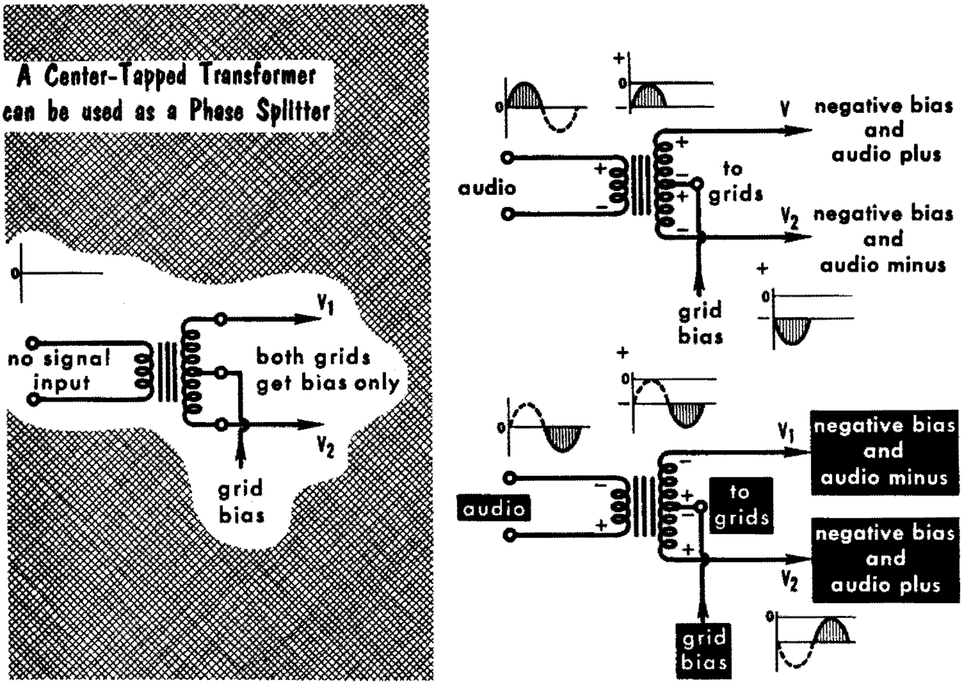
Amplifier Efficiency and Distortion (contd.)



With amplifiers in which the loading is important, trouble is sometimes caused because the load is not a pure resistance but possesses reactance in the form of inductance or capacitance along with the resistance. In a loudspeaker some of this reactance may be due to the inductance of the voice coil, whereas the resistive component is the d-c resistance of the coil. The actual working of the loudspeaker can cause additional components of resistance and reactance as well. Even if the total impedance of such a combination is the correct number of ohms, it may still cause distortion because the amplifier does not work well into a reactive load. This kind of distortion usually has the effect of "tilting" the waveform when viewed on a sawtooth timebase, or when viewed against the input, it opens out the slanting line trace into a distorted ellipse.

The Transformer-Type Phase Splitter

We have talked about using tubes in push-pull and said that to do this we need to have an audio input that makes one grid fluctuate positively while the other one fluctuates negatively. The question now comes, "How can we get this kind of drive or audio voltage for push-pull output tubes?" There are a number of these so-called ***phase-splitting circuits.***

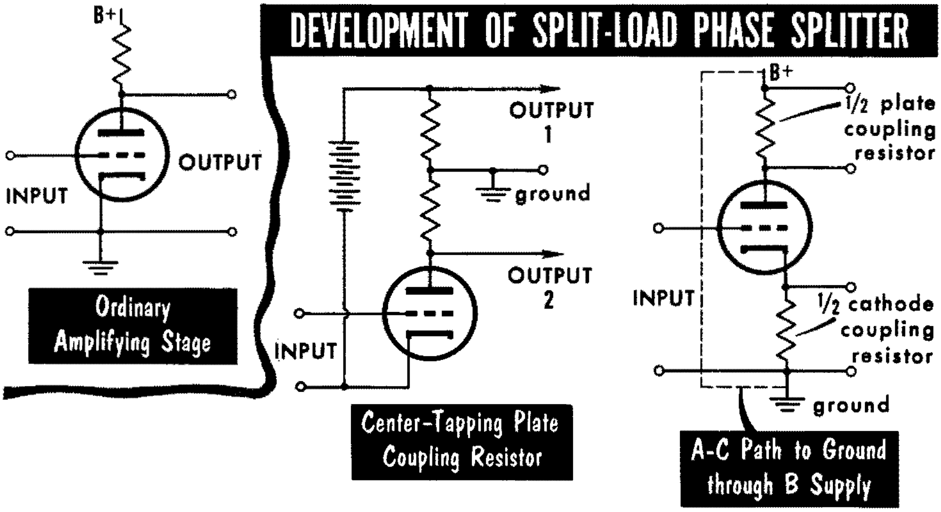


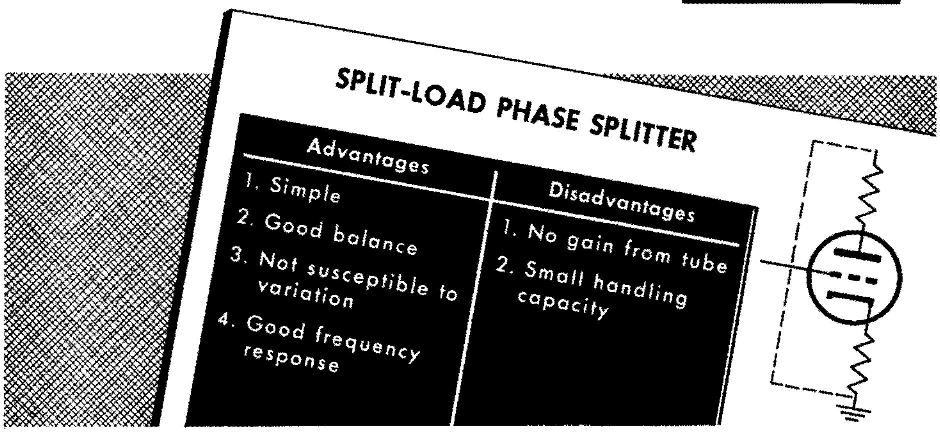
A Center-Tapped Transformer can be used as a Phase Splitter

The simplest arrangement uses a variety of audio transformer similar to the output transformer. The secondary, however, has more turns than the primary and is center-tapped, like the primary winding of the push-pull output transformer. The center tap is connected to the negative grid bias voltage. When no audio is passing, both ends of the secondary winding have the negative grid bias voltage for their respective grids. When an audio voltage is presented to the primary of the transformer, one end of the winding goes positive from the bias voltage while the other one goes negative, and vice versa.

This system has some very useful features. It is simple, and it gives good balance in voltages supplied to each grid, if it is well designed. It has, however, a disadvantage when used in a feedback amplifier, as will be explained later, and for this reason various phase-splitting circuits that do not employ transformers or chokes are preferred in modern amplifiers.

The Split-Load Circuit

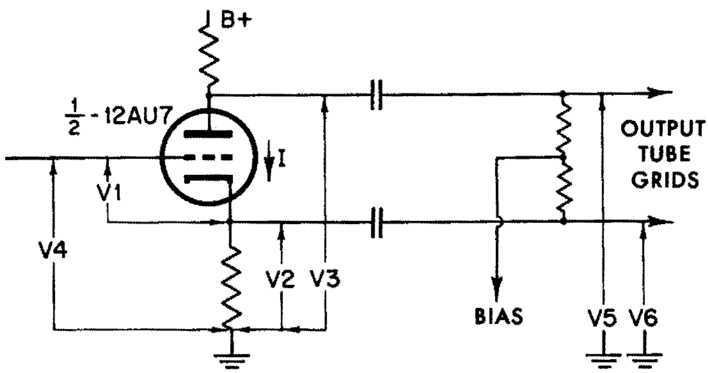




One simple way to make a phase splitter is to connect half of the plate load resistance between B+ and plate and the other half between cathode and ground. Since these resistances are equal and the same current flows through both, each will produce the same d-c voltage drop and the same audio fluctuations. When the fluctuation across the plate resistor goes positive, due to decrease in plate current, this same fluctuation will be negative across the cathode connected resistor because of the same decrease in current. This provides voltages of opposite phase, but we still must provide input to the tubes. The normal place to apply input voltage to a tube is between the grid and cathode, however, in this arrangement, half the total output voltage is between cathode and ground. This circuit uses the tube just to get phase *inversion* (to reverse the voltage between grid and plate) and does not achieve any useful amplification.

The Split-Load Circuit (contd.)

Using one-half of the 12AU7 as an example, an input fluctuation of 10 volts, measured between grid and *cathode,* will produce an output fluctuation of 100 volts—50 volts at the plate and 50 volts at the cathode. When the grid-to-cathode voltage goes 5 volts positive from its bias point, the current through the tube will increase. The cathode will go 25 volts positive from its d-c operating point, while the plate will go 25 volts negative from the d-c operating point. As the cathode has now gone 25 volts *more* positive from ground, and the grid-to-*cathode* voltage needs to go 5 volts positive to *cause*



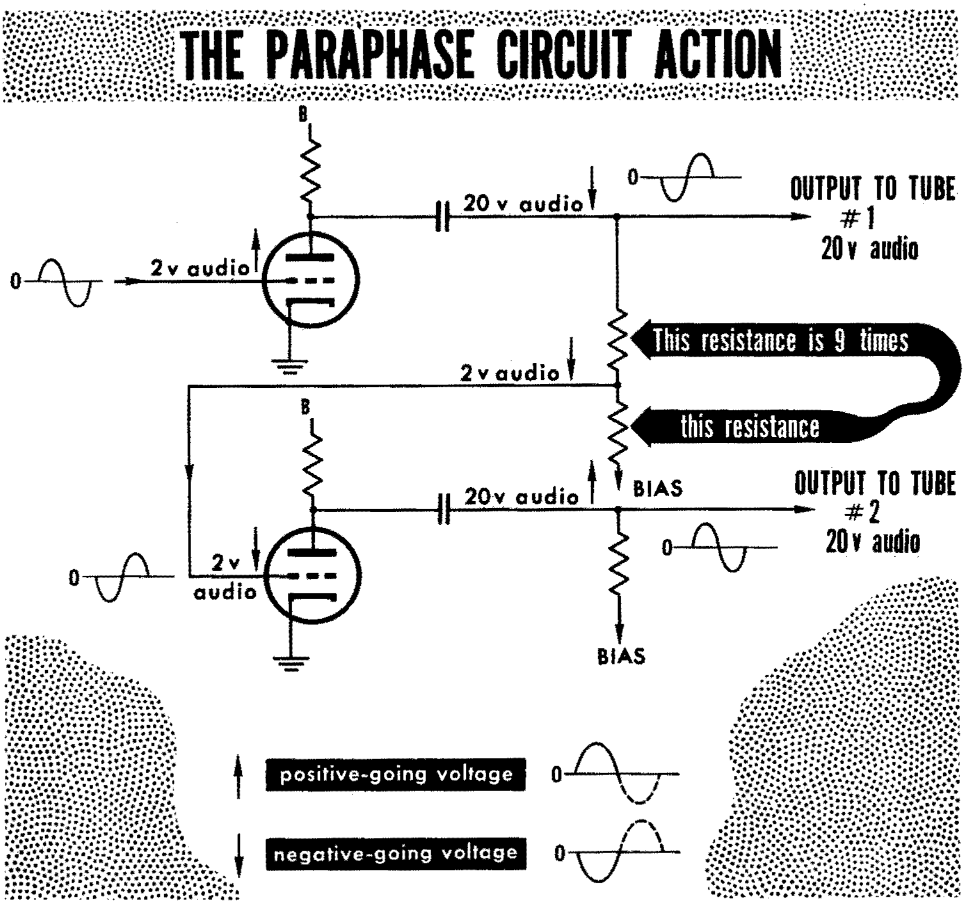
Currents and Voltages in Split-Load Circuit

Where Taken	TABLES OF VOLTAGES AND CURRENTS		
	No Audio	Positive Input Audio	Negative Input Audio
V1	-5V	0	-10V
I	5MA	7.5MA	2.5MA
V2	+50V	+75V	+25V
V3	+200V	+175V	+225V
V4	+45V	+75V	+15V
V5	BIAS	BIAS +25V	BIAS-25V
V6	BIAS	BIAS -25V	BIAS+25V

this, the grid-to-ground voltage itself must go 25 + 5, or 30 volts more positive to produce this swing. We therefore require a 60-volt peak-to-peak audio fluctuation from grid to *ground* to produce a 50-volt peak-to-peak fluctuation at plate and cathode, respectively.

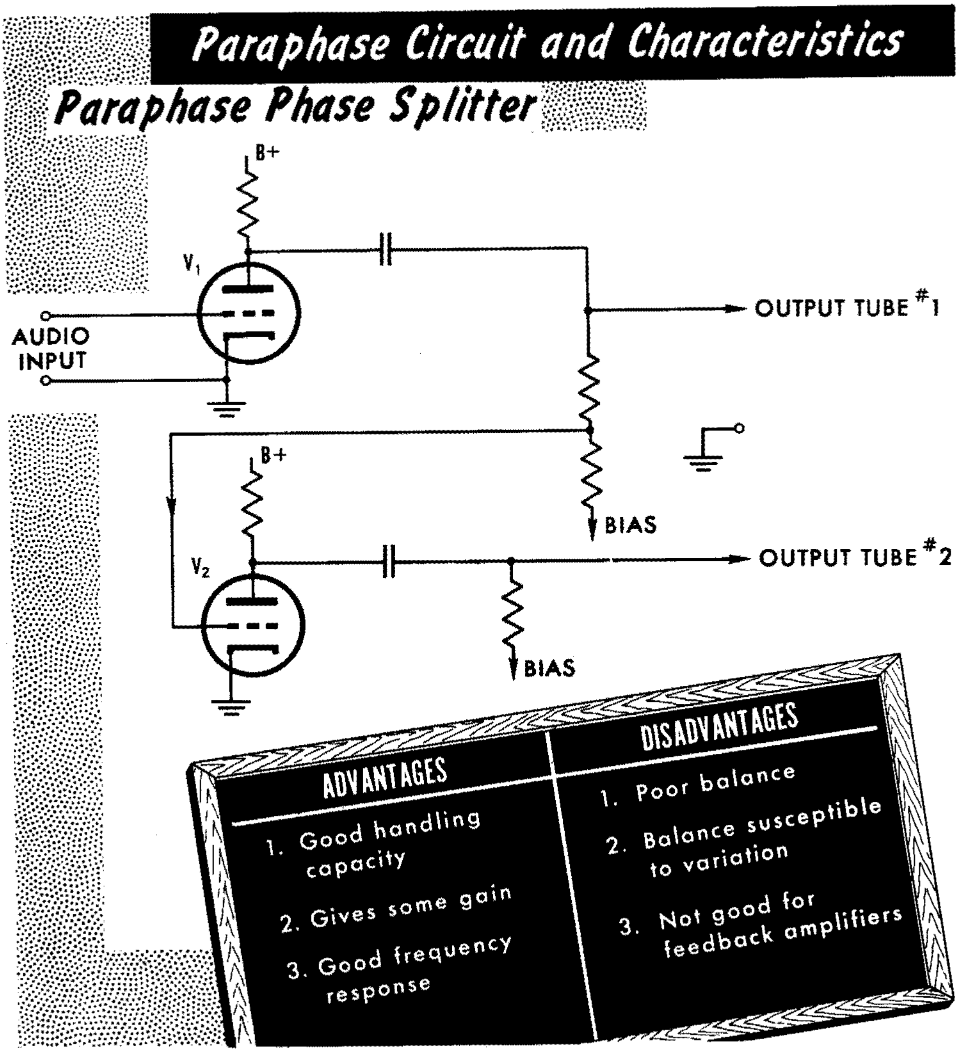
The Paraphase Circuit

Another phase-splitting arrangement is the so-called *paraphase* circuit. It uses two tubes. The output from the plate of one tube is fed by R-C coupling to the grid of one of the output tubes. From this same point, a voltage-divider arrangement cuts down the voltage and applies it to the grid of a second tube, which amplifies the voltage by as much as the resistance divider cuts it down, producing a voltage for driving the second output tube.



A positive fluctuation of 2 volts at the grid of the first paraphase tube produces a negative fluctuation at its plate of, say, 20 volts, which appears at the grid of one of the output tubes. This 20-volt fluctuation is also divided to provide a negative fluctuation of 2 volts for the grid of the second paraphase tube and becomes a positive fluctuation of 20 volts at the plate, providing positive fluctuation for the grid of the second output tube.

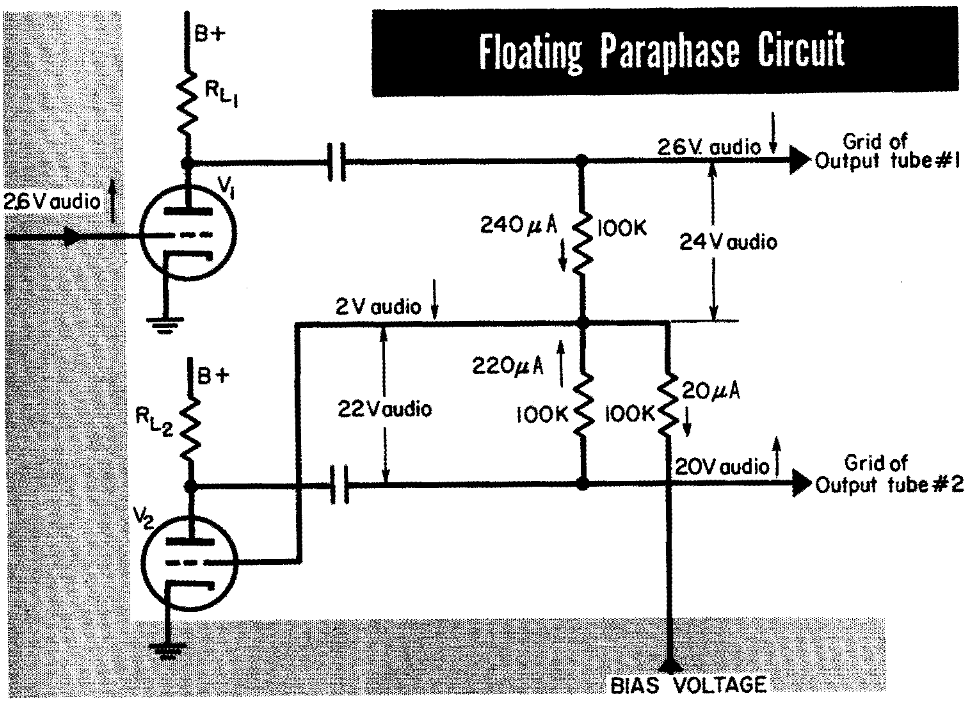
The Paraphase Circuit (contd.)



For this circuit to operate correctly, the voltage division produced by the resistors feeding the second tube must be exactly in the same ratio as the gain provided by the second tube. In the example given, the tube is multiplied by ten, and the voltage divider divided by ten. As tubes are subject to variation with line voltage, individual samples from production, and other differences, there is no guarantee that the amplification provided by the tube will be exactly the same as the voltage division provided by the resistors. The tube may amplify 9 or 11 times. Consequently this circuit is subject to deviation in its accuracy in a way that the two circuits discussed earlier were not.

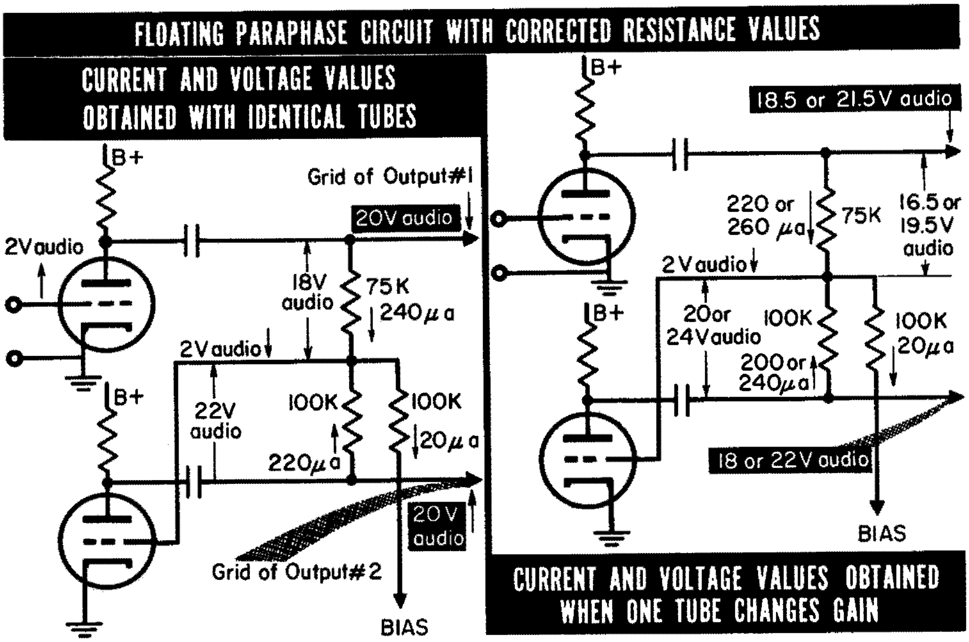
The Floating Paraphase Circuit

The next circuit aims at overcoming this problem. It is called *floating paraphase*. Instead of using a fixed voltage divider, the voltage divider is arranged to have a self-adjusting action that overcomes, to some extent, possible differences between individual tubes or changes in operating conditions. Instead of using a large resistor and a small resistor, as employed in the paraphase circuit, the tapping point used for connecting to the grid of the second paraphase tube is a junction point of large resistors from the output-tube grids and a third resistor from the bias point.



Assume that the three resistors have the same value, 100,000 ohms, and that the tubes have a gain of 10, as before. For the second tube to get its 2-volt grid fluctuation, the current through the common resistor must be 2/100,000 or 20 microamperes. The voltage at the grid of the second output tube will fluctuate 20 volts in the opposite direction, hence the resistor connecting this grid to the common point will have 22 volts across it and pass 220 microamperes. Because the resistor from the first output-tube grid joins this same point, it must have 220 + 20 or 240 microamperes flowing in it, which will produce a fluctuation voltage across it of 24 volts. As the common point is already fluctuating 2 volts in this direction, the total fluctuation at the first output tube grid must be 26 volts.

The Floating Paraphase Circuit (contd.)



The unbalance between output tube grids can be corrected by using a smaller resistance value from the first output tube, so that 240 microamps only drop 18 volts. This requires a resistor of 18/0.00024 = 75,000 ohms, in place of 100,000 ohms.

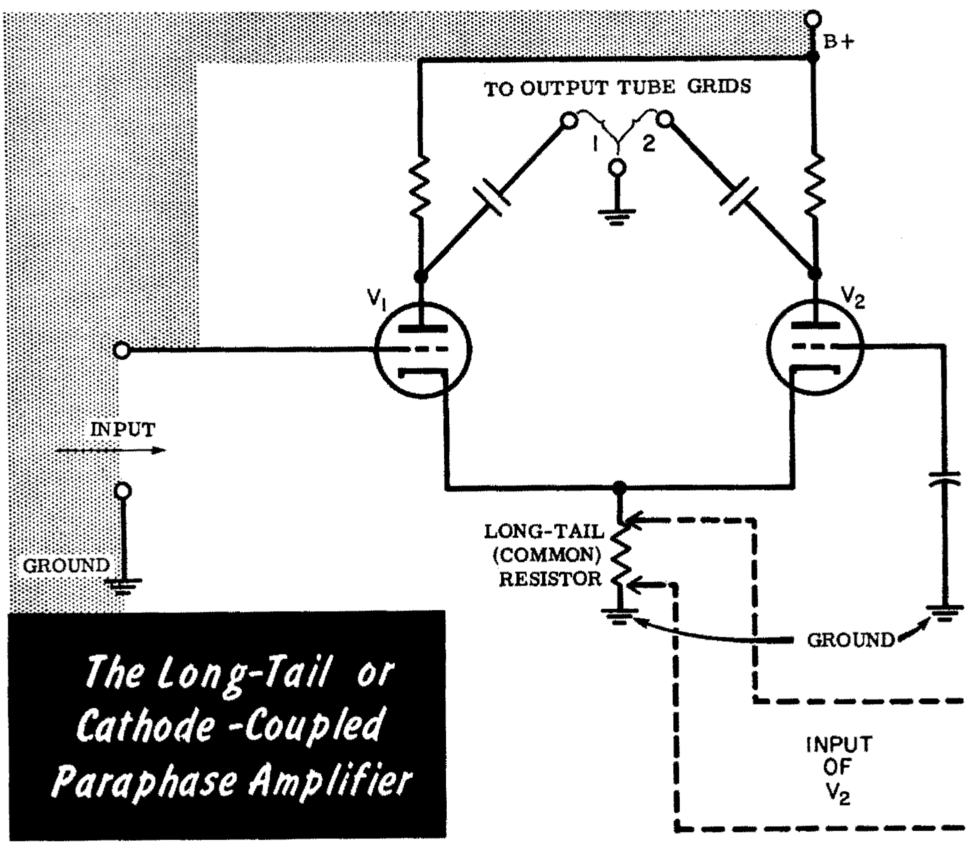
Now what happens if the second tube changes gain to either 9 or 11? If it still gets 2 volts on its grid, there will be either 18 or 22 volts fluctuation at its plate, so the current in the resistor from the second output tube grid to the common point will be either 200 or 240 microamperes. The current in the resistor from the other output tube grid will need to be 220 or 260 microamps, instead of the original 240 microamps. This means that the potential at that grid will be (75,000 × .00022) or 16.5 volts, or (75,000 × .00026) or 19.5 volts, from the grid to the common point. This represents 18.5 or 21.5 volts total fluctuation at this grid, to compare with 18 or 22 volts at the other grid.

With the ordinary paraphase, the grid of the first output tube would remain at a potential of 20 volts, while the second output tube would get 18 or 22 volts. Here, when the second tube gets 18 volts, the first has 18.5 volts; when the second gets 22 volts, the first gets 21.5 volts, which reduces the imbalance. If the first grid continued to get 20 volts, the second grid would get within 19.5 to 20.5 volts with this much change in gain, instead of 18 or 22 volts with ordinary paraphase.

The Long-Tail Circuit

Another variation puts the common resistor that carries the plate current of both tubes in the circuit between cathode and ground. The input voltage to the second tube is the audio voltage developed across this common resistor, because the second tube grid is connected to ground either directly or through a large capacitor.

It doesn't matter which way we measure a voltage, whether from grid to cathode or cathode to grid. In either case it will be the same *voltage*, different only in polarity or phase. Hence the grid voltage for the second tube will be due to the difference in the audio components of plate current in the common cathode resistor. This means that the audio plate current of the first tube must be a little higher than the corresponding component of the second tube. If identical plate resistors are used, the first tube will produce a slightly greater audio voltage than the second tube. This can again be overcome by using plate resistors of different values so that the audio voltages become equal.



The Long-Tail or Cathode-Coupled Paraphase Amplifier

QUESTIONS AND PROBLEMS

1. What are the advantages of working tubes in push-pull?

2. Explain how using tubes in push-pull enables them to be used (a) with less distortion, (b) to give more power (higher efficiency)?

3. Give the relative advantages of operating tubes in (a) class A, (b) class B, (c) class AB.

4. What is power drive, and how does it alter the requirements of the preceding stage?

5. What is the most informative way of displaying the distortion performance of an amplifier using an oscilloscope?

6. What is a phase splitter and when is it needed? Discuss the relative features of the following different types, with special reference for each to (a) accuracy of the two outputs, (b) uniformity of performance at different frequencies, (c) economical use of components:
   (i) Transformer type
   (ii) Split-load type
   (iii) Paraphase type
   (iv) Floating paraphase type
   (v) Long-tail type
   (vi) Paraphase from output transformer.

7. Compare different output circuits, using triode or pentode tubes, in various classes of operation, with regard to the power output that can be obtained for a given dissipation rating, to probable distortion content, to critical load requirements, and to the requirement of critical adjustment in operation.

8. In spite of the fact that it is evidently a good phase-splitting device, the transformer is very seldom used in modern amplifiers. Why?

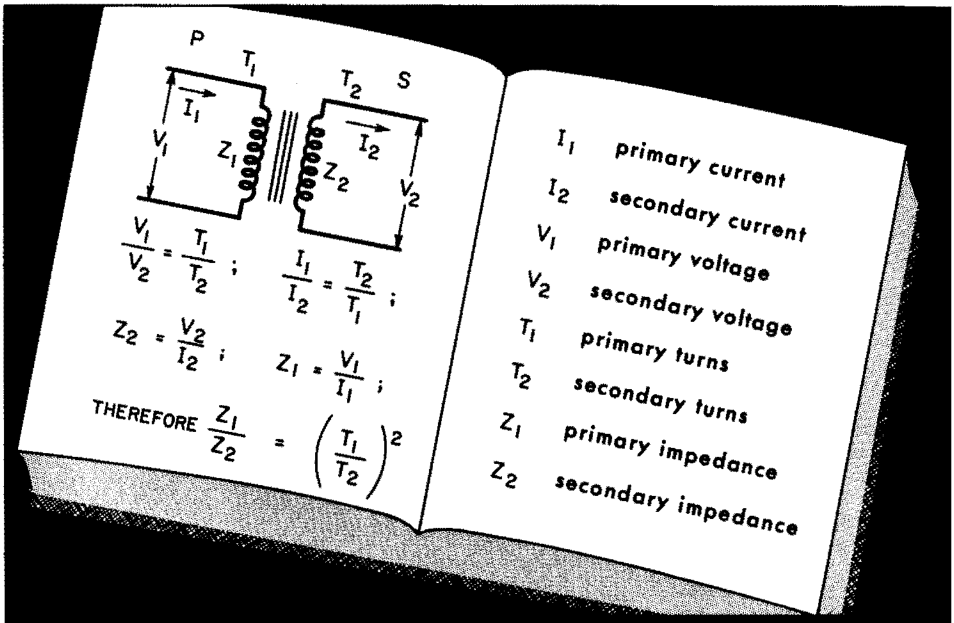
9. The split-load circuit is criticized for not having any useful gain. Explain why this is. Why is its handling capacity less than some other types?

10. Describe, with numerical illustration to prove your point, why the floating paraphase maintains betters balance than the simple paraphase circuit.

11. In a floating paraphase circuit, the section of grid resistor common to both output tubes has one-half the average value of the other two, which are slightly unequal, so that when the gain of the phase-splitting tube is 20, both output tubes get equal drive. Using values to illustrate this condition, find how much error will result from the phase-splitter tube's losing working gain down to 15.

Review of Transformer Action

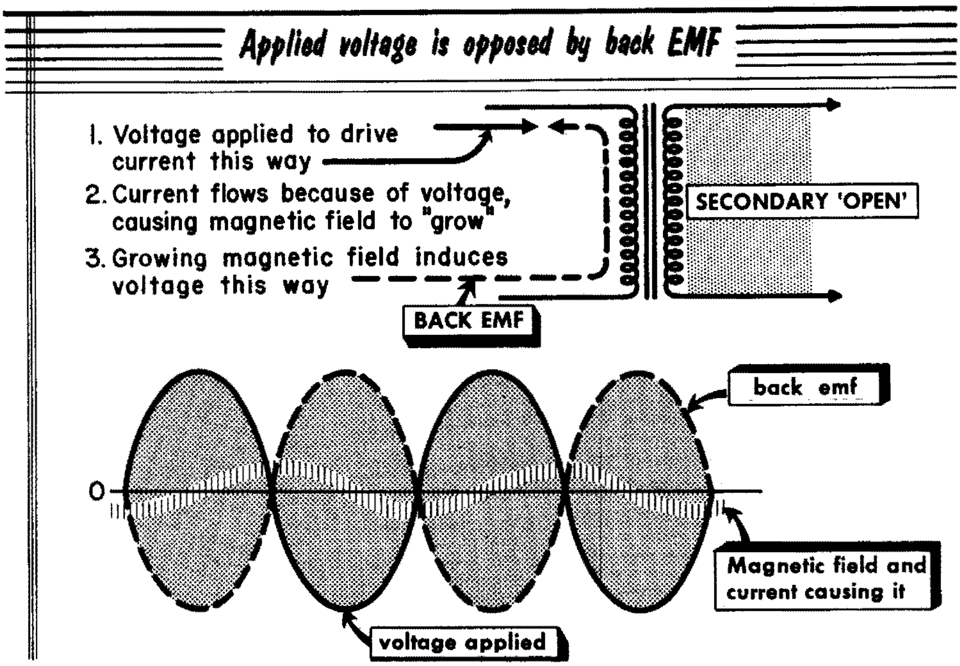
Audio transformers are used to match one circuit to another. For an ideal transformer, the ratio of the primary and secondary voltages varies directly with the ratio of the number of turns in the two windings, the ratio of primary and secondary current varies inversely with the turns ratio, and the primary and secondary impedance varies directly with the square of the turns ratio.



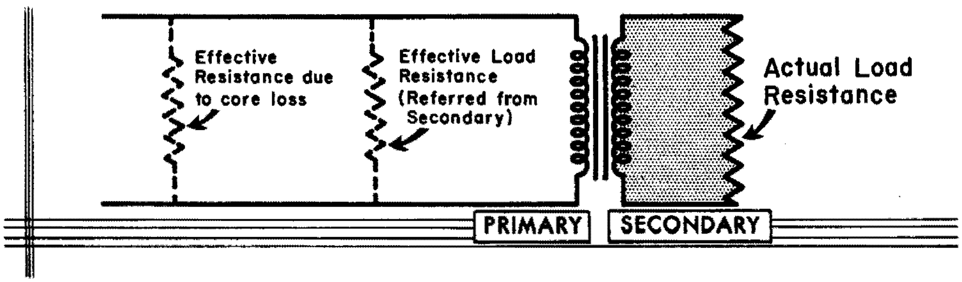
SUMMARY OF TRANSFORMER ACTION

These relationships may be explained as follows: the first two arise from the fact that (in an ideal transformer) the magnetization (measured in ampere-turns) of the core is zero at all times. If current is drawn from the secondary, the product of this current and the number of secondary turns must be cancelled by the product of an equal and opposite product of primary current and the number of primary turns. Since the ideal transformer neither adds nor absorbs power, the product of the primary voltage and current must equal the product of the secondary voltage and current. Thus current varies inversely with the turns ratio and voltage varies inversely with current or directly with the turns ratio. Since the secondary impedance is measured by the ratio of secondary voltage to secondary current, and the voltage has *increased* by the turns ratio while the current has *decreased* by the turns ratio, the impedance must change by the square of turns ratio.

Core Losses

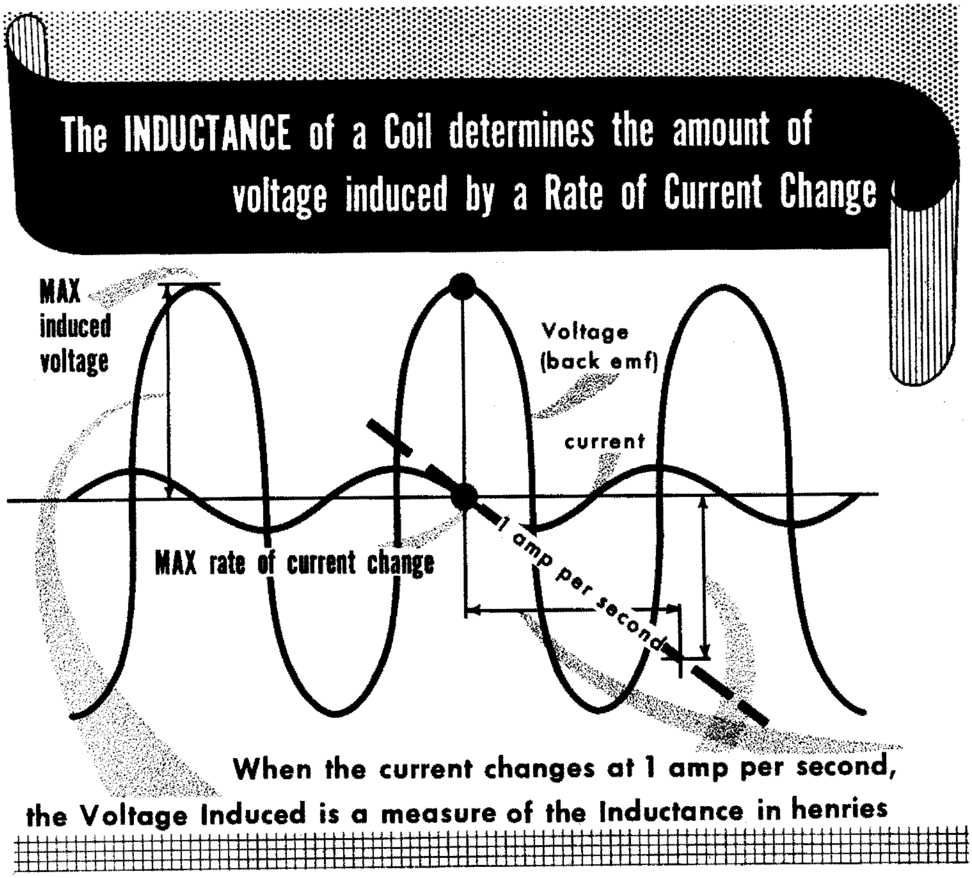


Real transformers are not the same as ideal ones, however. Some current flows in the primary even when no secondary current flows. This current induces a voltage *(back emf)* in the primary that opposes the applied voltage. This opposing voltage prevents the flow of excessive current in the low-resistance primary and, since the magnetization of the core depends on the *current* in the windings, it limits core magnetization as well. It is because the back emf does not completely cancel the applied voltage that primary current flows in a real transformer. In a properly designed transformer, this current is quite small, but it represents loss that can be visualized as a resistor connected across the primary that acts in addition to the resistance referred from the secondary load.



Back Emf and Saturation

The magnitude of the back emf developed in the primary by an applied a-c voltage varies with the rate at which the current developed by the voltage changes. The more rapidly this current changes, the greater the back emf produced. The property of the primary that relates the rate of change of the current in the primary windings to the resultant back emf is known as the *primary inductance.* When the primary inductance is high, the back emf is high, and the magnetizing current is small.

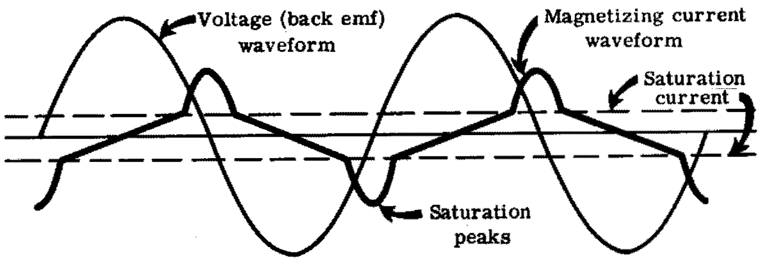


There is a limit to the magnetization that a core can take, called *saturation.* Below this point, the primary inductance remains nearly constant and is relatively high. When saturation occurs, the relation between voltage induced and the further increase in current needed to produce it ceases to be approximately constant. It is as if the magnetic material disappeared and we had only an air-core inductor.

The Effect of Saturation

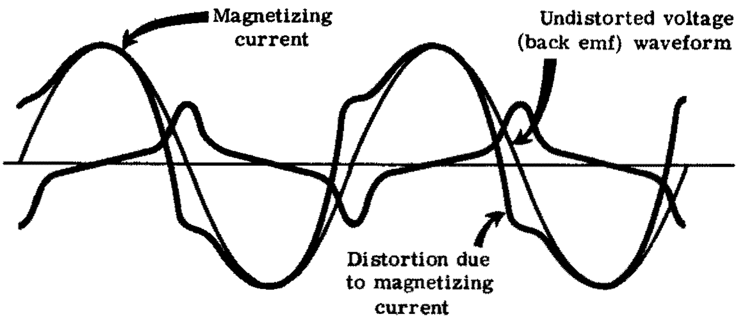
Saturation begins at a specific magnetization density. The back emf produced by this density depends on frequency, being directly proportional to it. This means that the power-handling capacity of a transformer is approximately proportional to the square of frequency. It is not a constant figure! Below saturation the relation between voltage and magnetizing current approximates that of simple inductance. Above the point on the current wave where saturation occurs, the current peaks up rapidly.

Waveforms when saturation occurs



If the primary of the transformer were connected to a source of voltage having zero resistance, the sudden rise in magnetizing current due to saturation would make no difference. Drawing this heavy current from a voltage source that possesses internal resistance, however, causes a voltage drop coincident with the current peaks, distorting the primary voltage waveform as well.

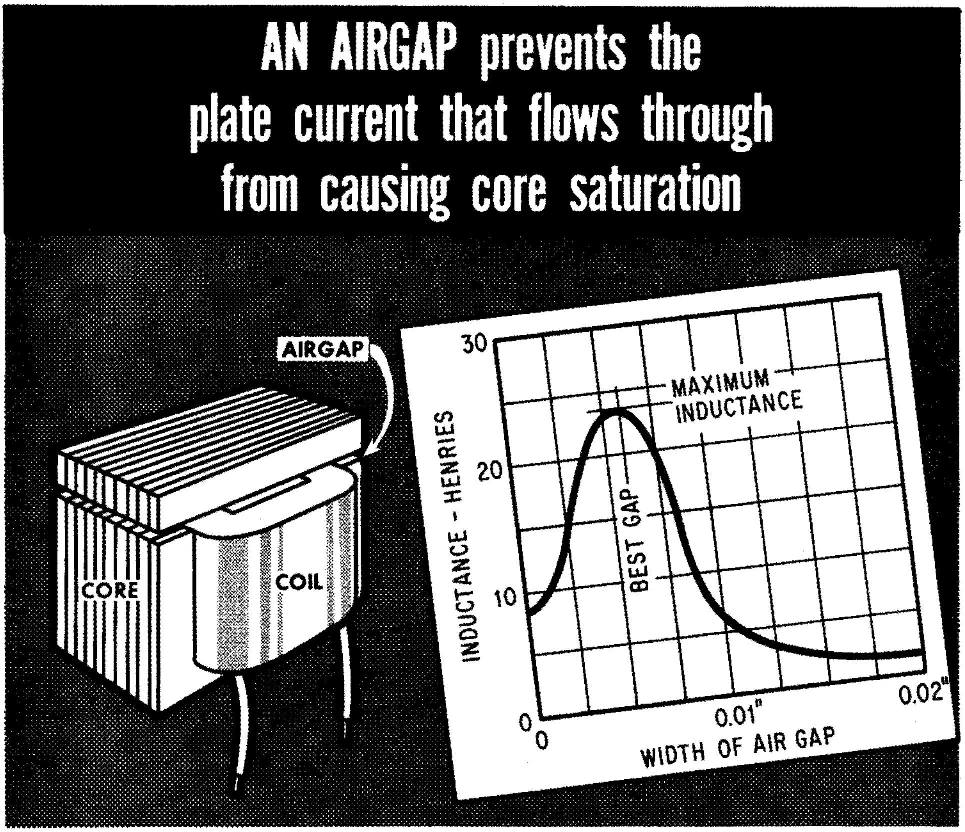
Effect of saturation



Frequency response below saturation is controlled by the relationship between primary inductance and the associated source and local resistances. Regardless of the saturation point, however, primary inductance causes a low-frequency loss at all voltage levels.

The Use of an Air Gap

With a good transformer core, very little current is needed to produce saturation. This means that passing the plate current through the primary winding will saturate the core unless some means is used to prevent it. Usually, an air gap is used for this purpose. It very considerably reduces the magnetization produced by any current in the primary winding, but it also reduces the primary inductance. The air gap has to be adjusted so that its effect in avoiding saturation is better than its effect in reducing primary inductance.



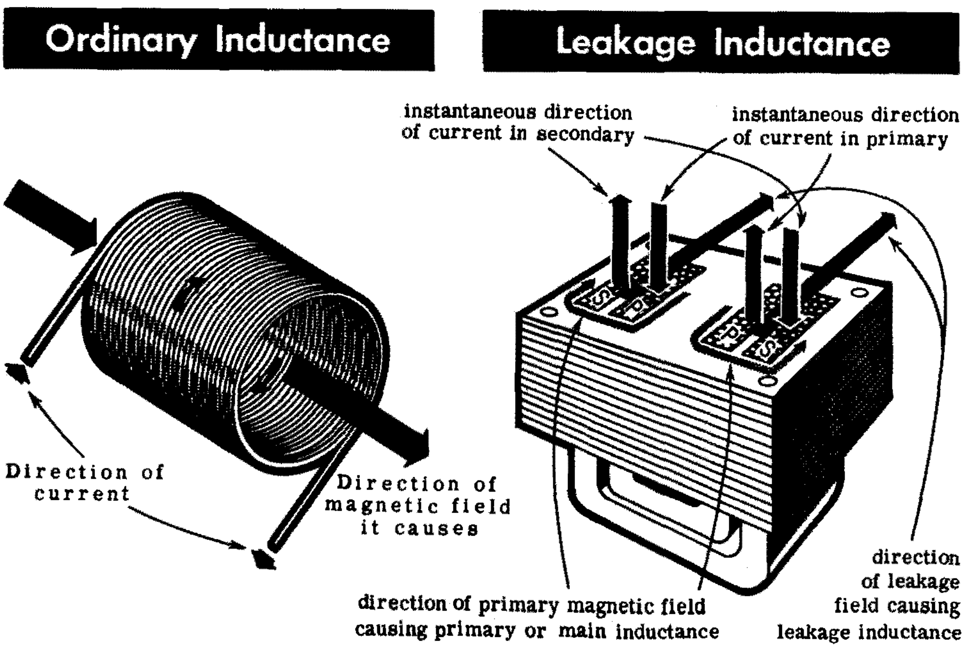
Using too small an air gap, or no air gap at all, means the primary current will cause the core to saturate and thus the inductance will practically disappear. Using too large an air gap amounts to making the transformer almost air cored, which again will result in a very low inductance. There is always an optimum size for the air gap, which will result in the biggest practical value of inductance for a transformer that carries direct current in one of its windings.

Other Losses

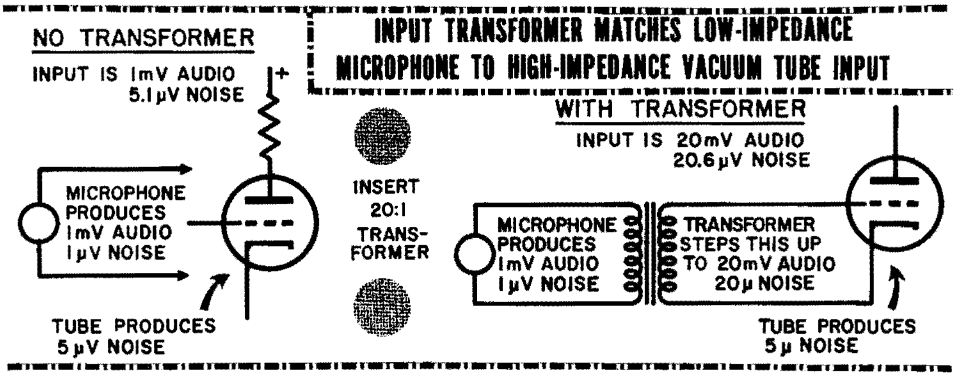
Primary inductance and core losses, together with distortion, are the principal things that complicate the behavior of an audio transformer at low frequencies. At high frequencies the behavior of the transformer is complicated by winding capacitance and leakage inductance. Winding capacitance allows minute audio currents to pass between points of audio voltage where there is no direct connection, in the way that stray capacitance allows audio current to pass at extremely high frequencies.

The magnetic field causing primary inductance follows a path through the magnetic core that embraces both windings. The current in the windings themselves, however, induces a magnetic field around each winding (apart from the bigger one in the core) that tends to "leak down" between the two windings. This means that the current in one winding will not produce a field that counter-balances the current in the other winding. Consequently, these uncounterbalanced fields will produce different voltages in their respective windings. Thus *leakage inductance* is like an inductance in series with the winding, because it allows an additional voltage to be developed between the terminals that is not accounted for by the counter-balancing effect of current in the two windings.

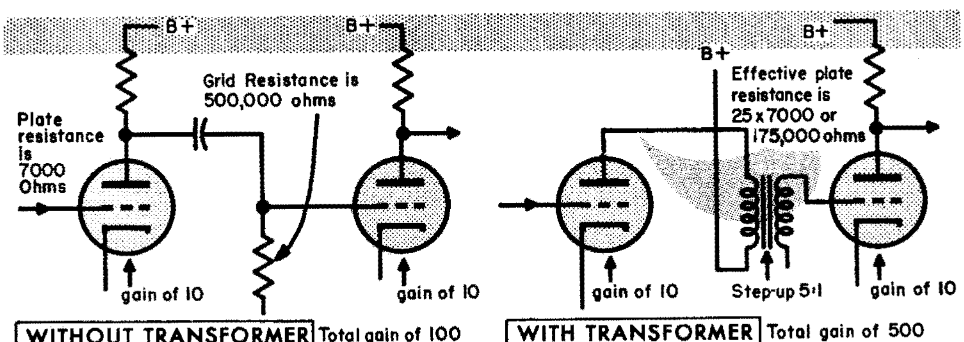
A well-designed audio transformer has to take into account all of these factors to make sure that its performance as an impedance-matching device is consistent over the frequency range for which it is intended.



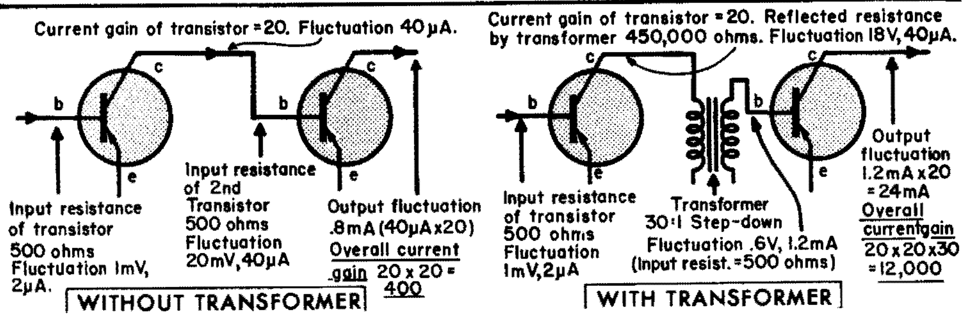
Input and Interstage Transformers

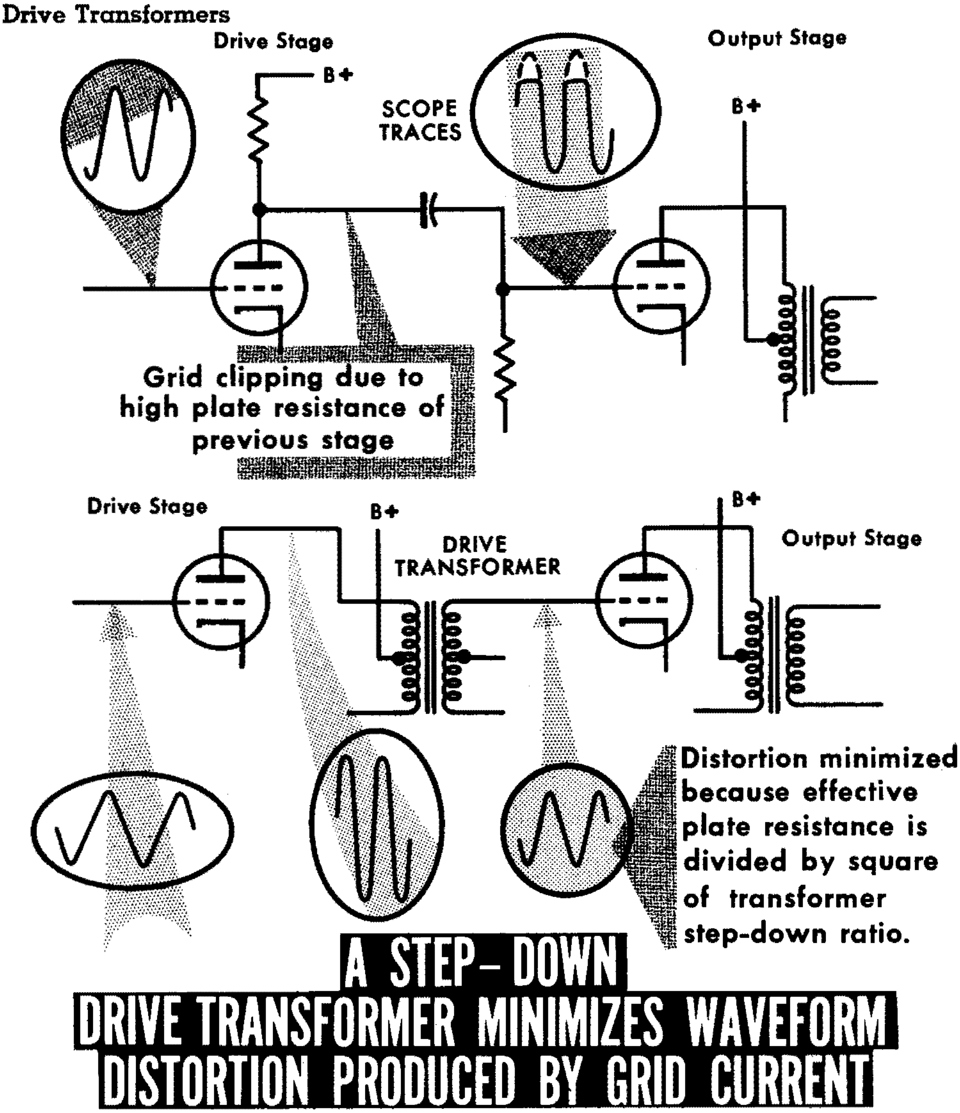


Audio transformers are used where the impedance is not naturally what is needed for best performance. For low-impedance pickups or microphones working into tube grids, the transformer steps up the minute voltages (in this case impedance is incidental) to achieve better margin over tube noise. Sometimes interstage transformers are used for essentially the same purpose. The limit to what can be done occurs when the impedance connected to the primary is stepped up to such a value that it adversely affects performance in some other way.



Interstage transformers can be used to increase the effective gain of an amplifier





Step-down transformers can also be used in the drive stage that feeds a push-pull output stage, where the grids of the output tubes are driven into the positive voltage region during part of the audio waveform. This means there will be current peaks which tend to distort the audio voltage at the grid, caused by a drop in the high source resistance (usually the plate resistance of the preceding stage). The apparent source resistance can be reduced down by a step-down drive transformer, which thus minimizes the waveform distortion produced by the grid current.

Output Transformers

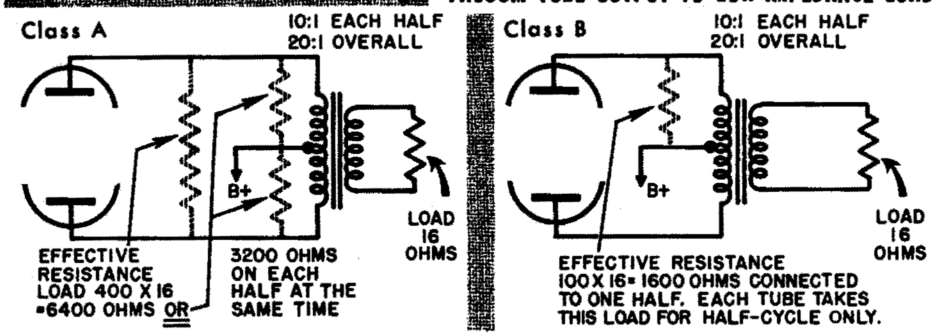
The most common form of audio transformer is used in the output of an amplifier to match the actual load impedance (usually that of a loudspeaker) to that required by the output tubes. The transformer here serves the additional purpose of avoiding both supply and audio losses because the winding resistances are low compared to their respective impedances. The way impedance reflects in a push-pull transformer depends to some extent on the way tubes are operated. In class A, both tubes are delivering part of the power throughout the cycle, so the load is shared between them. If the ratio make 16 ohms actual impedance equivalent to 6400 ohms at the primary, each tube has a load of 3200 ohms average.

But in class B, only one-half of the primary works at a time. The other is inactive for that half-cycle because its tube is cut off. Consequently the impedance transformation is based on the ratio to each half-winding. If the whole ratio is 20:1, this is 10:1 each half. So 16 ohms connected to the secondary makes a load of 1600 ohms for each tube, but the tube takes the load for only half a cycle.

A further advantage of push-pull operation in the transformer is that the magnetizing effect due to steady plate current cancels, whether the tubes are operated class A or class B. This means that no air gap is necessary to prevent saturation. In turn, this allows a much smaller core to be used for providing an adequate primary inductance with the available turns in the primary winding.

Of particular importance in output transformers is the leakage inductance between windings. Due to the load current in both windings, the leakage flux induces a voltage difference between the ideal ratio and what you might actually measure. Being an inductance, this voltage difference becomes larger with higher frequency. Its effect is to "uncouple" the load from the tubes at these higher frequencies, resulting in high-frequency loss. This uncoupling effect of the leakage inductance is like connecting an inductance in series with the load impedance.

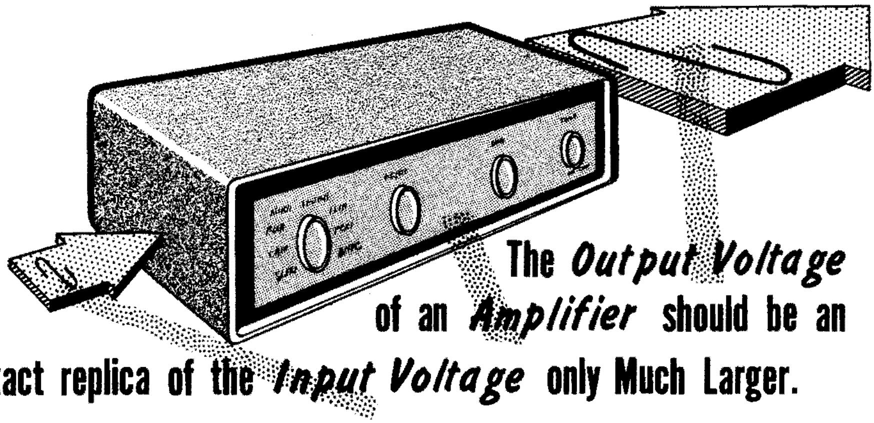
OUTPUT TRANSFORMER MATCHES HIGH-IMPEDANCE VACUUM TUBE OUTPUT TO LOW-IMPEDANCE LOAD

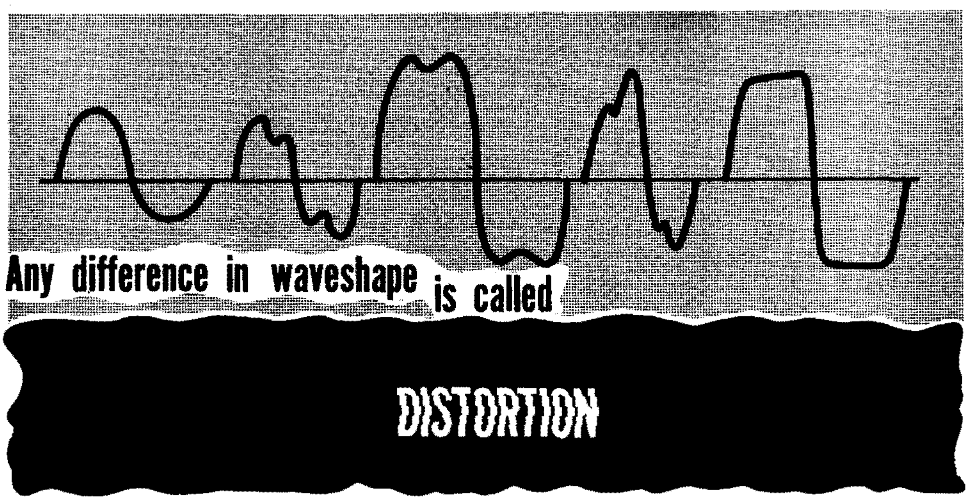


Harmonic Distortion

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

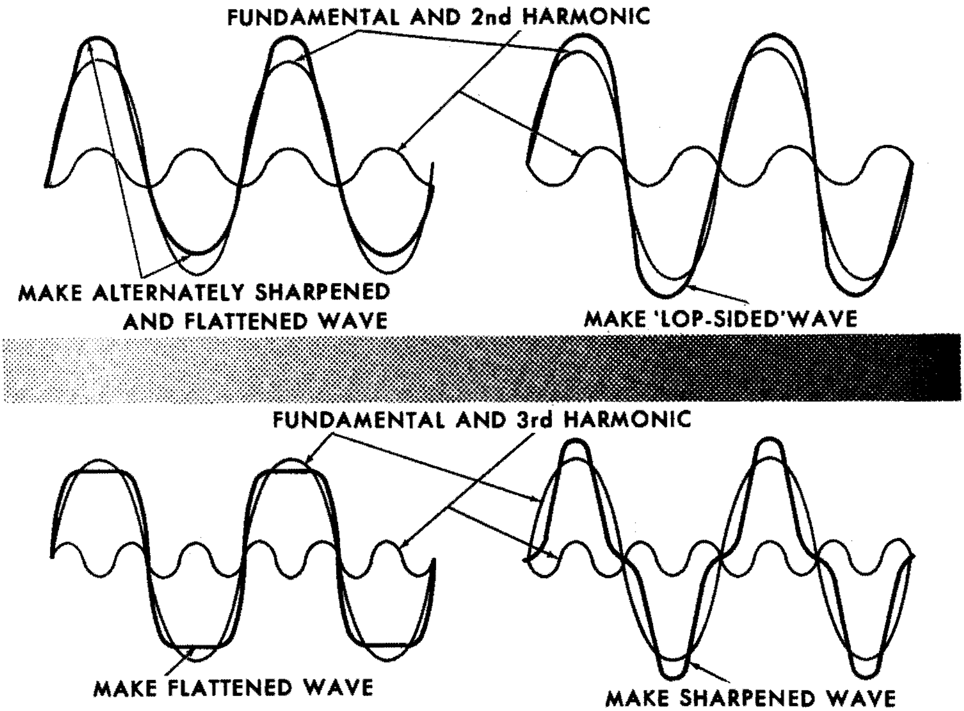
When the word distortion is used without qualification, it is usually taken to mean the kind of distortion due to curved or nonlinear characteristics in the amplifier. The fact that a change in audio voltage at the input is not accompanied by an exactly corresponding change at the output, at different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





Harmonic Distortion (contd.)

Distortion due to 2nd or even-numbered harmonics



Distortion due to 3rd or odd-numbered harmonics

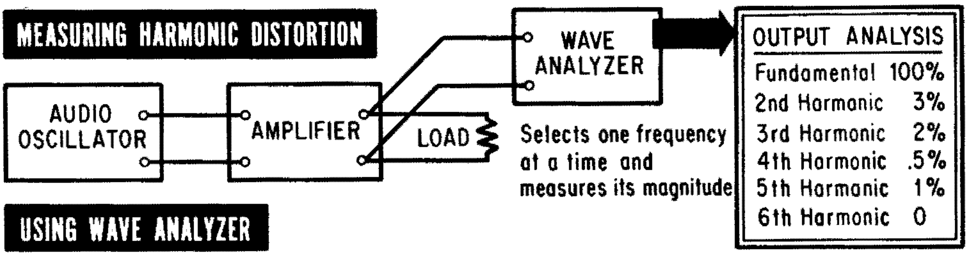
If we consider what this kind of curvature does to the amplification of a wave of single frequency, we can see how it introduces harmonic distortion —the presence of overtones of the fundamental frequency that are not present at the input.

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

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

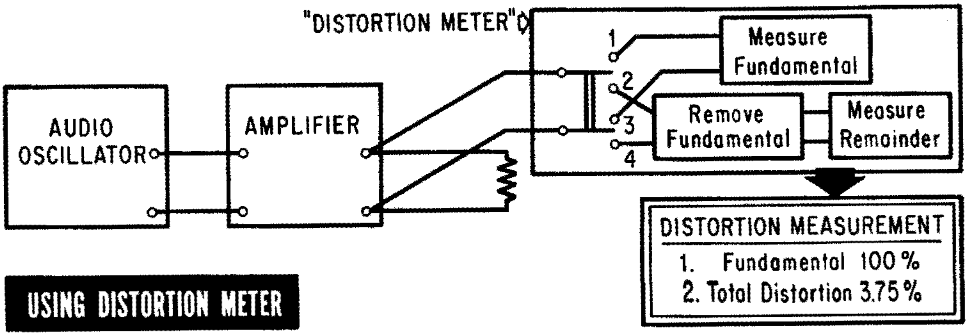
Measurement of Harmonic Distortion

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



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

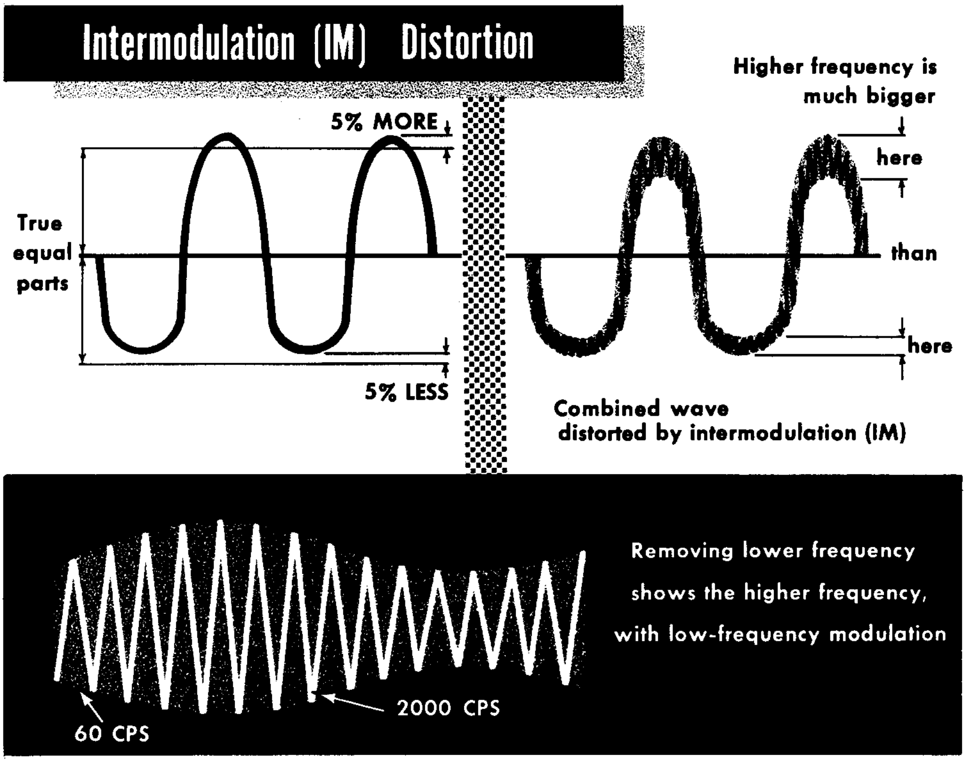
In the early days of audio amplifiers 5% harmonic was considered a good figure of distortion. At that time, tests were made which showed that human hearing could barely detect 5% of second harmonic. If the distortion is third harmonic, about 1.5% is just audible. At higher harmonics lower percentages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



Intermodulation Distortion

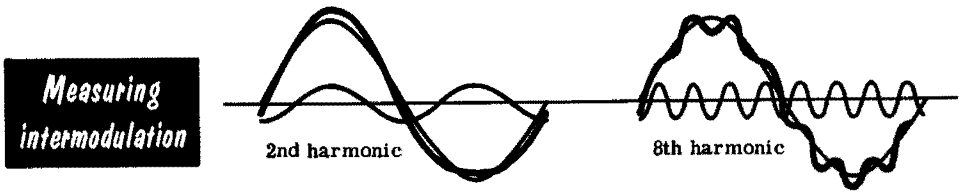
Low percentages of harmonic distortion may be inaudible, as such, but the same curvature in the amplifier characteristics causes another kind of distortion, called *intermodulation distortion* (IM for short). The effects of this kind of distortion can be audible when the harmonic distortion is not. There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

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



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

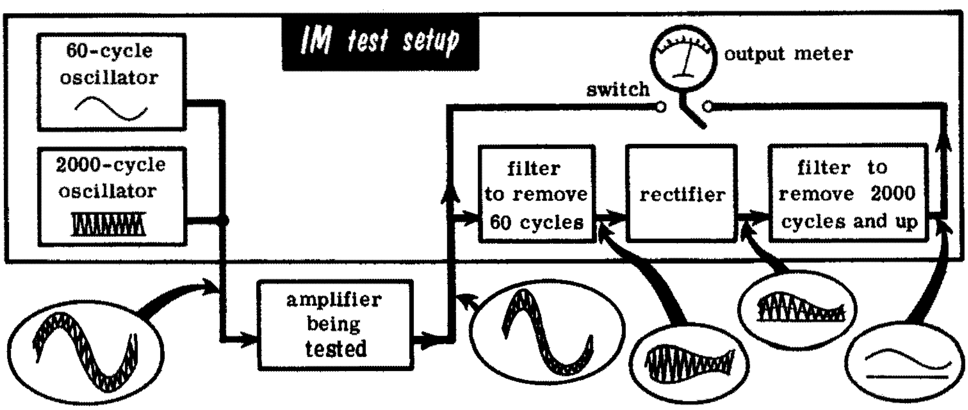
Intermodulation Distortion (contd.)



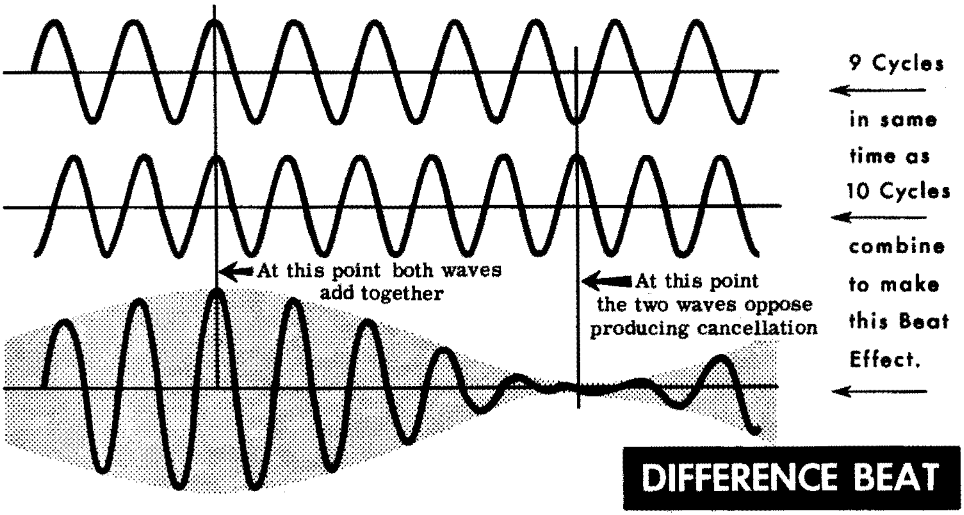
Both harmonics are same magnitude, but the higher numbered one has much bigger effect on the shape of the combined waveform.

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

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

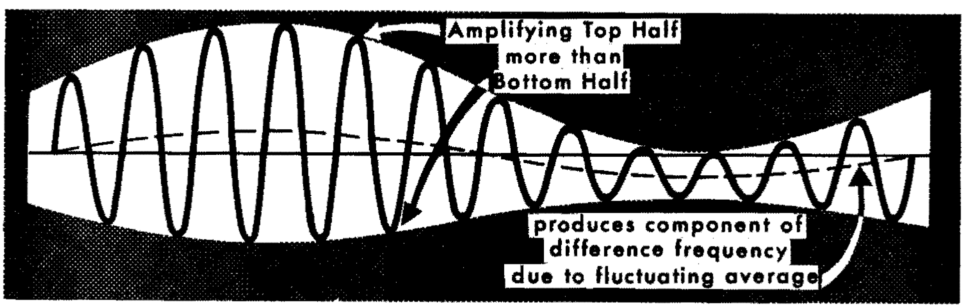


Intermodulation Distortion (contd.)



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

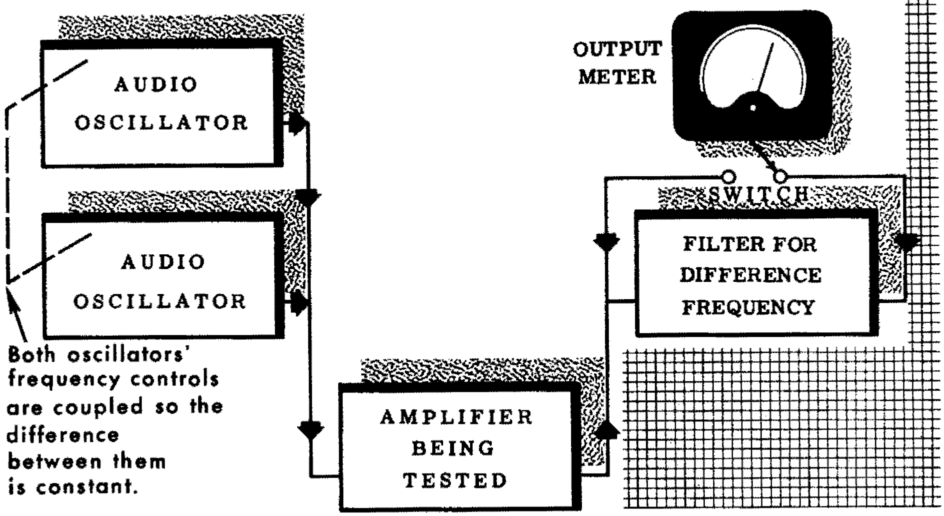
If this combination is applied to an amplifier without distortion, the waveform will be faithfully reproduced as would any other waveform. If the amplifier introduces any asymmetrical distortion, however, the upper part, at the peak in the combined waveform due to addition of the two components, will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



Intermodulation Distortion (contd.)

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

Test Setup to Measure IM Distortion Caused by Two High-Frequency Components

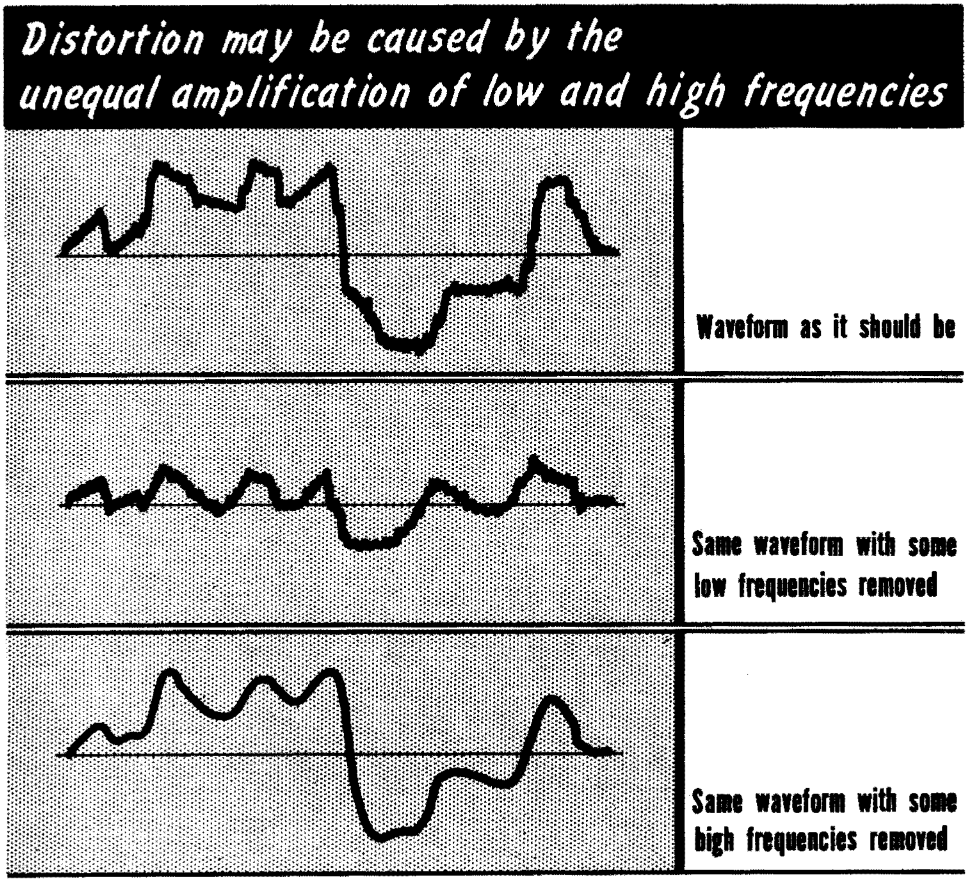


The method of testing for this kind of distortion is to use two oscillators that generate audio frequencies differing by a fixed amount. For example, if we decide to use the 200-cycle difference frequncy, we would arrange that one oscillator give 4000 cycles and the other one 4200 cycles. Or, to test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

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

Frequency Response

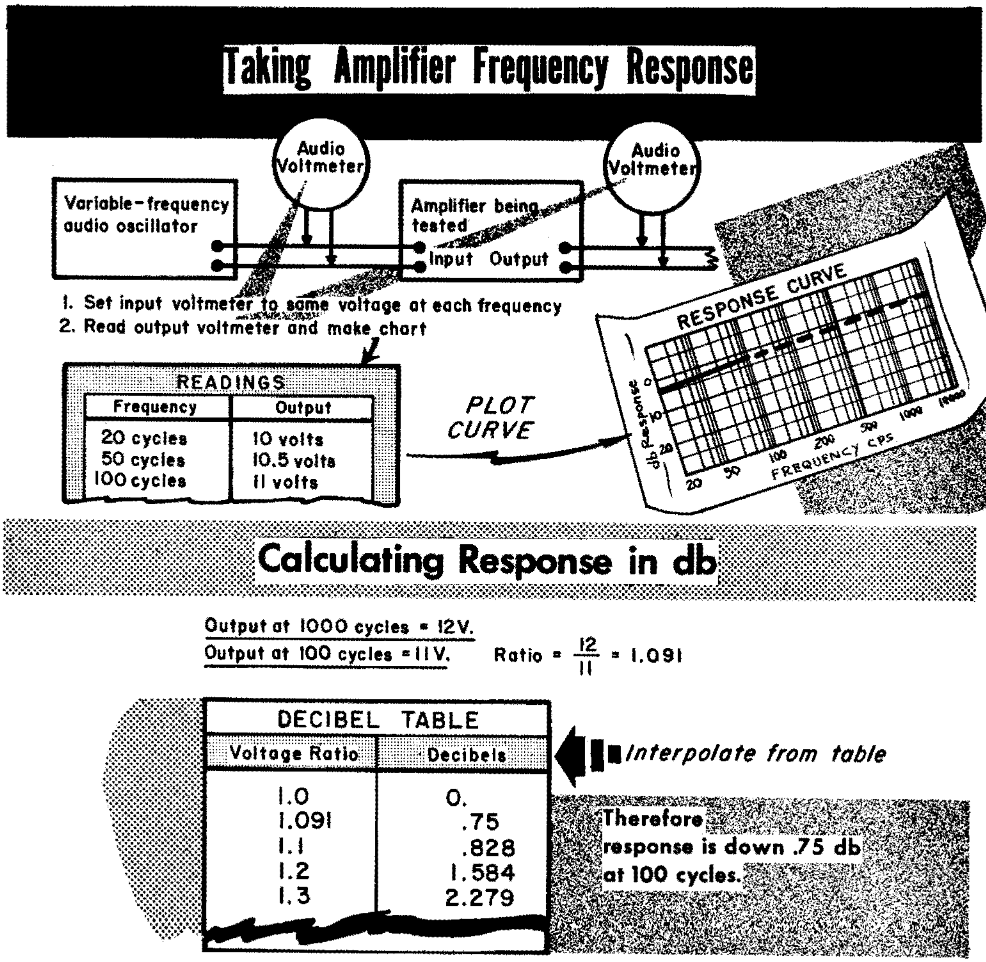
Amplifiers produce another kind of distortion because they do not amplify all frequencies uniformly (by the same amount). Low frequencies are reduced in amplitude by the effect of coupling capacitors. Various stray capacitances and the leakage inductance in the output transformer result in loss of high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



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

Frequency Response (contd.)

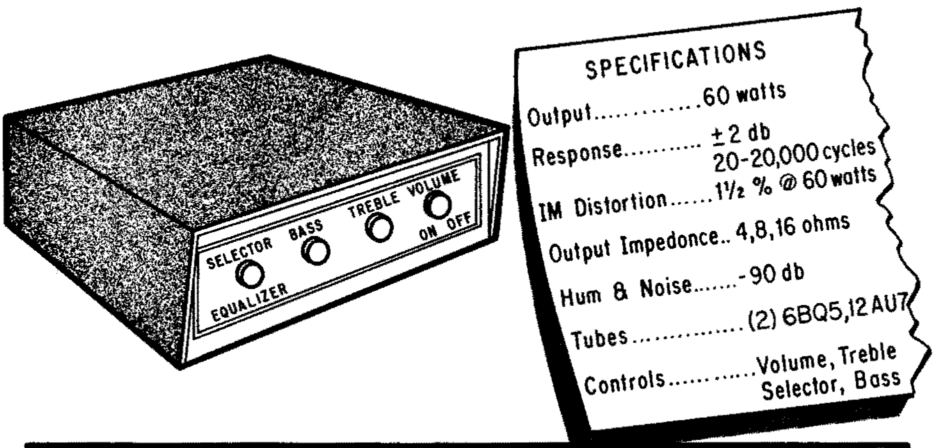
This form of distortion can readily be measured by taking frequency-response measurements of the amplifier. Audio voltages are fed into the amplifier at different frequencies, from the lowest to the highest, and the output voltage is carefully measured to see how closely it corresponds with the input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

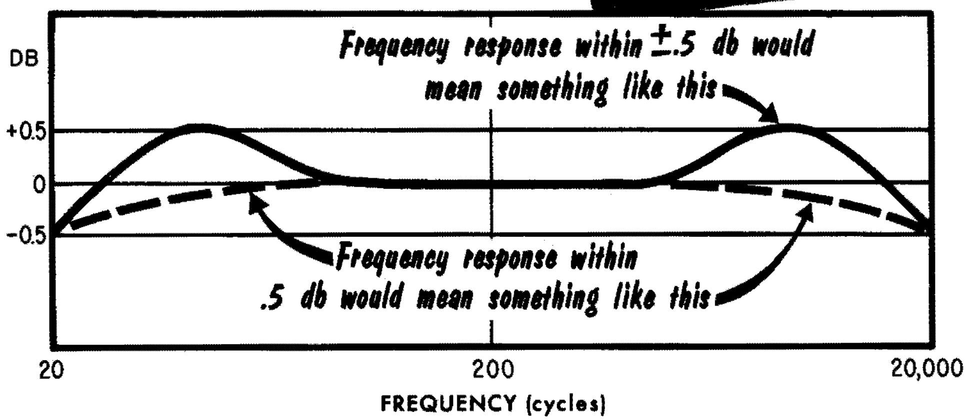


Amplifier Specifications

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

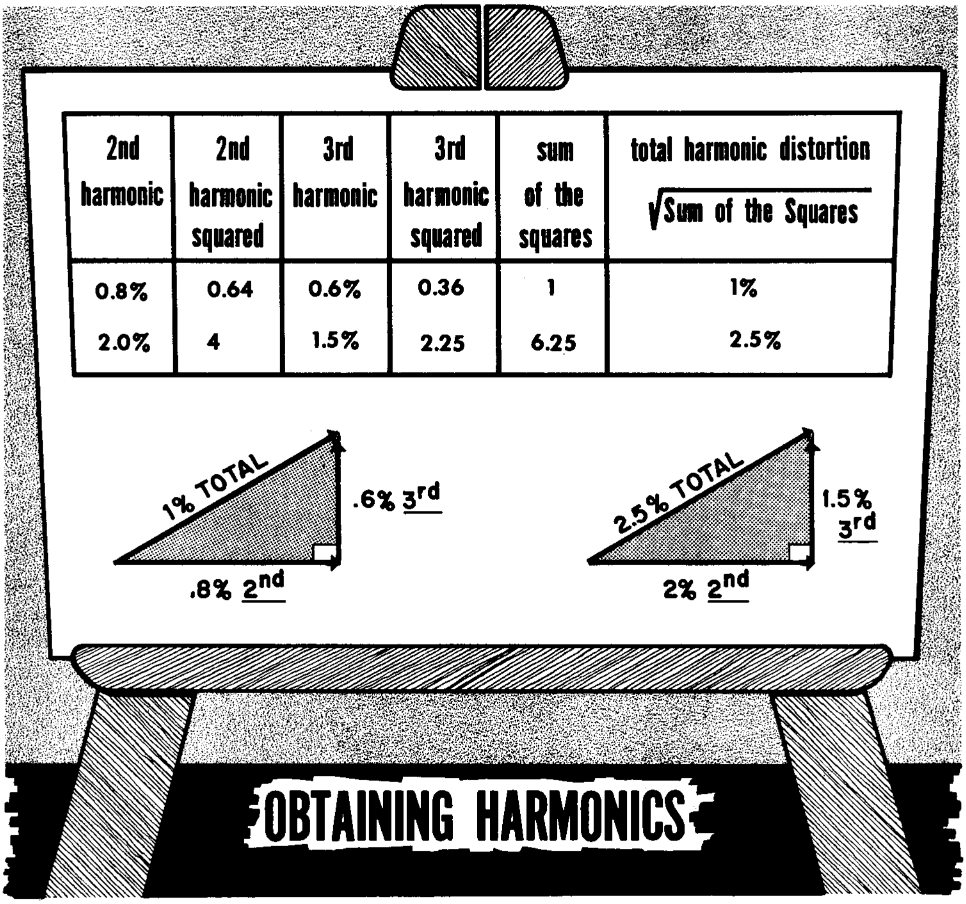
A distortion figure is also given and, unless otherwise specified, this indicates the amount of harmonic distortion. Unfortunately, it is not usual or convenient to specify what kind of harmonic distortion the figure given may be—entirely second harmonic, third harmonic, or it may be a composite of higher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





Amplifier Specifications (contd.)

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

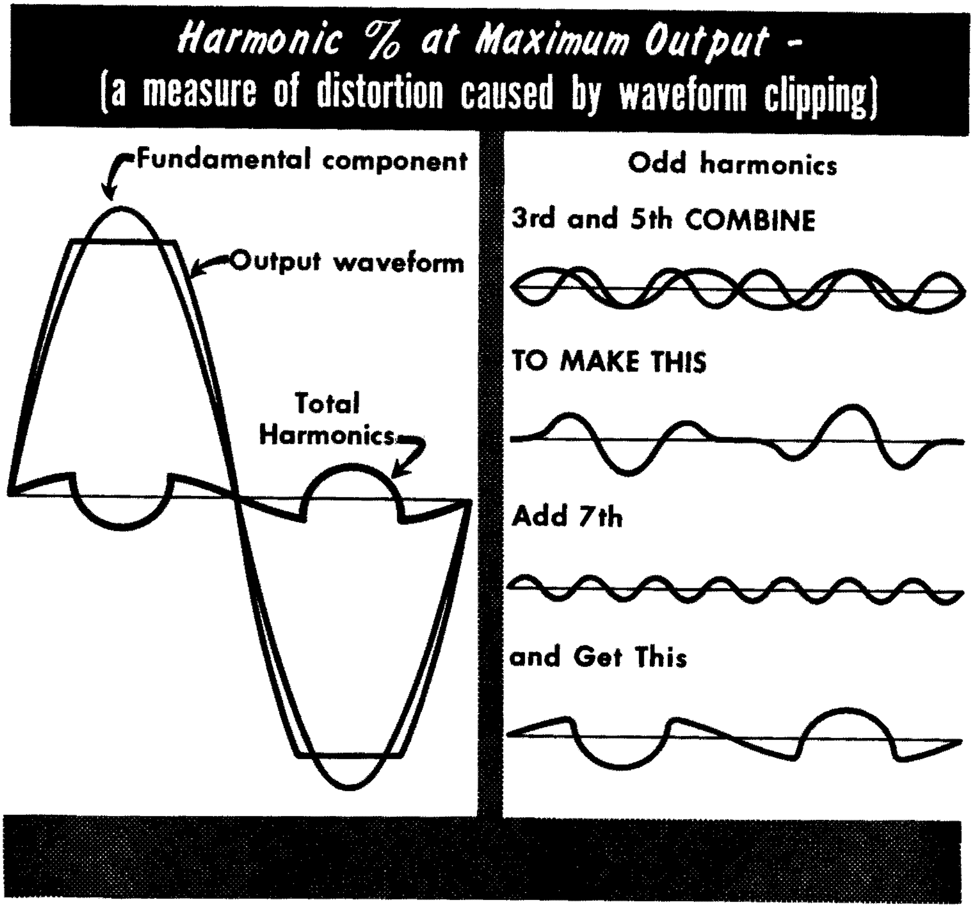


2nd harmonic	2nd harmonic squared	3rd harmonic	3rd harmonic squared	sum of the squares	total harmonic distortion $\sqrt{\text{Sum of the Squares}}$
0.8%	0.64	0.6%	0.36	1	1%
2.0%	4	1.5%	2.25	6.25	2.5%

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

Amplifier Specifications (contd.)

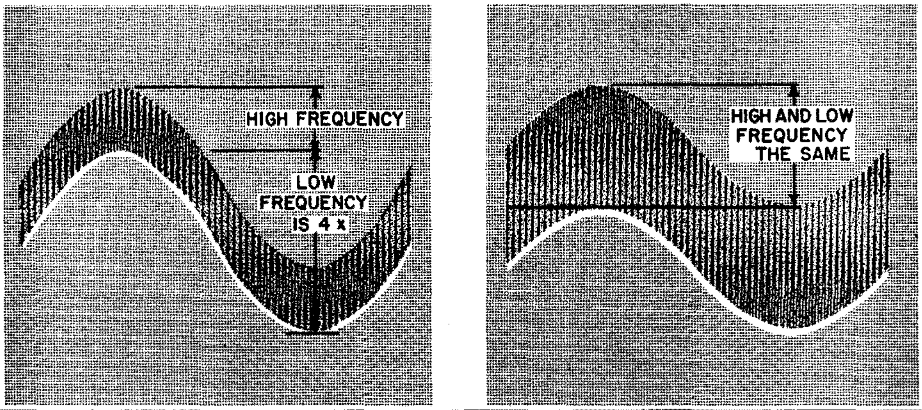
In modern amplifiers in which the harmonic percentage is specified at maximum output, the distortion usually takes the form of clipping on the tops of the waveform, due to the beginning of grid current.



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

Amplifier Specifications (contd.)

DISTORTION MAY BE SPECIFIED AS INTERMODULATION (IM)



Different possible combination of Amplitude 4:1 or 1:1

and

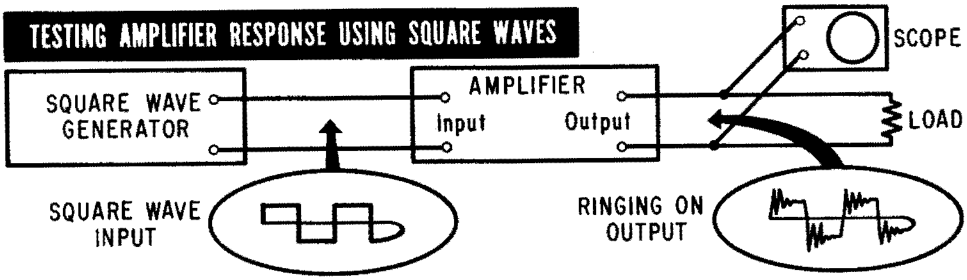
Different combinations of Frequency can be used

40 cps -- 3000 cps
100 cps -- 2000 cps
40 cps -- 2000 cps
60 cps -- 7000 cps, etc

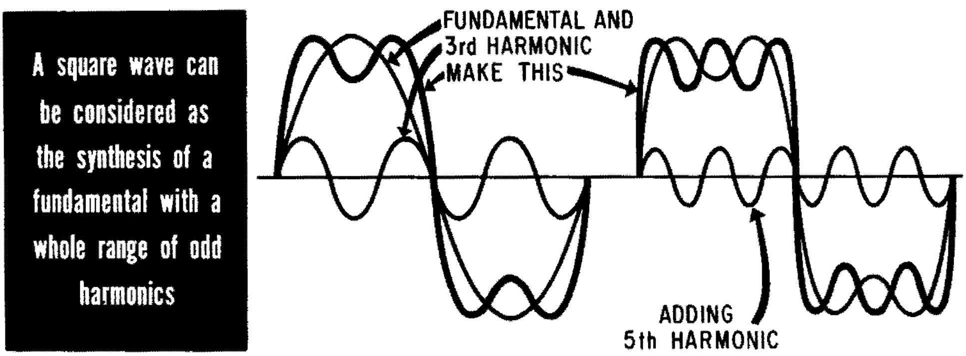
Sometimes the distortion is specified as IM (intermodulation). Unless further information is given about the frequencies used for the test, such a specification is valueless. If the first method of IM test is used, the low frequency (its actual value in cycles per second) is important, as well as the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

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

Transient Distortion



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



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

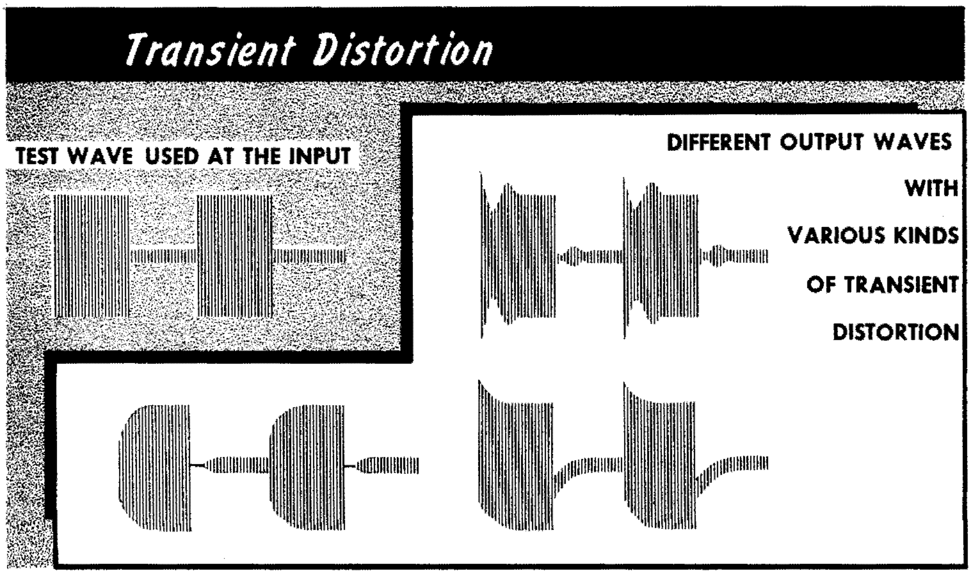
If the time delay to all components of the audio waveform is the *same*, the output waveform will still be a square wave, but if it is different for the higher frequency components than it is for the low-frequency fundamental and its lower harmonics, the waveform gets altered. This is one way in which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

Transient Distortion (contd.)

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

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

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



QUESTIONS AND PROBLEMS

1. Audio transformers are used to produce step-up of voltage or current, to transform impedances to different values, or to make maximum use of available power (as in a loudspeaker). Comment on the relationship among these functions.

2. What are the basic differences between primary inductance and leakage inductance in a transformer?

3. Why is an air gap used in a transformer core? What determines the optimum size of the gap?

4. Can a flat-topped waveform at low frequencies be due to saturation? Comment on the relationship between frequency response, and the limitation imposed by saturation.

5. Would distortion due to saturation be more apparent in an amplifier using triodes or in one using pentodes? Why?

6. What features must be considered in the design of a good audio transformer?

7. Explain how a transformer can help performance (a) at the input to an amplifier, (b) at its output, (c) between tube stages, (d) between transistor stages.

8. What is harmonic distortion? Show how you would identify the presence of different harmonics on the waveform.

9. Explain two methods of measuring harmonic distortion. What would be considered an acceptable figure?

10. What is intermodulation distortion? Describe two different forms it can take.

11. Why are there inconsistencies among the results obtained in different distortion measurements?

12. Write what you would consider to be a good amplifier specification, and comment on the limited validity of any figures you may quote.

13. If a harmonic analysis of a waveform showed 1.6% second, 1.5% third, and 1.2% fifth harmonics, what would you expect the total distortion to be (assuming no higher harmonics are present in measurable quantity)?

14. Comment on the significance of a harmonic measurement when the distortion mainly takes the form of clipping.

15. Comment on the use of square waves for testing (a) as to facts not brought to light that are revealed in other tests, and (b) as to forms of distortion that square-wave testing would not reveal.

16. Discuss some aspects of transient distortion that affect amplifier performance, which are not always included in the specification.

VOL. 2 INDEX

basic audio

by NORMAN H. CROWHURST

VOL. 3

JOHN F. RIDER PUBLISHER, INC.
116 West 14th Street • New York 11, N. Y.

Library of Congress Catalog Card Number 58-13425

Printed in the United States of America

PREFACE

Audio is like Topsy: it wasn't born, it just growed. Whatever Topsy may have been like, Audio has grown like a gawky child—not always in proportion! Originally associated with radio and later with high fidelity, audio now finds application in many other places—to name a few: computors, automation, ballistics and guidance for missiles, sonar detection for navigation, ultra- and infra-sonics for medicine, both diagnostic and therapeutic, as well as geophysical and other work. In fact Audio is now one of the largest and most basic divisions of electronics.

Courses in audio were nonexistent not too many years ago. Since then, textbooks and courses have appeared. But their approach follows the principle of many professors: "I learned it the hard way—you'll have to!" It's like learning watchmaking from a bridge-building man.

My wide experience in various aspects of audio has shown the need for a better way. In industry, in academic education, and particularly in working with graduates from college, this need is evident. My extensive technical writing for magazines and consultant work in the industry have also shown me audio's educational needs.

Many competent "practical men" find themselves hindered by lack of academic background in the subject. They can do their job in their own established "groove." But they do not have—and find it impossible to acquire—the background to enable them to expand outside this groove. These people need help in closing the gap between "theory" and "practice."

Engineers are conversant with the accepted "technical language," but they read the literature with only an "intuitive comprehension" (or should it be apprehension?). Their education dragged them past many "awkward spots" about which they have never felt really "comfortable." Like the King of Siam in "The King and I," they find many facts of which they wish they were more certain they are sure.

Very important are the new students, technicians, and audiophiles. They will need a basic education in audio to enable them to add their contribution to progress (and to earn themselves a living!). Why make it difficult? They'll do much better if they can get a good start.

All-in-all, it is time that certain roundabout approaches to this key subject were eliminated. We need a direct, meaningful way to take the place of the difficult detours. Then each of our three groups can not only "learn audio," but also understand it! This three-volume book results from the author's extensive education research. The finished arrangement achieves a completely new directness.

Let me give just one example: how many understand the behavior of a coupling capacitor, particularly its contribution to amplifier transient performance, and what sometimes happens to feedback? This has always been based on the concept of capacitive reactance, which does not adequately *explain* all the effects. We have adopted a practical "what happens?" approach.

As a result, someone who learned this the old way may miss the familiar landmark of the reactance concept—when he expects it; a closer examination will reveal the reason for postponing it: the whole presentation has been arranged to avoid the "dead spots" left by the traditional approach.

Inevitably such a change of approach will mean a change of stress. I make no apology for this. I know from practice that it is far more successful in getting Basic Audio "across."

It would be impossible to acknowledge the very many who have, knowingly or unknowingly, contributed to my experience, making this book possible. But I would like to express my thanks to the John F. Rider staff for their cooperation in "packaging" it in a form that interprets my intentions so well.

NORMAN H. CROWHURST

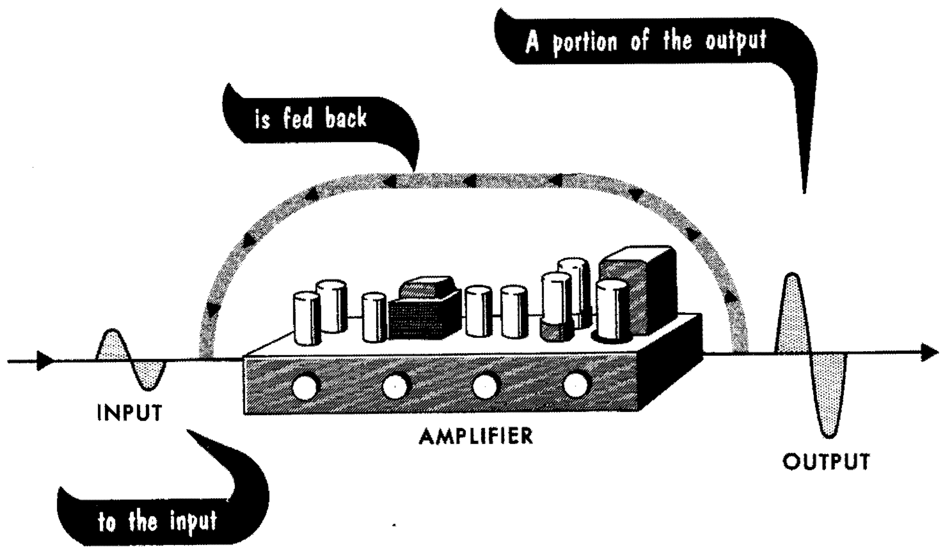
New York, N.Y.
August 1959

CONTENTS

VOL. III — BASIC AUDIO

Feedback and Distortion

FEEDBACK -- A MEANS OF REDUCING DISTORTION



Distortion in the fed-back signal acts to reduce distortion in the output

As the development of audio amplifiers produced better and better performance, the struggle to reduce distortion became steadily greater. A point of diminishing returns was reached, beyond which it did not seem possible to go. Then came the idea of *feedback.*

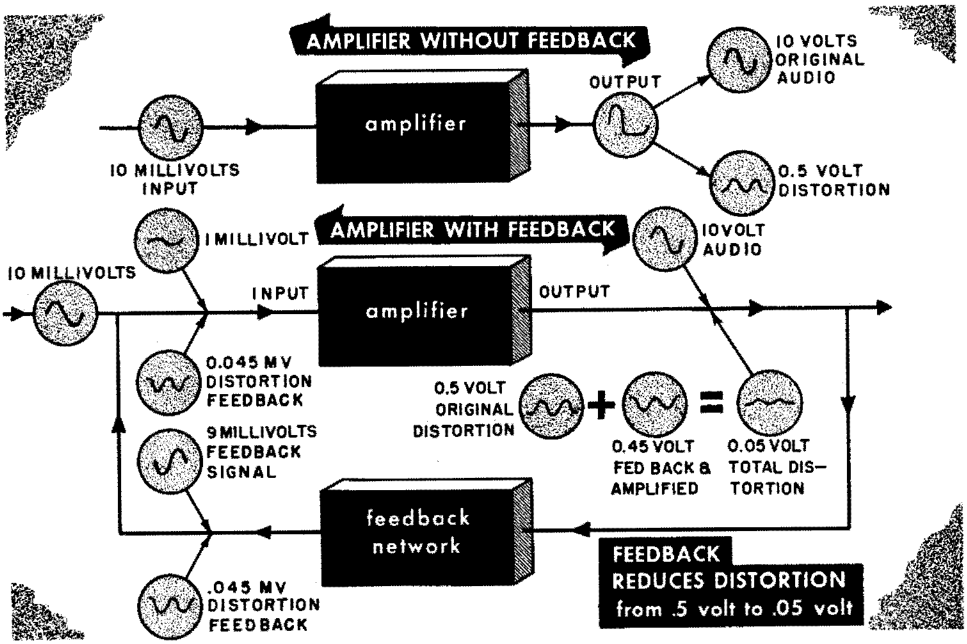
Tube development has made it easily possible to get more *gain* from an amplifier. What proved difficult was getting the output to be a more exact replica of the input. Feedback uses some of the additional gain, which can easily be obtained, to achieve this objective. A portion of the output is fed back or returned to the input. The difference between the input and output waveforms, which represents the distortion component, acts to reduce the amount of the distortion.

Feedback and Distortion (contd.)

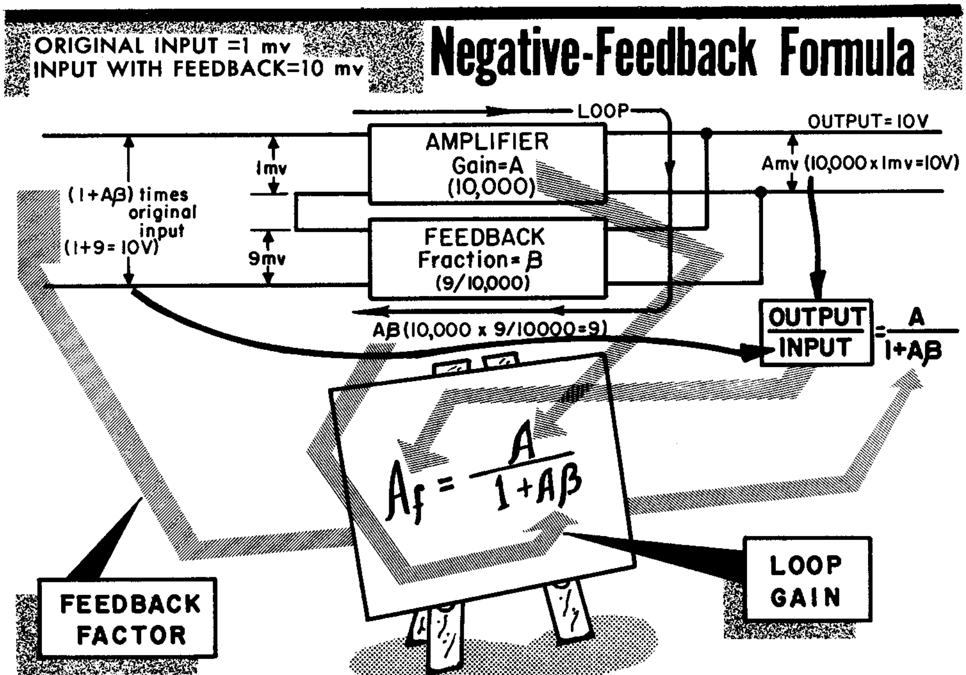
Suppose that an amplifier originally gives an output of 10 volts for an input of 10 millivolts and that the 10 volts contain 5% distortion. This will be 0.5 volt of some frequency (the distortion) that was not included in the 10-millivolt input.

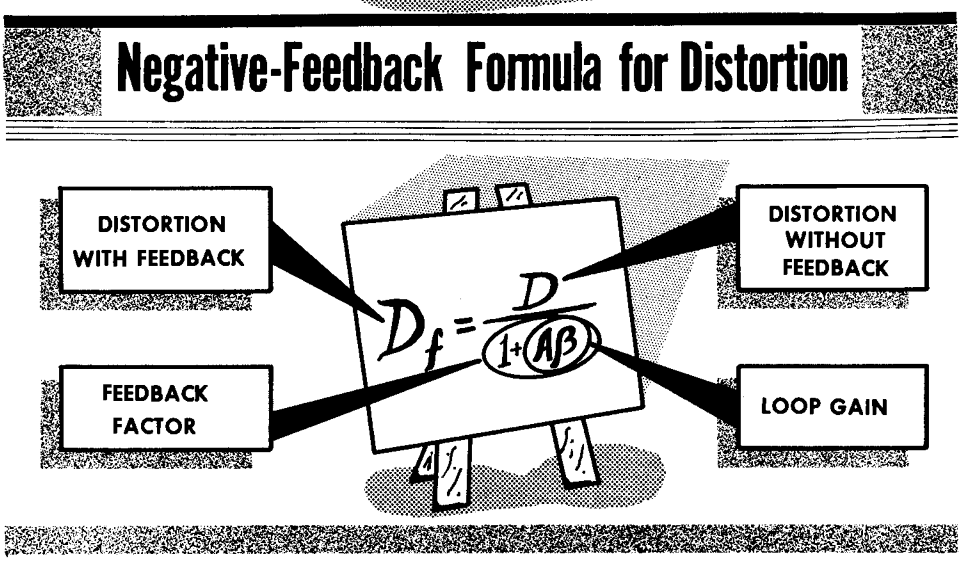
It is comparatively easy, by using an extra tube, to increase the gain of an amplifier so that an input of only 1 millivolt will produce the 10-volt output. Bearing this in mind, if we take a 9-millivolt sample of the output and *subtract* this from the original input of 10 millivolts, we shall have the required 1-millivolt input. This input consists of 10 millivolts original audio *minus* 9 millivolts fed back from the output, which is introduced into the signal by the amplifier. If there is 5%, or 0.5-volt distortion in the 10-volt output, the 9 millivolts fed back to the input will include 0.45-millivolt distortion.

The original audio *almost* cancels itself by the feedback — for the 1 millivolt actually fed into the amplifier, 9 millivolts are fed back to offset the 10 millivolts of audio fed into the complete arrangement. The distortion component, however, which originally was 0.5 volt in the output, has no original input to "offset," hence the whole distortion component fed back from the output will get amplified again and thus come out as 9/10 of its original size in the opposite direction. Thus, the ultimate amount of distortion left will be 1/10 of the original 5%, 0.5%, or .05 volt.



Negative Feedback

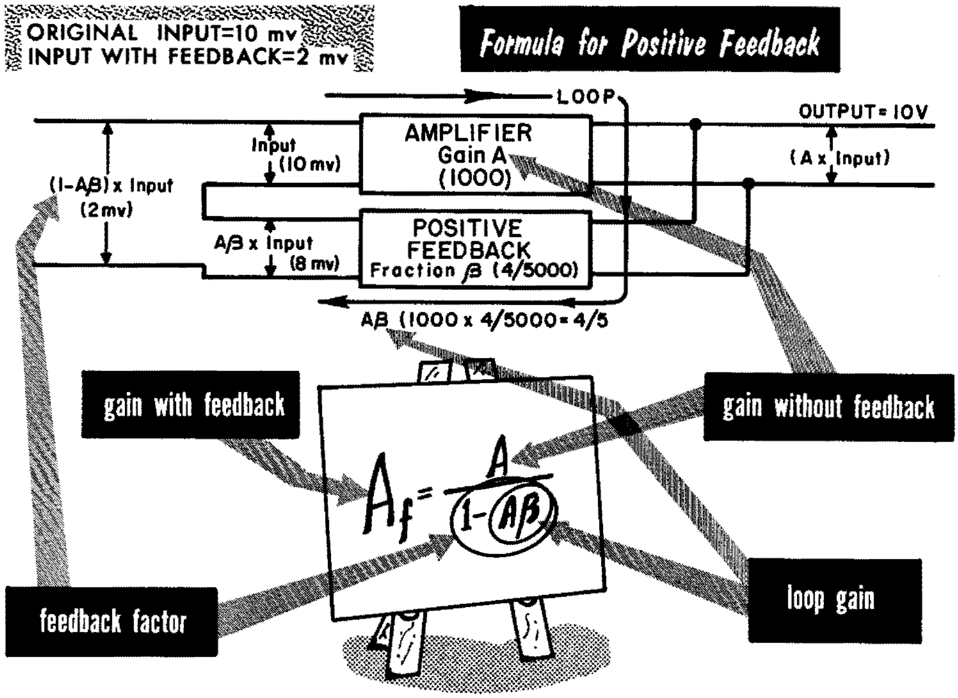




This is the formula for *negative* feedback, because the sample fed back from the output is in the *opposite* direction or phase from the input. In addition, feedback can also be *positive*.

Positive Feedback

Using the original amplifier as a starting point, with a gain of 1000 (a 10-millivolt input produces 10-volt output), we could take the sample from the output the other way round, so it *provided,* say, 4/5 of the required input. This would mean 8 millivolts would be taken from the 10-volt output and fed back to the input, hence the actual input only needs to provide the *remaining* 1/5, or 2 millivolts. The addition of positive feedback thus *increases* the gain of the amplifier, instead of reducing it as with negative feedback. When positive feedback is used, the formula for gain with feedback is rewritten as $A_f = A/(1 - A\beta)$.

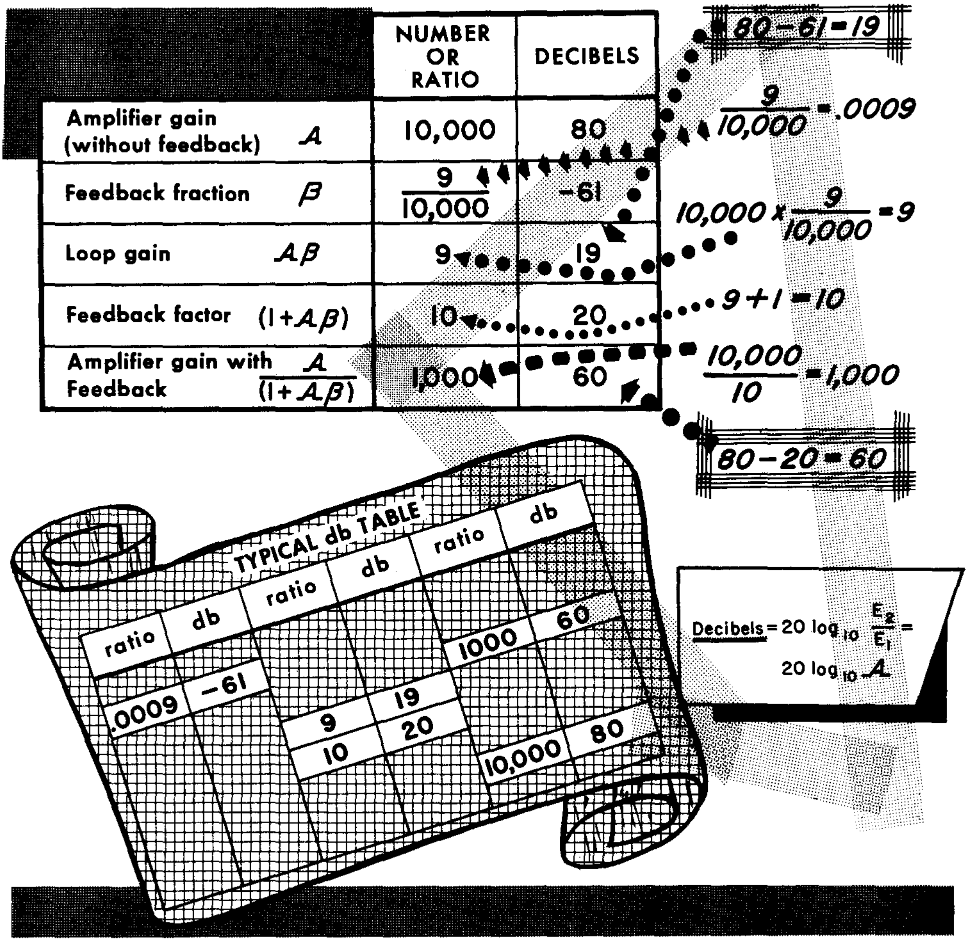


A represents the gain of the amplifier and β (a Greek letter called "beta"), represents the *fraction* fed back. The quantity A × β or Aβ is called the *loop gain.* It is the *net* gain of the combined arrangement, measured from the input to the amplifier, through to the output, and back through the feedback to the input again. The denominator of the fraction (1 + Aβ or 1 − Aβ) is called the *feedback factor,* or sometimes just *feedback.* It represents the amount by which the gain is divided by connecting the feedback.

For positive feedback, the feedback factor is (1 − Aβ). Thus, it is always a fraction. Hence, the gain will be *greater* than it was originally, because any number divided by a fraction becomes greater than the original number.

Decibel Calculations

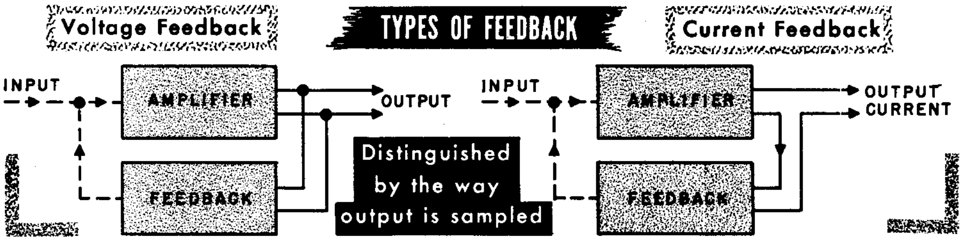
	NUMBER OR RATIO	DECIBELS
Amplifier gain (without feedback) A	10,000	80
Feedback fraction β	$\frac{9}{10,000}$	−61
Loop gain $A\beta$	9	19
Feedback factor $(1+A\beta)$	10	20
Amplifier gain with Feedback $\frac{A}{(1+A\beta)}$	1,000	60



Amplifier gain figures are often given in *decibels*, and we may encounter feedback referred to in decibels as well. It is, however, always more convenient to make the calculation in *ratios* and convert the *result* to decibels. This avoids confusion.

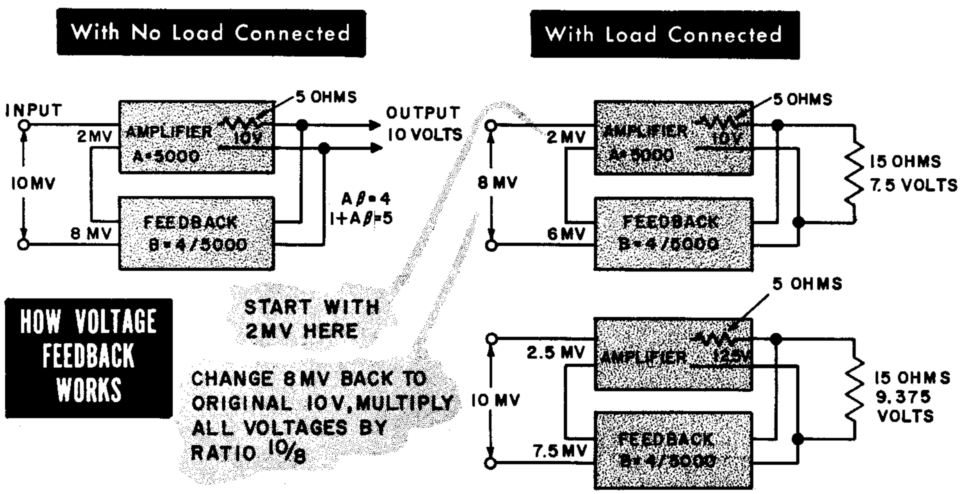
In the negative feedback example that we discussed, the gain of the amplifier without feedback was 10,000 or 80 db. The feedback fraction was 9/10,000, or approximately −61 db. Hence, the loop gain $A\beta$ is $80 - 61 = 19$ db. This corresponds to the loop gain ratio of 9. The feedback factor is 1 + 9 or 10, which is 20 db. So in this case, the loop gain is 19 db and the feedback is 20 db. For negative feedback, the bigger the loop gain, the nearer the two figures come to coinciding. If the loop gain were 99 (39.9 db), the feedback would be 100 (40 db).

Voltage Feedback



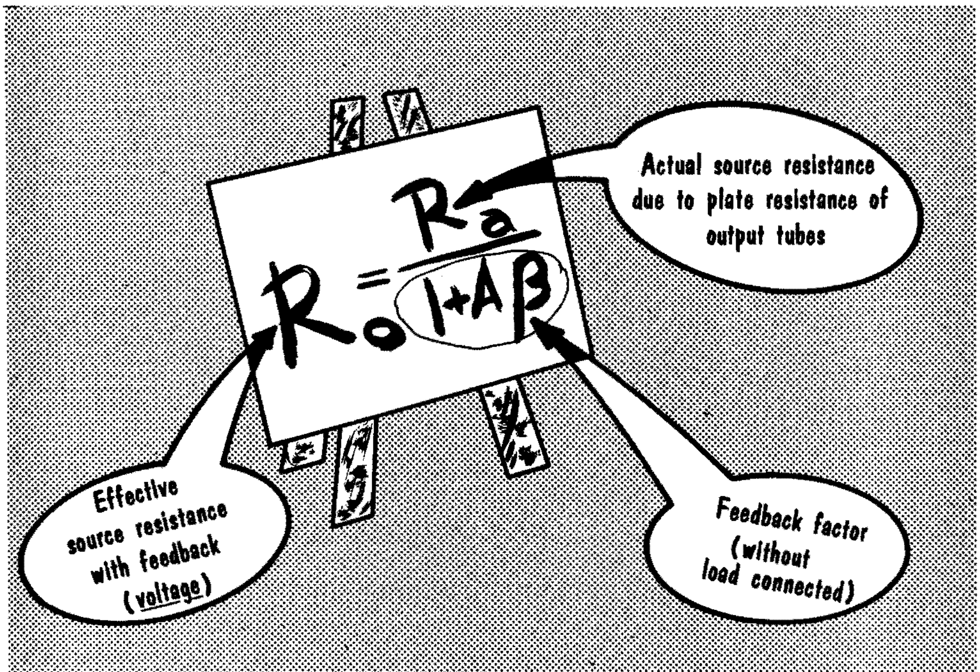
There are several ways in which feedback may be distinguished. For example, either the output *voltage* or the output *current* can be used as a basis for the output "sampling." The chief difference as far as the operation of the amplifier is concerned is in the effect on the amplifier's effective output impedance.

Let us work out a typical example, using voltage feedback. Suppose that an amplifier has to work with a 15-ohm resistance as its output load. It has an internal resistance due to the plate resistance of the output tubes of 5 ohms. Without the output load connected, a 10-millivolt input produces a 10-volt output. Without feedback, a 2-millivolt input would produce the 10-volt output. This means the gain of the amplifier, A, is 5000, or 74 db. The feedback, β, is 8 millivolts fed back for 10 volts at the output, 8/10,000 of 4/5000. The loop gain, Aβ, thus is 5000 × 4/5000 or 4. The feedback (1 + Aβ) is 1 + 4 or 5, which corresponds to 14 db. This means the gain and distortion will be divided by 5. Connecting the 15-ohm load causes the same 2 millivolts at the input of the actual amplifier to produce (10 volts × 15 ohms/20 ohms) 7.5 volts output.



Voltage Feedback (contd.)

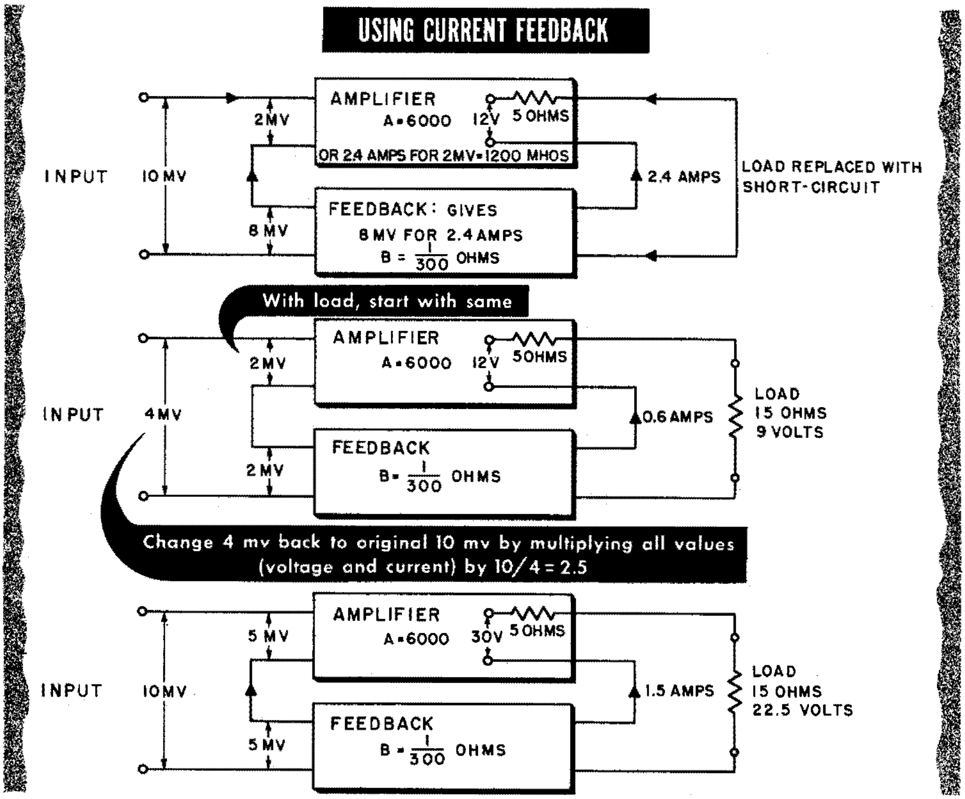
Because β is .0008, a 7.5-volt output will produce 6 millivolts of feedback. The total input required to give a 7.5-volt output with the load as well as the feedback connected is (2 + 6) or 8 millivolts. By simple proportion we can deduce the output with a 10-millivolt input. It will be ($7.5 \times 10/8$) or 9.375 volts. Thus the *effective* drop in the internal resistance (because feedback changes the *actual* input) is from 10 volts (without the load connected) to 9.375 volts (with the load connected), a difference of 0.625 volt. As it takes a load of 15 ohms to do this, the internal resistance must be $0.625/9.375 \times 15$ ohms, or 1 ohm. The connecting of feedback, which has reduced the gain by a ratio of 5, has also reduced the effective internal resistance of the output by the same ratio. Without feedback, the internal resistance was 5 ohms; with feedback it is 1 ohm.



Negative Voltage Reduces Effective Source Resistance

We could use the formula instead of the numbers and come out with the same answer — that the internal resistance with voltage feedback is divided by the same factor that the gain is reduced by. An important point to notice (not stated in many textbooks on the subject) is that we must use the gain *without the load connected* in this formula.

Current Feedback



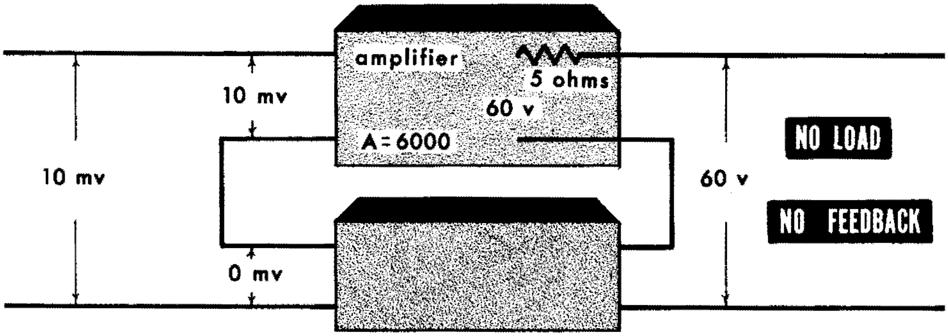
The other way in which the output can be sampled is called *current feedback.* Assume that we have an amplifier that without any load connected gives 12 volts output for an input of 2 millivolts. It has an internal resistance of 5 ohms. When the output is short-circuited, the internal resistance will be the only thing to limit the current—12/5 = 2.4 amperes. Assume that this 2.4-ampere output produces 8 millivolts of feedback and that the current feedback is always in this ratio.

This means the amplifier, complete with feedback, requires 10 millivolts input to produce a 2.4-ampere output on short-circuit. Now assume that the 15-ohm load is connected to the amplifier, still with 2 millivolts at the input. The original 12 volts at the output will produce a current of (12/20) or 0.6 ampere (which produces 9 volts across the 15-ohm load). Because the 2.4 amperes cause 8 millivolts of feedback, the 0.6 ampere will produce 8 × 0.6/2.4) or 2 millivolts feedback. Thus the input required for this condition will be a total of (2 + 2) or 4 millivolts. Assuming that we still apply the original 10 millivolts of input, the output current will then be (0.6 × 10/4) or 1.5 amperes.

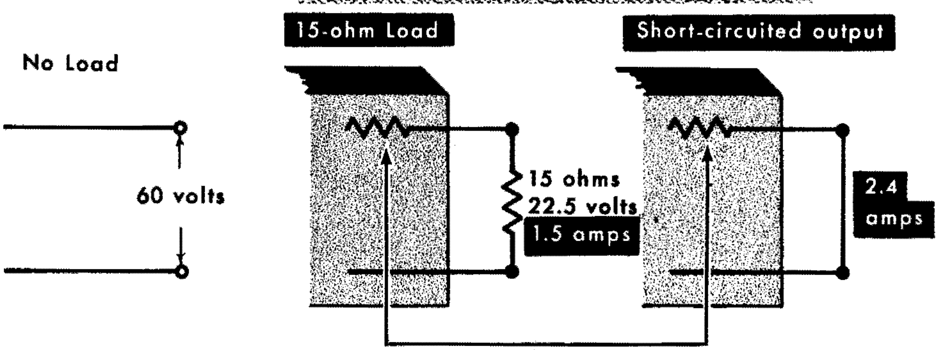
Current Feedback (contd.)

Now we have a basis for comparison. With the output short-circuited, the feedback amplifier with a 10-millivolt input gives a 2.4-ampere output. With the load of 15 ohms on the output, the same 10 millivolts input produces a 1.5-ampere output.

-- *No load results in no feedback* --



CURRENT FEEDBACK RAISES SOURCE RESISTANCE

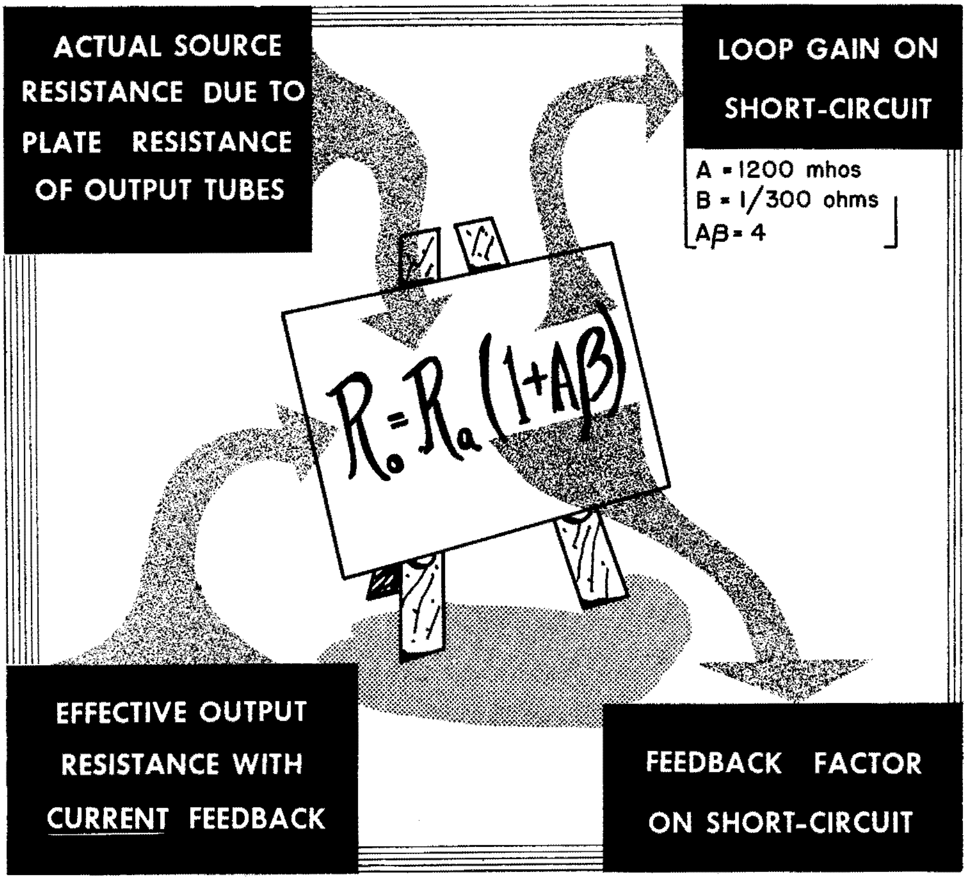


Equivalent to 25 ohms

We can take the matter one stage further and consider the output with no load at all. If there is no current, there will be no feedback and a 10-millivolt input will produce ($10/2 \times 12$) or a 60-volt output. With the current feedback applied and a 10-millivolt input, the effective open-circuit output voltage is 60 volts, whereas the short-circuited current is 2.4 amps. This means the *effective* internal resistance is 60/2.4 or 25 ohms.

Current Feedback (contd.)

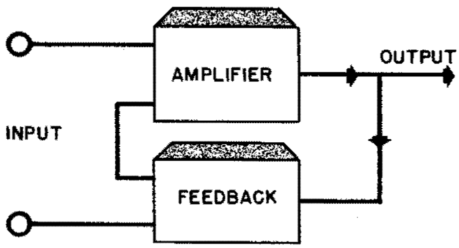
This checks with the loaded condition as well. When the 15-ohm load is connected, the total effective resistance is (25 + 15) or 40 ohms. This resistance connected across an effective 60-volt source will allow a current of 1.5 amperes, as calculated earlier. The current feedback has multiplied the effective source resistance by 5. (The actual value is 5 ohms and the effective value is 25 ohms.) The reduction in gain caused by the feedback is 5, *only when the amplifier operates short-circuited.*



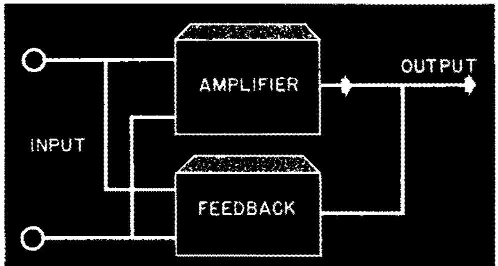
At normal loaded condition, the reduction in gain due to feedback is 2. (Without feedback, 2 millivolts produces the same output as 4 millivolts does with feedback.) Disconnecting the load entirely results in no feedback. Thus, for *current* feedback, the *effective* output source resistance is the actual source resistance multiplied by the feedback factor when the amplifier output is short-circuited.

Series Injection

Feedback Circuits can be classified according to the Method of Feedback Injection



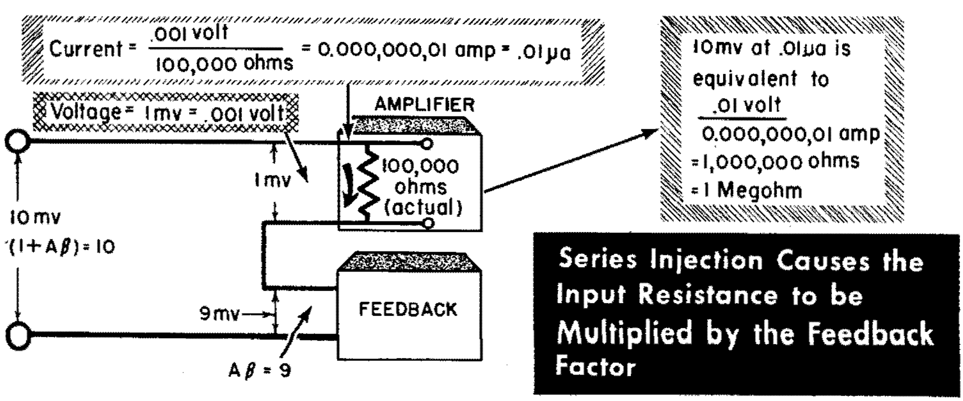
SERIES INJECTION



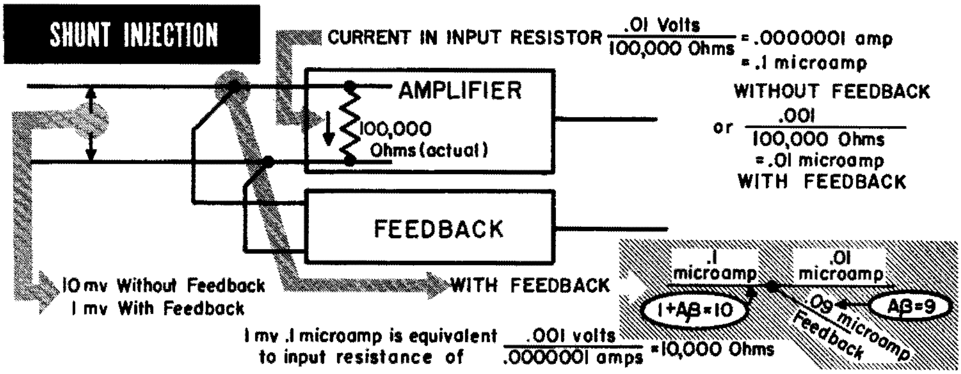
SHUNT INJECTION

Another way of distinguishing between different types of feedback is in the way in which the feedback is injected at the input end of the amplifier. It can be injected either in series or in parallel (shunt) with the input resistance of the amplifier itself. The effective *input* resistance of the amplifier is affected by the method of injection. When feedback is injected in series with the input to the amplifier (as we have assumed it to be thus far because this system is the simplest to follow), we assume that only voltages are being considered. There are always small currents as well.

When the 10-millivolt input is applied, the input resistance to the amplifier will draw a current, depending upon its value. If the input resistance were 100,000 ohms, 10 millivolts applied across it would produce a current of 0.1 microampere. Due to the feedback of 9 millivolts, however, the input resistance of 100,000 ohms has only 1 millivolt across it, which means it will draw only .01 microampere. Thus, from the viewpoint of the 10-millivolt total input, the resistance appears to be not 100,000 ohms, but 1,000,000 ohms, or 1 megohm. We may conclude that series injection causes the input resistance to be multiplied by the feedback factor — in this case, 10.

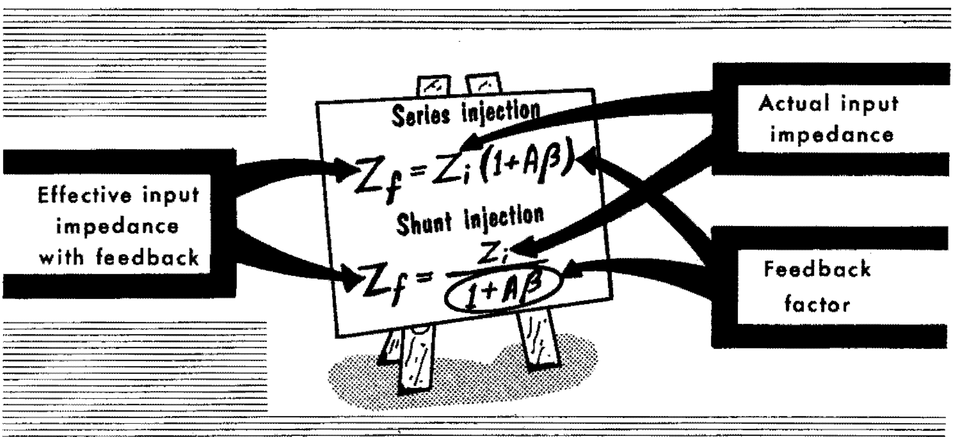


Shunt Injection



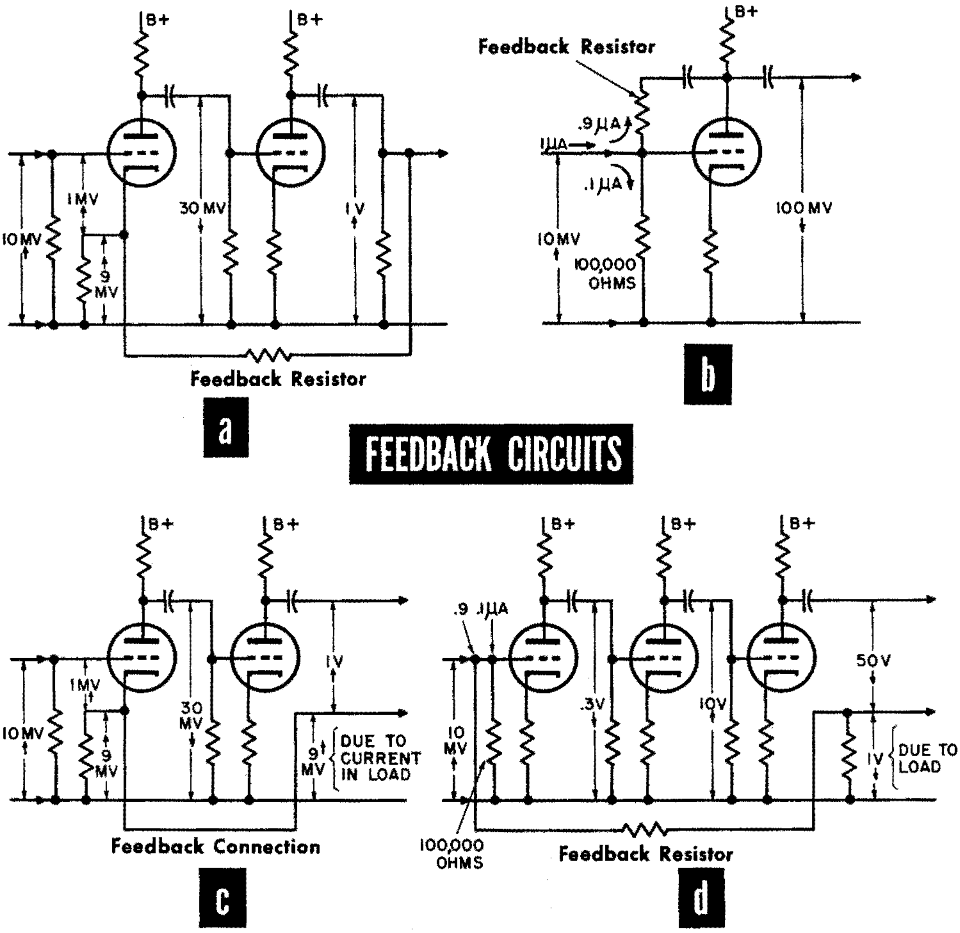
In shunt injection we must also consider the input current as well as the input voltage. Again assuming an input resistance of 100,000 ohms and an input current of 0.1 microampere, this would produce an input of 10 millivolts. However, the feedback current will provide 9/10 of the total input current in this case, or .09 microampere, leaving a current through the input resistance of only .01 microampere. This current would produce a voltage drop of only 1 millivolt.

Thus, although the input *current* is still 0.1 microampere, the input *voltage* is only 1 millivolt instead of the 10 millivolts it would be in the absence of feedback. The *effective* resistance now is 10,000 ohms instead of 100,000 ohms, and shunt injection has *divided* the input resistance by the feedback factor.



We can now analyze some practical feedback circuits to see how they classify under these different distinctions.

Feedback Amplifier Arrangements



CIRCUIT		a	b	c	d
Kind of feedback	Current or voltage	VOLTAGE	VOLTAGE	CURRENT	CURRENT
	Series or shunt	SERIES	SHUNT	SERIES	SHUNT
Effect on gain		$\div(1+A\beta)$			
Effect on distortion		$\div(1+A\beta)$			
Effect on output source impedance		$\div(1+A\beta)$	$\div(1+A\beta)$	$\times(1+A\beta)$	$\times(1+A\beta)$
Effect on input impedance		$\times(1+A\beta)$ CONNECTED GRID – CATHODE	$\div(1+A\beta)$	$\times(1+A\beta)$ CONNECTED GRID – CATHODE	$\div(1+A\beta)$

QUESTIONS AND PROBLEMS

1. What simple facts about amplifiers led to the use of negative feedback to reduce distortion? If feedback reduces the gain by a ratio of 5:1, by how much should distortion be reduced?

2. What is the basic difference between positive and negative feedback? Show, with a sketch, the difference between loop gain and feedback factor.

3. What is the maximum loop gain that can be used with positive feedback without causing oscillation?

4. Explain the effect of (a) voltage feedback and (b) current feedback on amplifier output impedance. State clearly any condition that should be specified concerning the amount of feedback in figuring this effect.

5. Explain the effect of feedback, using (a) series injection and (b) shunt injection on input impedance in an amplifier.

6. The loop gain in an amplifier using negative feedback in 12 db. What is the feedback factor, in db?

7. Without feedback, an amplifier has an output source resistance that is 3 times its nominal load resistance. What will be the value when voltage feedback is used that reduces gain by 26 db without the load connected? Or with current feedback that reduces gain by 6 db with the output short-circuited?

8. With the output load connected, a pentode output amplifier shows a reduction in gain of 20 db when voltage feedback is connected. What is the approximate source resistance (as a multiple or fraction of load impedance) with feedback connected? How could you evaluate the source resistance without feedback connected?

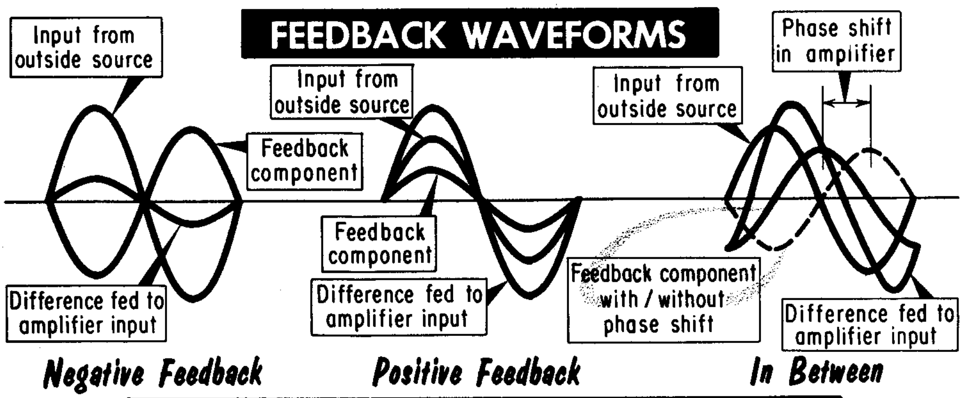
9. Without feedback, an amplifier shows 8% distortion unloaded, and 5% loaded, at maximum rated output voltage. Changing the output loading requires an input change of 8 db to maintain the voltage. Feedback is connected that reduces gain by 20 db in the loaded condition. Calculate: (a) the source resistance (as multiple or fraction of nominal load) without feedback; (b) the feedback, using the same circuit unloaded; (c) the source resistance with feedback; (d) distortion with feedback, unloaded; (e) distortion with feedback, loaded.

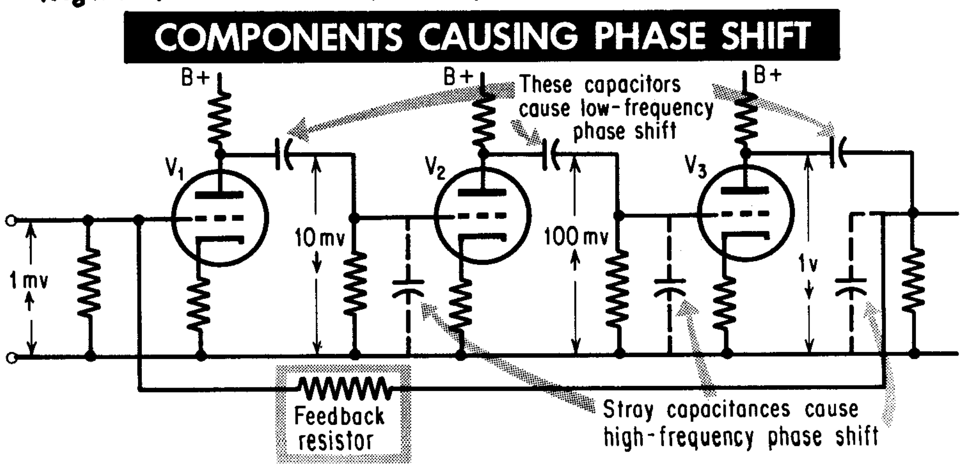
10. An amplifier uses overall feedback, voltage derived at the output, and series injected at the input. Without the load connected, removal of the feedback connection increases gain by 46 db. With it connected, the reduction is 26 db. An input resistor of 100,000 ohms is across the basic amplifier input (inside the feedback loop). Calculate: (a) the source resistance without feedback; (b) the source resistance with feedback; (c) the input impedance with feedback, the output being unloaded; and (d) the input impedance with feedback, the output being loaded.

Phase Shift Due to Feedback

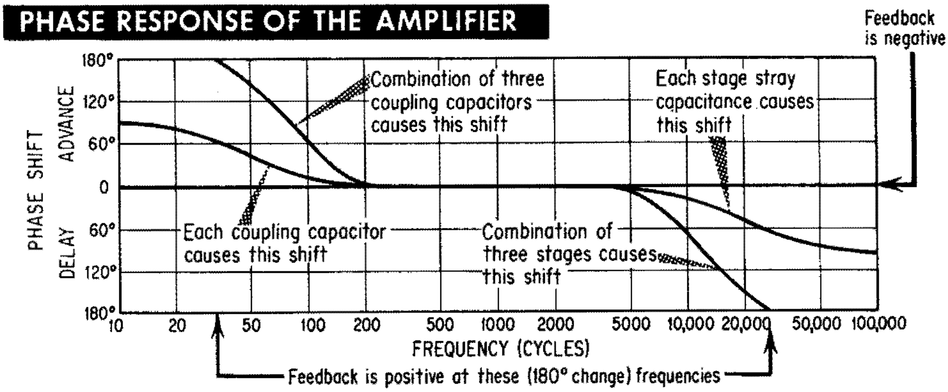
Thus far we have talked about feedback as being either *positive* or *negative,* as if the fed-back voltage (or current) *either* adds to *or* subtracts from the original input. This has helped us to lay down the simple rules of feedback. Practical circuits, however, do not behave quite as simply as this. At some frequencies, the fed-back audio is neither precisely in phase nor precisely out of phase with the original input, but somewhere in between. The resulting voltage fed to the amplifier input is what is called the *vector difference* between the input and the fed-back audio. We can see by examining the waveforms that their relationship is very much like that between voltage and current in a series circuit containing resistance and capacitive reactance.

All amplifiers possess some reactance, such as coupling capacitors, which affect the low-frequency response, and stray capacitances, which affect the high-frequency response. The output voltage gets out of phase with the ideal arrangement of negative or positive feedback, as the case may be, at these frequencies.





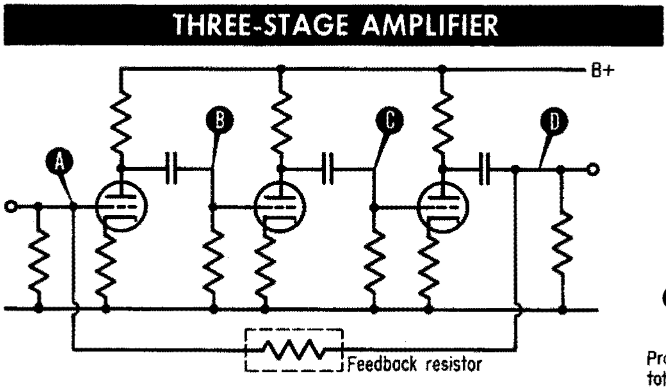
Phase Shift Due to Feedback (contd.)

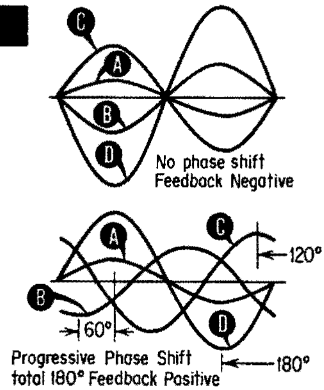


Take the effect of the coupling capacitors on the low-frequency response. If each of these capacitors produces a phase difference of 60° at the same frequency, the total phase shift adds up to 180°. This amounts to phase reversal. If the feedback starts out being negative at this particular frequency, it will convert into positive feedback.

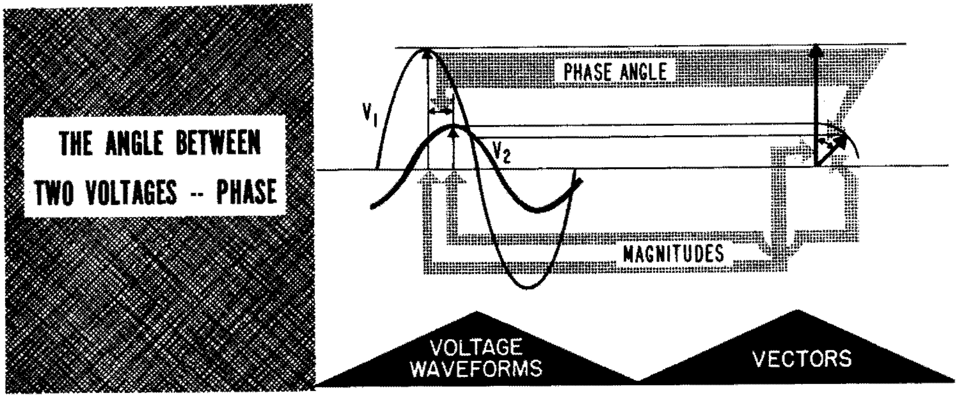
The dangerous thing now is that, with positive feedback, the amplifier will oscillate because the output voltage will equal the input voltage; consequently, no *external* input voltage is needed — the amplifier will continue to amplify its own output at this particular frequency.

This critical point is a loop gain $A\beta = 1$. In the formula for positive feedback, the feedback factor (by which the gain is divided) is $(1 - A\beta)$. As we said before, this is always a *fraction,* so the gain is not reduced in fact, but increased. However, when $A\beta$ becomes equal to 1, then $1 - A\beta = 0$. *Anything* divided by zero is infinity. It means that the gain of the amplifier becomes infinite at this frequency, and will go on amplifying its own output indefinitely, without the need for any input.

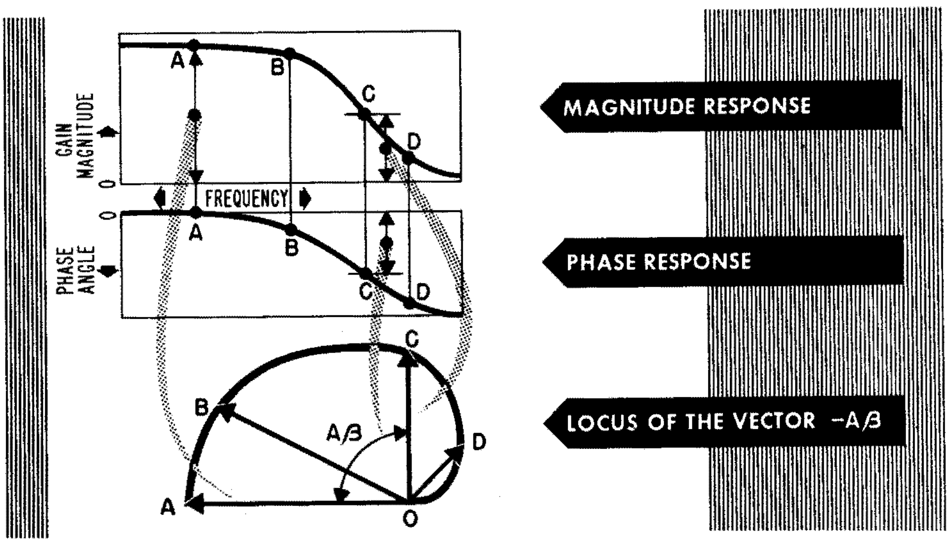




The Loop Gain — Nyquist Diagram

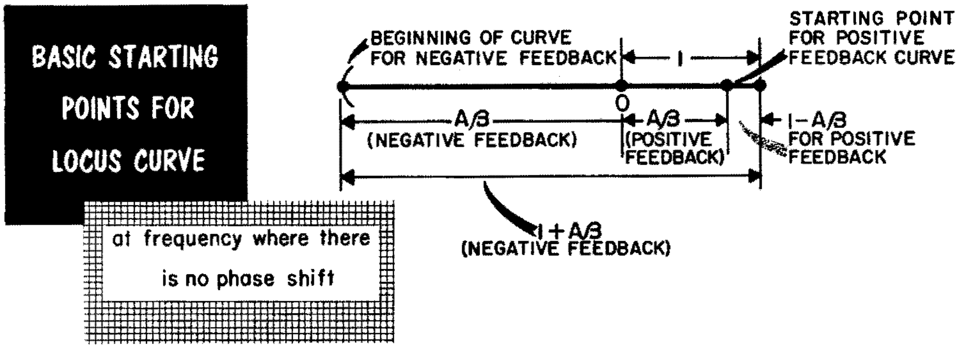


We are fortunate that this time difference between the voltage fluctuation at different points in an amplifier at various frequencies can be designated as a *phase angle*. This means that we can draw a "picture" of the voltage with what is called a *vector*. It is a line drawn at an angle to correspond with the phase angle from a starting point representing zero.



The loop gain Aβ is the important thing to consider. We can draw a series of lines from a starting point called O (for Origin). Each line represents both the magnitude and the phase angle of Aβ for a different frequency. Joining the outer ends of these lines produces a curve that represents all possible positions for the tip of the Aβ vector and shows how Aβ varies in phase and magnitude due to frequency changes. This curve is called a Nyquist diagram.

Nyquist Diagram (contd.)

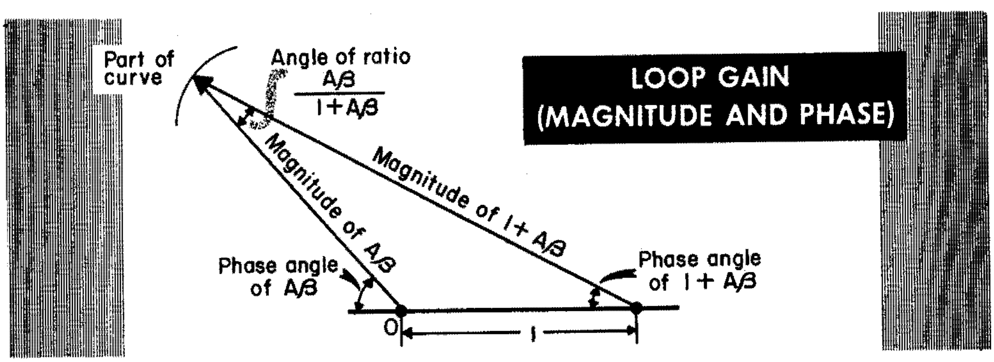


In the case of negative feedback, the feedback factor is $(1 + A\beta)$. First we draw a line in one direction from O, representing $A\beta$. Assuming that there is no phase shift, we can mark off 1 in the opposite direction from O and the total length of the line is then the feedback factor $(1 + A\beta)$. In the case of positive feedback, both $A\beta$ and the distance will be measured in the same direction from O and the disance between the end of the line representing $A\beta$ and 1 will represent the feedback factor $(1 - A\beta)$.

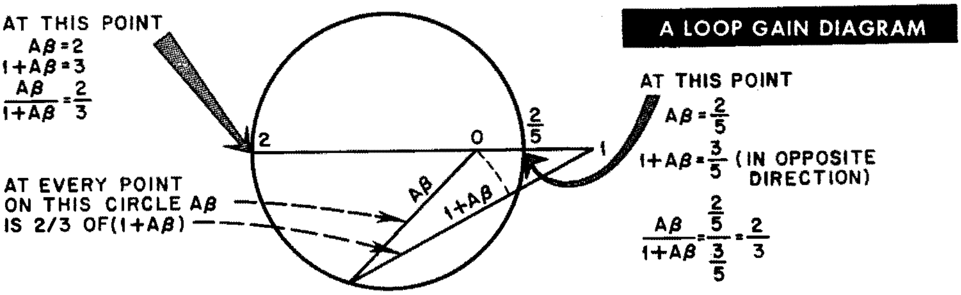
Either way, the feedback factor is given by the distance between the $A\beta$ locus and the point measured off 1 unit from O. The convenience of this method is that the distance between any other point on the locus and this position, measured 1 unit from O, also represents both the magnitude and phase angle of the feedback factor.

Thus the feedback factor can be called $(1 + A\beta)$; bear in mind that $A\beta$ is not just a simple number now, but that it includes a phase angle. When this phase reaches 180°, $A\beta$ has become negative instead of positive, and represents fully positive feedback.

A rigorous proof of this would involve mathematics beyond the scope of this book, and an exact explanation would call for a knowledge of complex numbers. Looking at it as a simple geometrical diagram, however, can give us a good picture of what happens without knowing all the mathematics.



Nyquist Diagram (contd.)

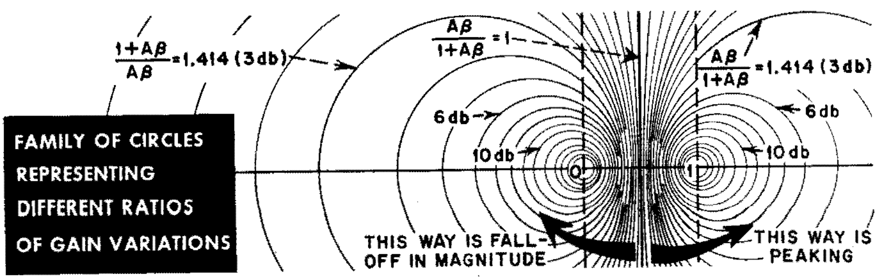


Now let's look a little more at the geometry relative to this diagram. The formula for the gain of an amplifier with feedback is $A_f = A/(1 + A\beta)$.

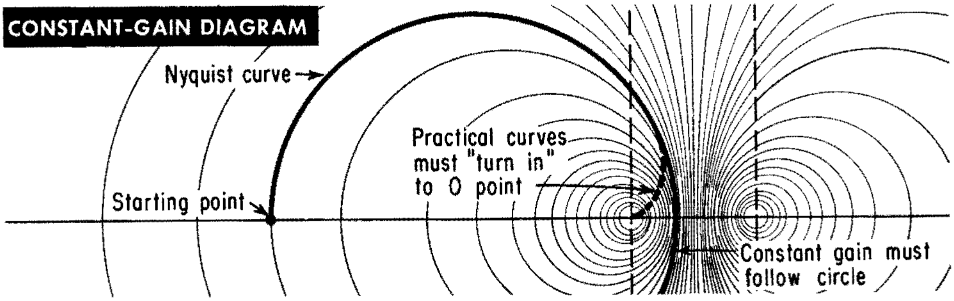
In most amplifiers, the feedback fraction β is constant; it does not change either in magnitude or phase as we change frequency. The internal amplifier gain A is the part that changes with frequency and produces the phase shift. We could multiply the top and bottom of the fraction giving the amplification with feedback by β and still have the same results. We obtain a part of the formula for amplification with feedback that does not change with frequency $1/\beta$, and a part that does change with frequency $A\beta/(1 + A\beta)$.

In the locus diagram, we have a curve representing the locus of a point whose distance and angle from the point O represent the value of $A\beta$ in magnitude and phase. Also, the distance and angle from the point measured off 1 unit from O represents the magnitude and phase of $(1 + A\beta)$. So the *ratio* of the distances of any point on the curve from these two points gives us the *frequency-varying part of the formula for amplification with feedback.*

It is a fact in geometry that all the points whose distances from two fixed points are in a fixed ratio form a circle. If we draw a family of circles representing different ratios of gain variation $A\beta/(1 + A\beta)$, we have a background that will help us interpret this locus diagram curve. In the background circles shown here, the vertical straight line joins all points where $A\beta = 1 + A\beta$. From there, curves are drawn at 0.5-db differences in ratio up to 3 db either way $[A\beta = 1.414 \times (1 + A\beta)$ or $1 + A\beta = (1.414 \times A\beta)]$. From there to 10 db, the circles are at 1-db intervals, and from 10 to 20 db, at 2-db.



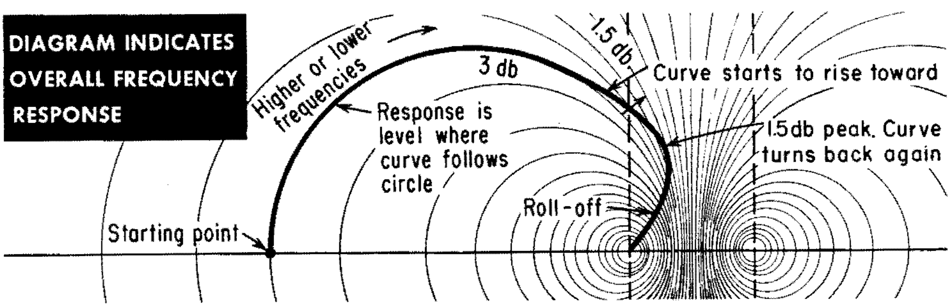
Nyquist Diagram (contd.)



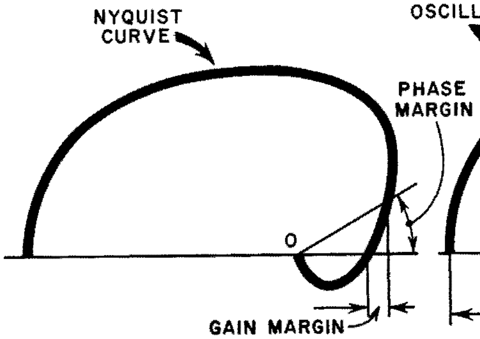
The locus vector curve itself is called a *Nyquist diagram.* If the curve representing the locus of Aβ follows one of the circles which represents a constant ratio of distances from O and 1, the gain of the amplifier with feedback would be *constant,* although there would be *phase change* in both Aβ and (1 + Aβ), as well as a transition from positive to negative value of Aβ.

This particular response is impossible with any practical amplifier. At some frequency, A *must* fall to zero. Usually this happens at a very low frequency *and* at a very high frequency, due to ultimate loss in the coupling capacitors at the low-frequency end and stray capacitance between stages at the high-frequency end. Either way, when Aβ falls right down toward zero, the curve must turn in and go to the point O. None of the circles representing constant ratio goes through the point O; all pass between the points O and 1 and out beyond them on opposite sides, one side or the other, except for the line that represents Aβ = 1 + Aβ which is a straight line perpendicular between the two points. As a practical amplifier has frequency limitations and eventually loses gain completely at extreme frequencies, it cannot follow *any* of these constant-gain lines (circles) *all the way.*

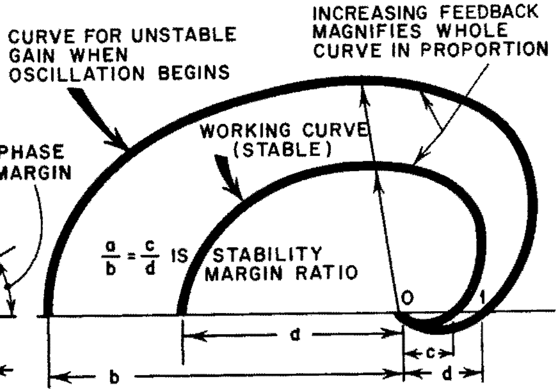
We *can* use this diagram, however, to predict the overall response of the amplifier (with feedback) by the way the locus curve criss-crosses the circles drawn to represent different values of constant ratio. A complete Nyquist diagram starts from O and finishes at O, representing frequencies zero and infinity. For simplicity, we have shown half, starting from a mid-range frequency where there is no phase shift either way.



Nyquist Diagram (contd.)



GAIN AND PHASE MARGINS



MEASURING GAIN MARGIN

Now we can see how this same diagram shows how much margin of safety we have between the way we are working and the beginning of oscillation. Increasing the value of Aβ at all frequencies uniformly multiplies up the whole size of the curve in proportion. If the curve goes round more than 180° before it turns into the source or origin point O, increasing its size eventually causes it to go through the 1 point, which means oscillation occurs. We can see how much margin there is between the point where the curve passes through the positive *direction* and the 1 *point*. As the whole curve multiplies up in proportion as the gain or feedback of the amplifier is changed, the *ratio* between this distance and 1 gives the amount by which the loop gain Aβ can be increased before oscillation commences.

This ratio, expressed in db, is called the *gain margin,* because an increase by this much gain starts oscillation. This margin can actually be measured by increasing the amount of feedback until oscillation commences and then calculating the difference in loop gain between the condition at which the amplifier actually works and the amount needed to make it oscillate.

The other *criterion of stability,* as these margins are called, is the *phase margin.* This can be shown on the Nyquist diagram, but it is not at all easy to measure. Hence, the gain margin is probably of the most value in assessing the performance of an amplifier. On the diagram, the phase margin is the angle of the vector Aβ at the point where the curve passes through a radius of 1 from point O, and point 1 from which distances (1 + Aβ) are measured. It means that increasing the loop phase shift in the amplifier (or anywhere in A or β) by this angle would cause the amplifier to oscillate.

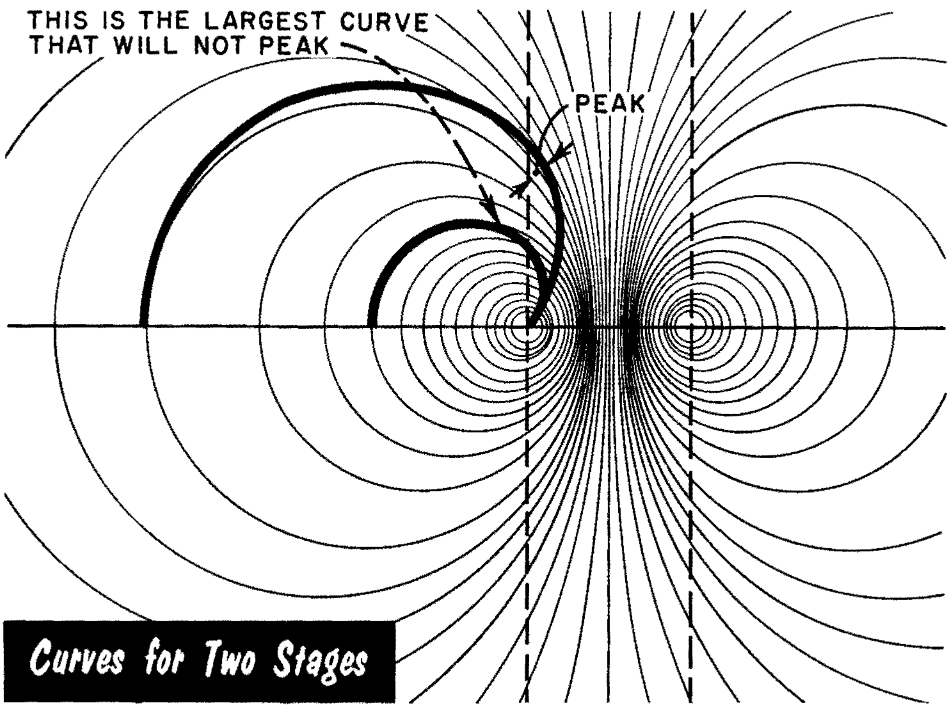
This statement assumes a change of the phase shift without a change of the amplification characteristic in any other way, which, in practice, is not possible. This is another reason why the phase margin is not a very practical criterion: it has no real significance, but now we have the tools.

How To Get The Right Answer

We like to use a lot of feedback but this results in conditions, if we are not careful, that cause oscillation, because Aβ will be so big at the starting point that it will be difficult to have it turn in sharply enough to get "inside" the 1 point. We can look at this on the graphs of db response and phase shift caused by each coupling network. We notice that, at the 3-db point, the phase shift is just 45°. The phase shift never quite reaches 90° but the db response keeps on falling off at almost 6 db for every octave in frequency.

If we add a second arrangement with the 3-db point at the same frequency, we would double these values. There would be a 6-db loss of magnitude and a 90° phase shift. The phase shift would never quite reach 180° but the db response keeps on falling off at almost 12 db for every octave. Although this two-stage arrangement can never get as far as causing oscillation, it can cause a peak in the frequency response.

This can be shown easily by drawing this curve on top of the background of circles representing constant ratios. Changing the size of the two-stage curve alters it from a condition where it follows one of the circles and then turns in to the "back" of the point O to one where it moves outward over the circles before cutting back across them. When it gets as big as this, the response indicated is one with a peak.

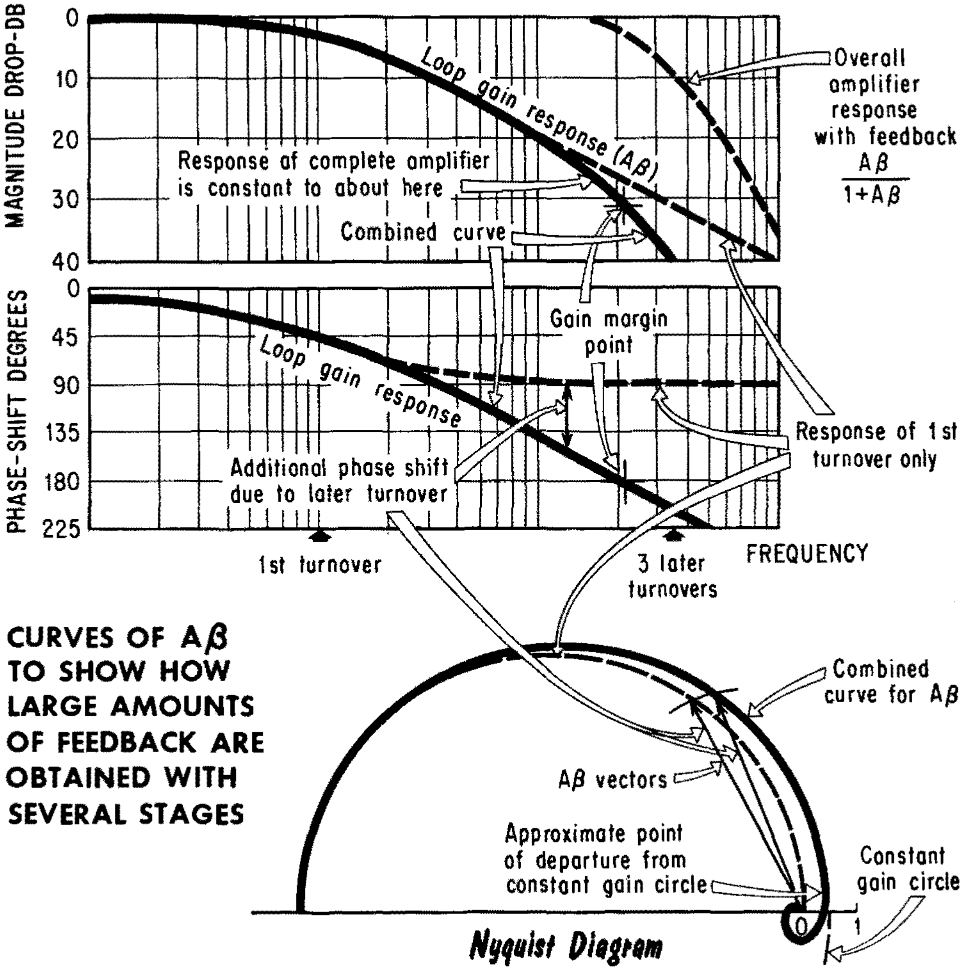


Curves for Two Stages

How To Get The Right Answer (contd.)

If more than two stages are used, the ultimate phase shift approaches 270°, so it must pass through 180° somewhere. The problem now is to make sure that the amplification has reduced, so Aβ is less than 1, by the time the phase shift reaches 180°.

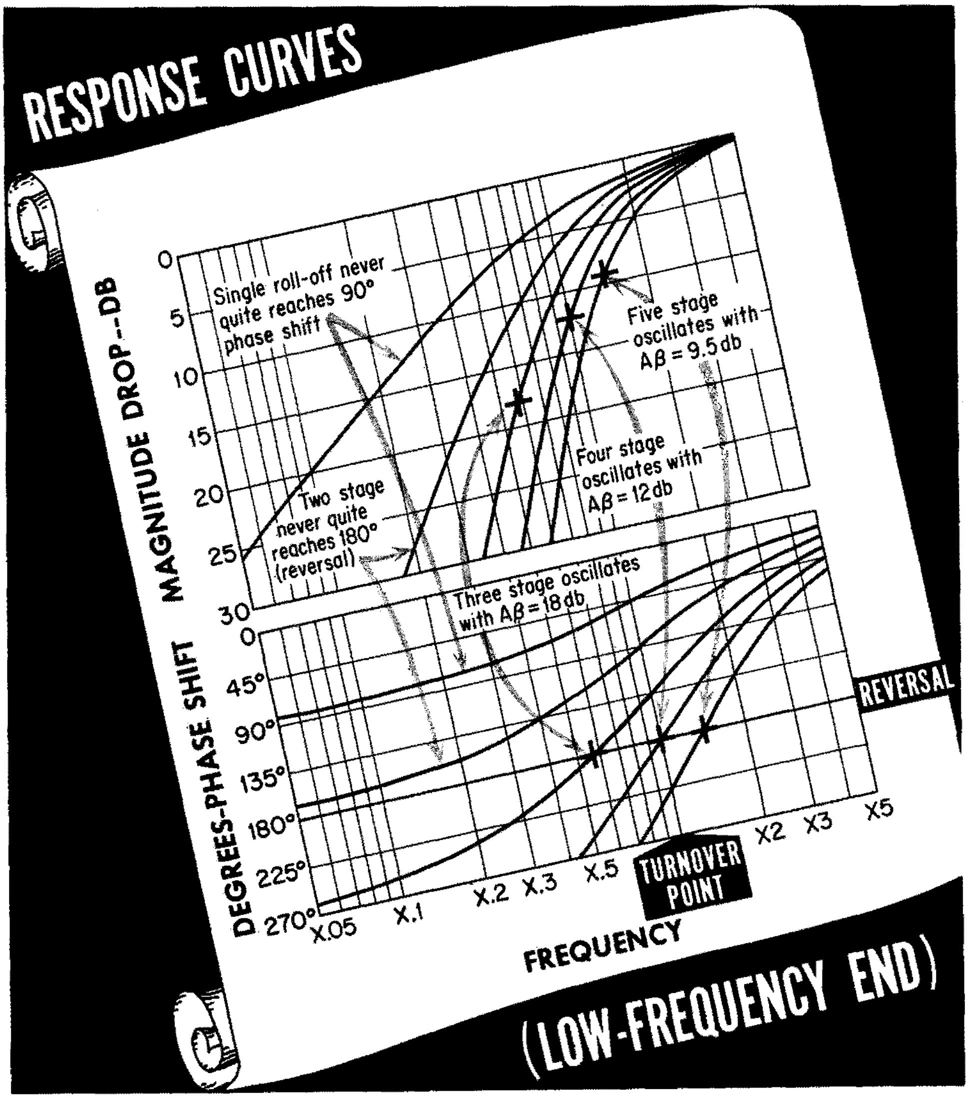
RESPONSE CURVES (HIGH-FREQUENCY END)



CURVES OF Aβ TO SHOW HOW LARGE AMOUNTS OF FEEDBACK ARE OBTAINED WITH SEVERAL STAGES

The best way of achieving this proves to be a choice in the combination of coupling capacitors values (for the low-frequency response) that causes one rolloff or 3-db point to occur at a much higher frequency than the other two (or more).

How To Get The Right Answer (contd.)

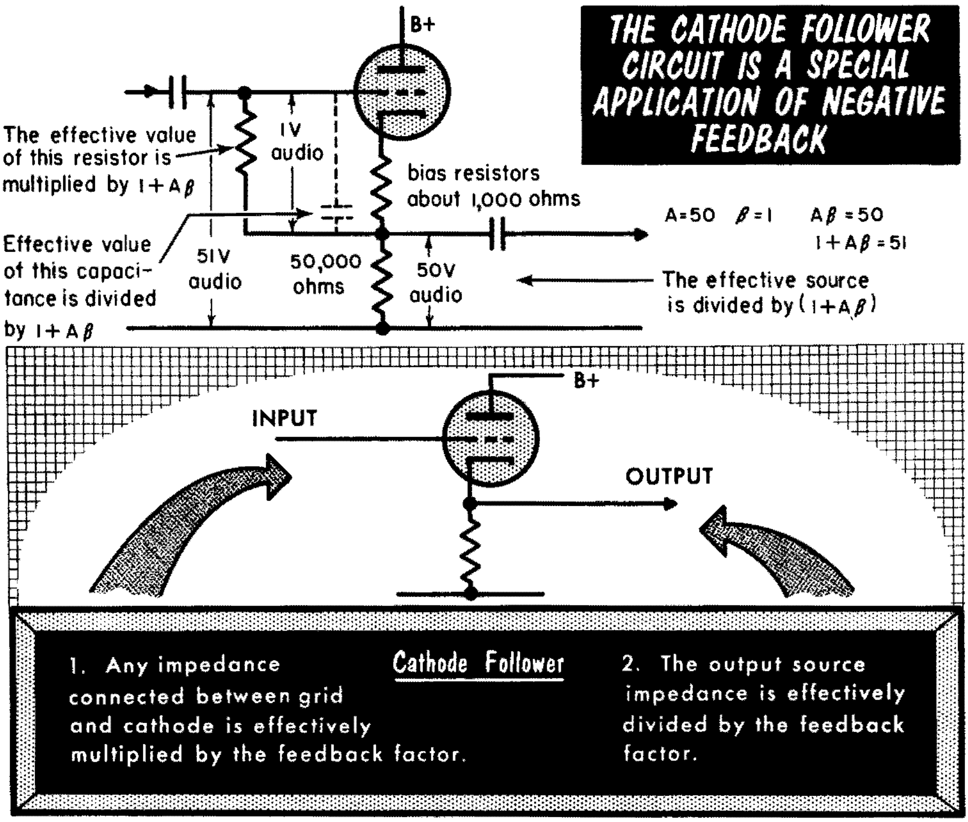


In this way the gain is reduced at the rate of 6 db for every octave with only a little more than 90° phase shift (because the other rolloffs are gradually starting) so it gets the magnitude of Aβ down to much less than 1 before the phase shift reaches 180°.

If similar rolloff points are used at each stage, a three-stage amplifier will start to oscillate when Aβ reaches 18 db (or a loop gain of 8); four stages will only reach a loop gain of 4 (12 db); and five stages limit it to 3 (9.5 db).

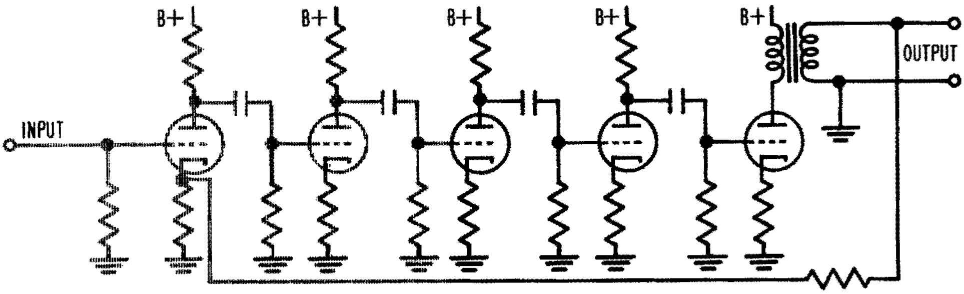
The Cathode Follower

The cathode follower may be regarded as a special example of feedback — all of the output voltage appears in the input circuit, hence $\beta = 1$. If the tube has a working gain of 50 with the resistor values used, an input fluctuation of 51 volts will give an output fluctuation of 50 volts. Following the feedback principle, the effective value in the input circuit of any resistor connected between grid and cathode will be multiplied by 51. This also applies to the reactance of any capacitance between grid and cathode of the tube. The reactance of a capacitance is inversely proportional to its capacitance value, hence the feedback divides the effective capacitance by 51.



This characteristic makes the cathode follower a very convenient tool for reducing the output impedance of the amplifier. The output impedance in the absence of feedback would be the plate resistance in parallel with the plate load resistor, in this case, about 50,000 ohms. However this gets divided by the feedback factor (in this case 51) to give an effective resistance of less than 1000 ohms.

Positive With Negative Feedback

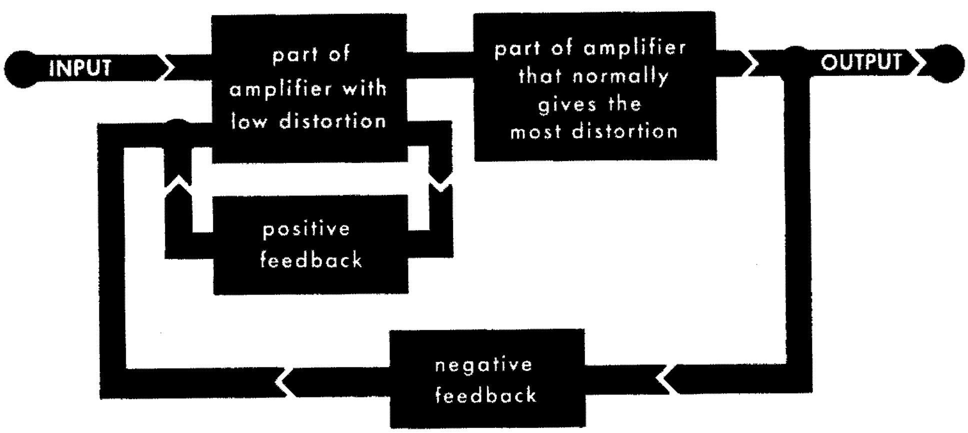


Adding Extra Stages gives More Gain, so more Feedback could be used, except for one thing -- EACH EXTRA STAGE ALSO INCREASES THE PHASE SHIFT PROBLEM

As amplifiers use more and more feedback to get the distortion down to even lower proportions, the problem arises of getting enough amplification to "throw away." The more gain required, the more stages that have to be added, and the more possibility of phase shift we encounter.

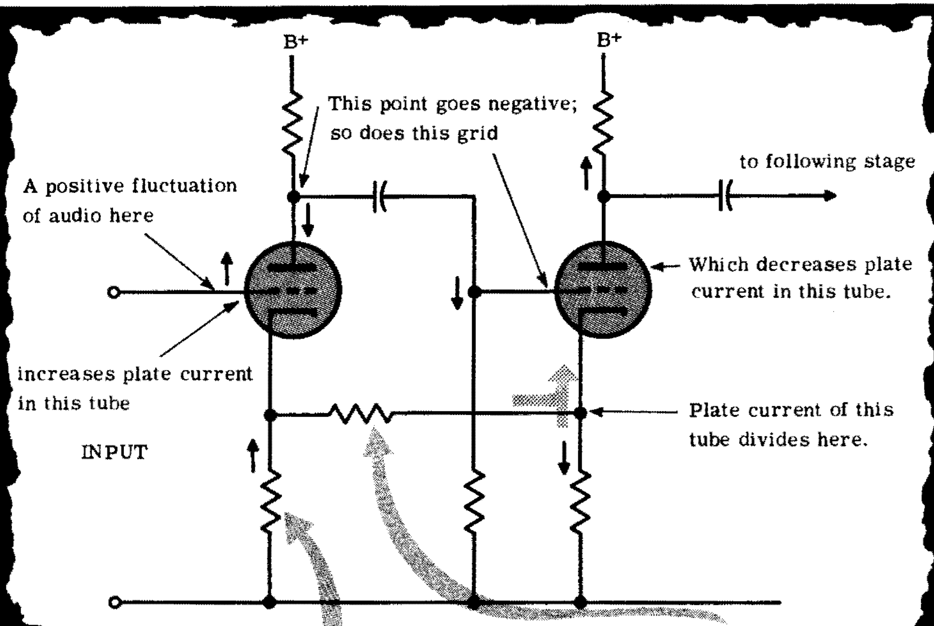
One way of achieving the extra gain without adding extra stages is to use *positive* feedback. Care must be taken to see that the beneficial effects of the negative feedback — reduced distortion — are not cancelled in the process. The secret is to use positive feedback to boost the gain in a part of the amplifier that has very little distortion. Then increased negative feedback can be used over the whole amplifier to reduce distortion elsewhere.

Use of Positive Feedback Avoids Adding of Extra Stages



Positive with Negative Feedback (contd.)

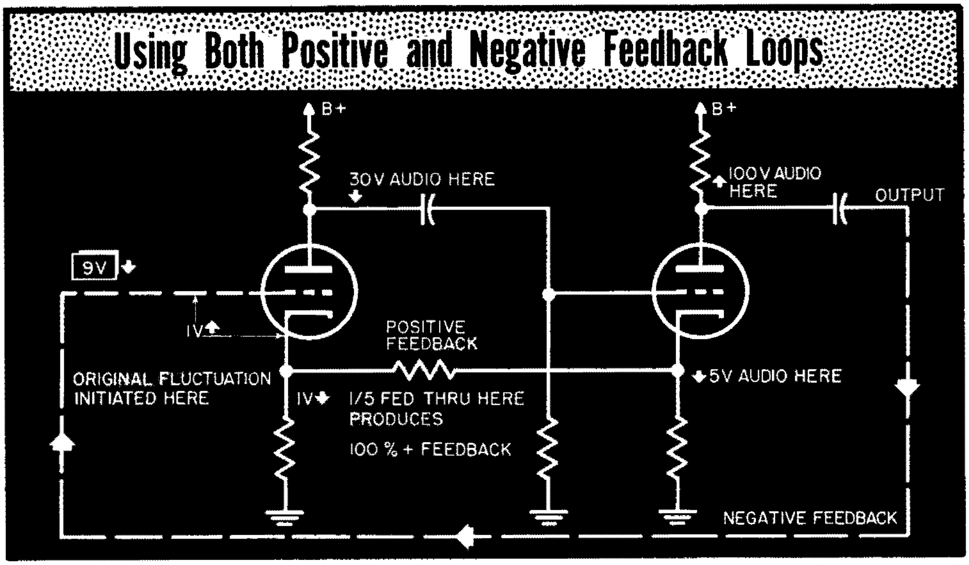
Positive feedback can only be used over one stage in order to avoid oscillation. An easy way to accomplish this positive feedback is to couple cathode bias resistors of two consecutive stages in the earlier part of the amplifier, where the distortion is small.



IF THE REDUCTION IN CURRENT THROUGH THIS POSITIVE FEEDBACK RESISTOR IS GREATER THAN THE INCREASE IN THE PLATE CURRENT FROM THE FIRST TUBE, CURRENT IN THIS RESISTOR WILL DROP (ALSO VOLTAGE), INCREASING THE VOLTAGE FLUCTUATION BETWEEN GRID AND CATHODE AND THUS CAUSING POSITIVE FEEDBACK

A momentary positive fluctuation at the grid of the first stage will produce a momentary negative fluctuation at the plate, which is passed on to the grid of the following stage. This produces another positive fluctuation at the plate of the second tube. At the same time, the negative fluctuation at the plate of the first tube, resulting from increased plate current, will be accompanied by a positive fluctuation at its cathode. Similarly, a negative fluctuation appears at the cathode of the second tube. The fluctuation at the cathode of the second tube is much bigger than that at the first stage. Connecting the resistor between the cathodes will allow some of the fluctuation from the cathode of the second stage to cancel the fluctuation at the cathode of the first stage.

Positive with Negative Feedback (contd.)

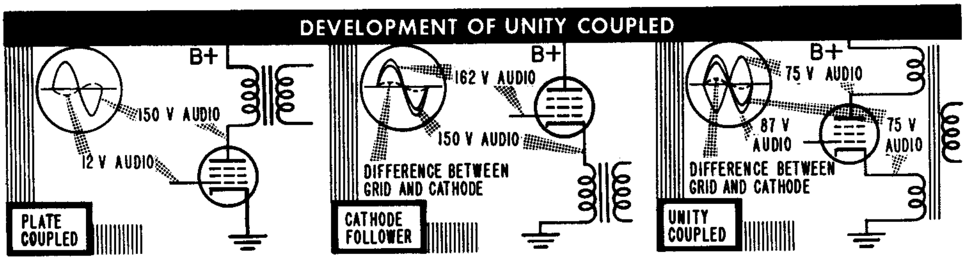


Because of amplification, positive and negative feedback,an input voltage of 9 volts is required.

If the voltage fluctuation fed back caused the cathode of the first tube to move negative by as much as the positive initial fluctuation at its grid, oscillation would take place. (A negative fluctuation at the cathode is equivalent to a positive fluctuation at the grid, hence feedback would provide the total input.) If we use two feedback loops, one positive and one negative, things are not so difficult. Without the negative loop, the initial voltage between grid and cathode of the first stage could feed back enough to cause oscillation, but negative feedback supplies a voltage at the grid that opposes the the initial fluctuation, preventing oscillation from taking place.

Assume that we have an initial fluctuation of 1 volt between the grid and cathode of the first stage. The positive feedback at the cathode that could cause oscillation would also be 1 volt. Now suppose that we provide an amount of negative feedback that *in the absence of the positive feedback* would give a 20-db (10:1) gain reduction. If the positive feedback were not present, we should require a total input of 10 volts instead of 1 volt, the additional 9 volts being required to offset 9 volts negative feedback to the cathode. When both the positive and negative feedback are used, however, the resultant voltage or the feedback effect at the cathode is only 8 volts of negative feedback. This means that the total input need only be 9 volts instead of 10 volts, and that the positive feedback has reduced the amount of gain lost by 10:9 (very nearly 1 db). Alternatively, we could say that when the negativ feedback is added, the effect of positive feedback that could increase the gain of the stage to infinity and cause oscillation is reduced to an increase of less than 1 db.

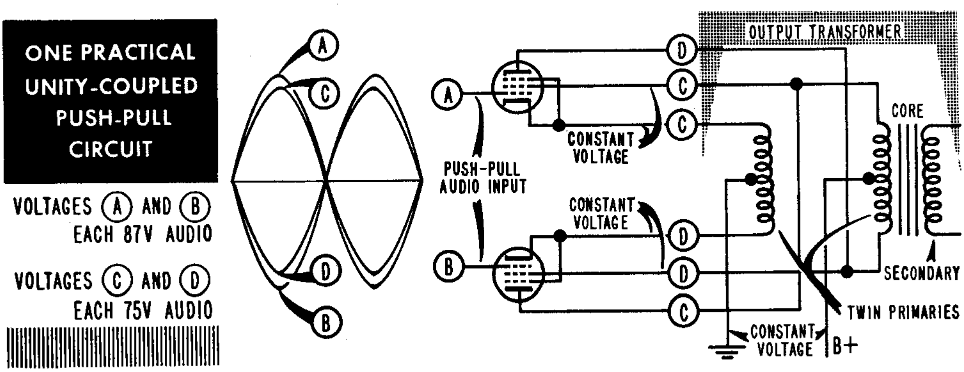
Special Output Circuits



Until feedback came along, the choice for output tubes was between triodes and pentodes. Pentode operation is much more efficient in terms of audio power output for the power input, but it is far more critical of being operated with exactly the right load resistance value than when the same tubes are triode connected.

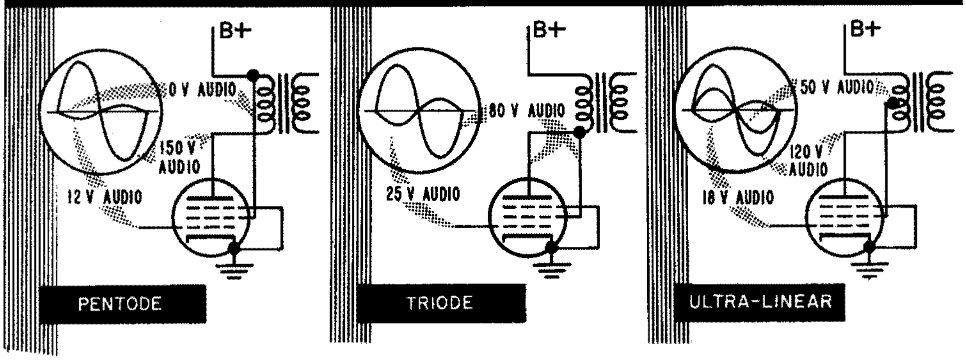
This led to two basic variations in output circuits, although many further minor variations have developed. The first, called "unity coupled," can best be thought of as a "half-way" cathode follower. Assume that we use a pentode tube that needs a 12-volt audio input to produce a 150-volt output across the load coupled to the plate. To go wholly cathode follower would require an input *voltage* of 150 + 12 = *162 volts* to get the power represented by 150 volts across the load coupled in the cathode circuit. But by coupling the load so that the plate circuit feeds half the power and the cathode half, an audio voltage of 75 volts will appear at each. Now the input audio voltage needed is only 75 + 12 = *87 volts.*

To work as a pentode, the screen must always be at a constant voltage "above" the cathode. This can be achieved by using a multiple-wound transformer. One push-pull primary connects to the cathodes of the tubes, with its center tap to the ground. The other, of exactly equal turns, connects to the screen, with its center tap to B+. This insures that the audio voltage on the screen is the same as that on the cathode. For the plate to deliver its half of the power, it must produce an equal but opposite voltage, so the plates are "cross-connected."



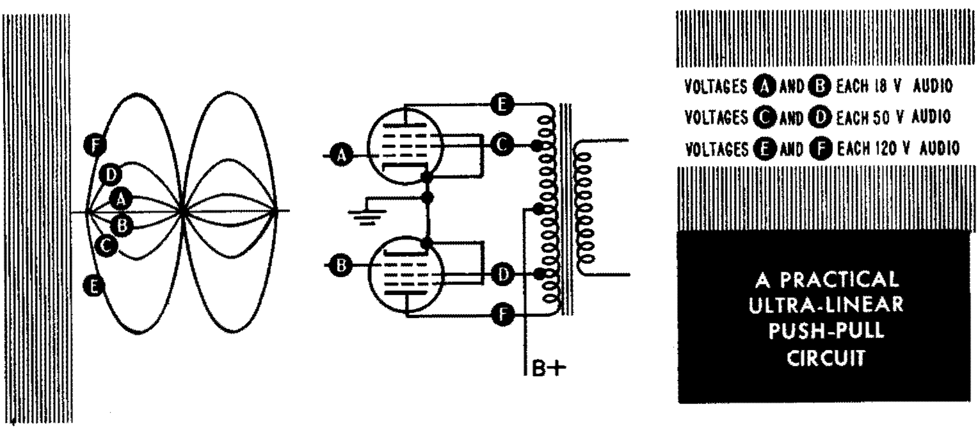
Special Output Circuits (contd.)

DEVELOPMENT OF ULTRA-LINEAR CIRCUIT



The other special output circuit is called *ultra-linear*. We can visualize the operation as half-way between pentode and triode connection. When a tube works as a pentode, the screen voltage remains steady. The only changing voltages in the tube are the grid and plate potentials. To make the same tube work as a triode, the screen is connected to the plate. This means that when the grid goes positive, the plate current rises, making the plate voltage drop. As the screen is also connected to the plate voltage, this goes negative, tending to offset the plate current rise.

Connecting a pentode to make it work as a triode is like applying negative feedback from the plate to the screen. For ultra-linear operation, the screen is connected to a tapping on the output transformer winding so its audio voltage swings the same *way* as the plate, but not as *much*. Thus ultra-linear can be regarded as using less negative feedback from plate to screen than occurs to convert the pentode to triode operation. This means that the advantages, too, are split. Most of the efficiency of pentode working is retained, without its being so critical of having the correct output load resistance.



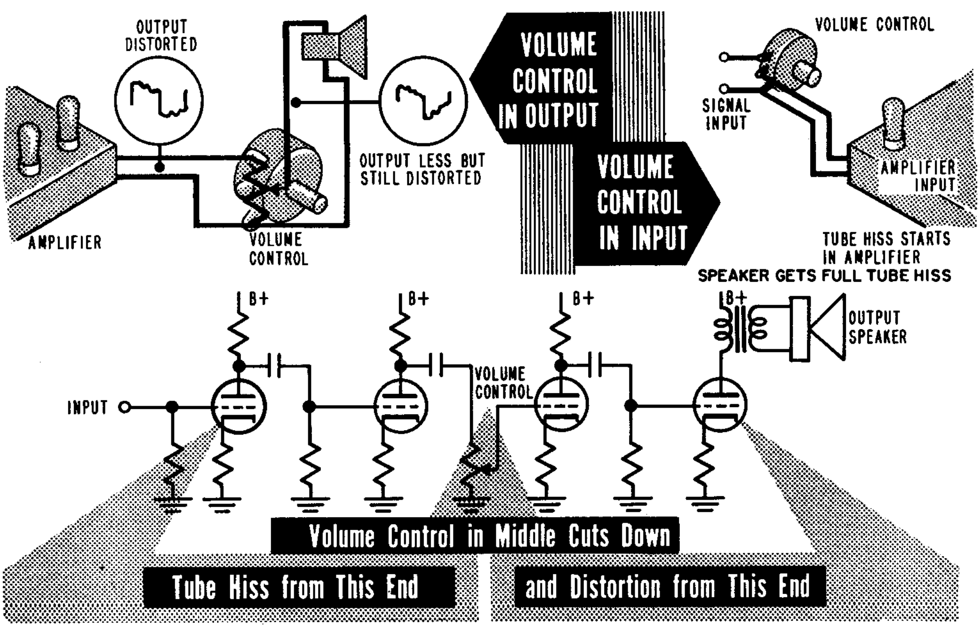
A PRACTICAL ULTRA-LINEAR PUSH-PULL CIRCUIT

Gain Control

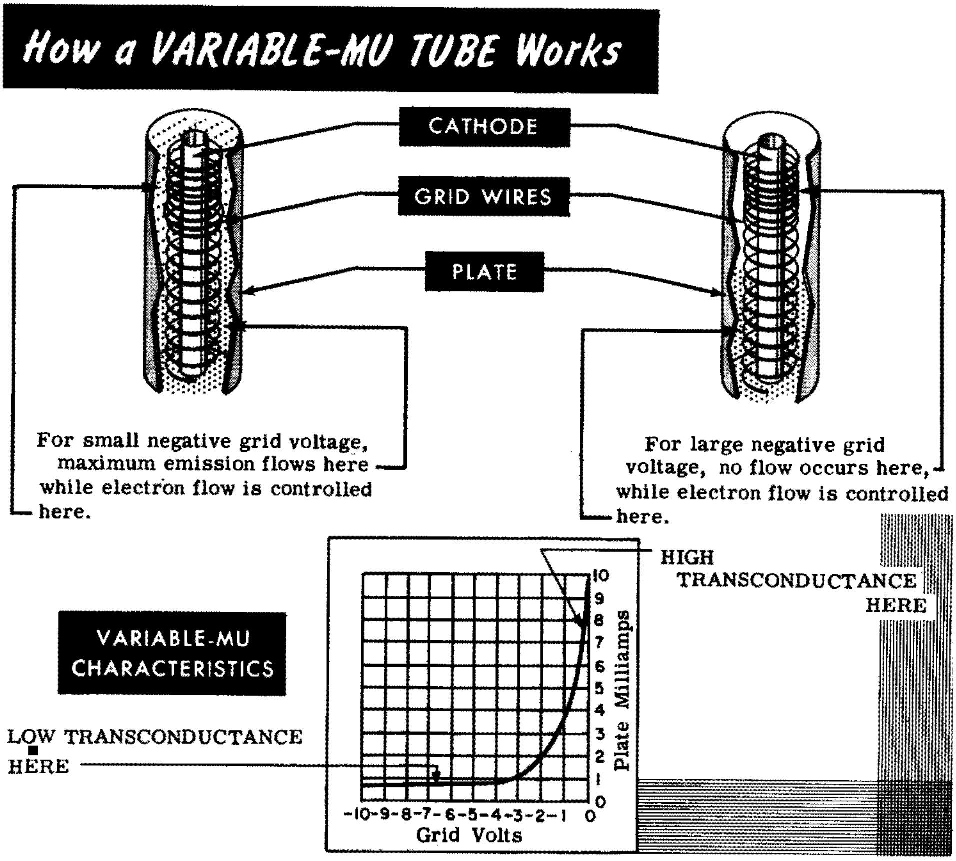
It is easy enough by one means or another to get as much amplification as we want. In fact, if we are not careful, we will go on adding amplification and end up with too much. For this reason, we need some means of controlling amplification, called a *gain* control or a *volume* control. A simple volume control consists of a resistor with a slider riding on it. It is called a *potentiometer*, because it is a device that allows the *potential* at the slider to be varied.

The first idea for a volume control was to put this potentiometer across the output of the amplifier and take the connection to the loudspeaker from the slider and one end. This varies the amount of power delivered to the loudspeaker, but there is one serious disadvantage. If the amplification is too great, the amplifier will distort the signal, and turning down the control will merely adjust the loudness of the *distorted* program.

The next obvious place to put the volume control would be at the input end of the amplifier. Turning down the volume would then eliminate distortion. However, all amplifiers have a limited *dynamic range*. At the input end, the problem is noise, not distortion. Putting the gain control at the front end of the amplifier, means the loudspeaker gets all the noise amplified up from the input stage (which is usually the point at which noise limits dynamic range). For this reason, the best place to connect a gain control is somewhere in the middle of the amplifier.



Gain Control (contd.)

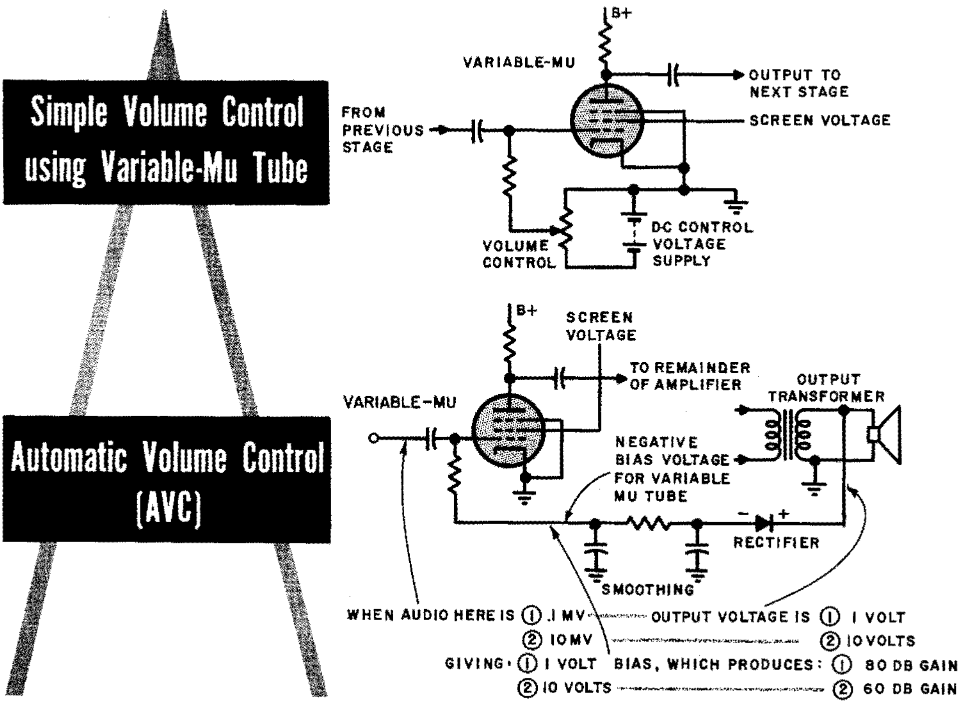


Gain can also be adjusted by altering the operating conditions of a tube, especially if we use a special kind of tube with a considerable amount of curvature in its characteristic. These tubes do not have uniform spacing of the wires that make up the control-grid. This construction produces a tube whose transconductance varies considerably with grid bias voltages. When the grid voltage is only slightly negative, it has practically no effect on electrons passing through the wider part of the grid mesh, but it does influence the number of electrons passing through the closer part of the grid mesh. (In this range, the tube has a high transconductance.)

Making the grid more negative prevents any electrons at all from passing through the closer part of the grid mesh. Those passing through the wider part are subject to less control, and the tube acts as if it had a much lower transconductance. If the change in the spacing of the grid wires is gradual, adjustment of the grid voltage will give a very smooth change of gain as the bias is changed.

Automatic Volume Control (Compressor)

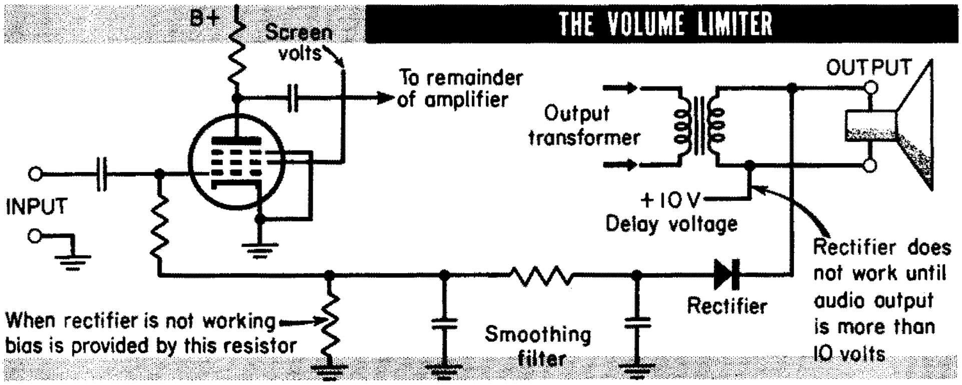
The circuit shown will provide changes in gain with variation in the grid bias of the tube. It can be easily converted to an automatic volume control by sampling the output and rectifying it to produce a d-c voltage proportional to the loudness at the moment. This d-c is then used as a bias for the variable gain stage with the result that louder program material produces a bigger negative bias reducing the amplification.



What this circuit does is to reduce the dynamic range of the program. Suppose that an input of 0.1 millivolt will produce an output of 1 volt, which biases the tube to 1 volt negative. This is not sufficient to change its transconductance, so the amplifier works at full gain — in this case, 10,000.

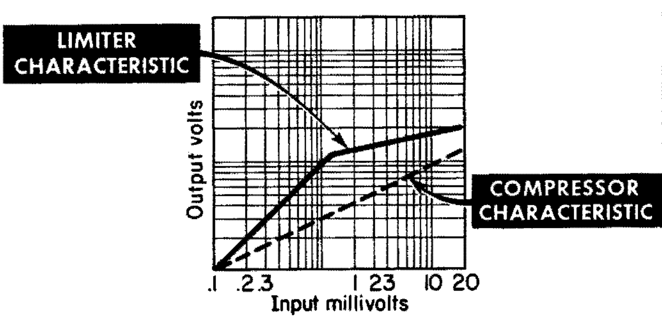
When the output reaches 10 volts, however, the bias for the variable-gain tube has increased to −10 volts. This will reduce the overall gain of the amplifier by a factor of, say, 10:1. The overall gain is now only 1000, instead of 10,000, and the input required to produce the 10-volt output is not 1 millivolt, but 10 millivolts. In this way, the "compressor" has "squeezed" the program material so that a range from 0.1 millivolt to 10 millivolts at the input (40 db) is compressed into a range from 1 volt to 10 volts at the output (20 db).

The Volume Limiter



This type of circuit can also work as a volume *limiter,* rather than a volume compressor. It may not be desired to restrict the dynamic range, but merely to make sure the output stage of an amplifier does not overload, causing distortion. What we want to do is turn the gain control down whenever a very loud passage comes through. In this case, the output is rectified in the same way as before to produce a negative bias. The negative bias is compared to another d-c voltage that corresponds to a point below the maximum output that can be allowed. No change in the bias of the gain-control tube occurs, until this *delay voltage* is reached. As soon as the output exceeds the delay voltage, the grid of the gain-control tube goes negative quite rapidly, turning the volume back, and insuring that no distortion occurs.

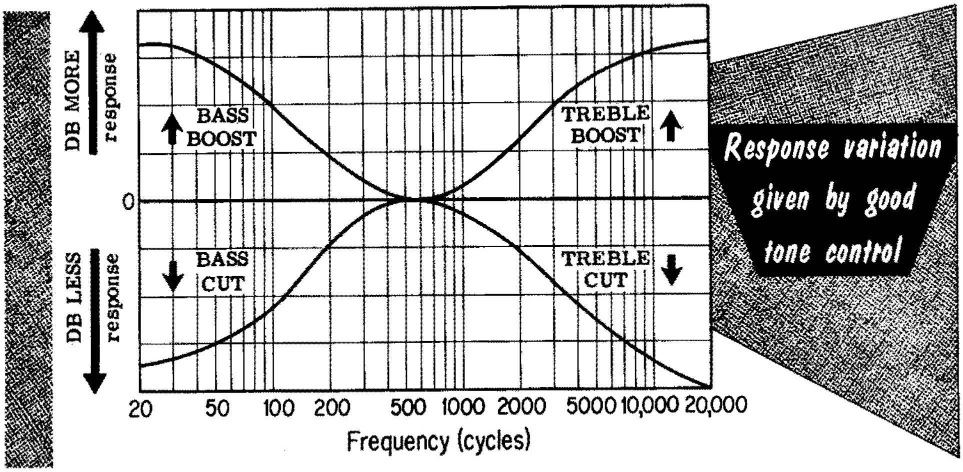
Both the volume compressor and the volume limiter really use a kind of feedback. The difference from the ordinary feedback that we have discussed is that the audio itself is not fed back, merely a d-c voltage taken from it. The fact that this voltage is fed back means that care has to be taken in the design of this kind of circuit (as it is with a feedback amplifier) to make sure that oscillation does not occur. The audio has to be properly rectified, and any a-c components filtered out so that the d-c applied to control the grid will not start oscillation.



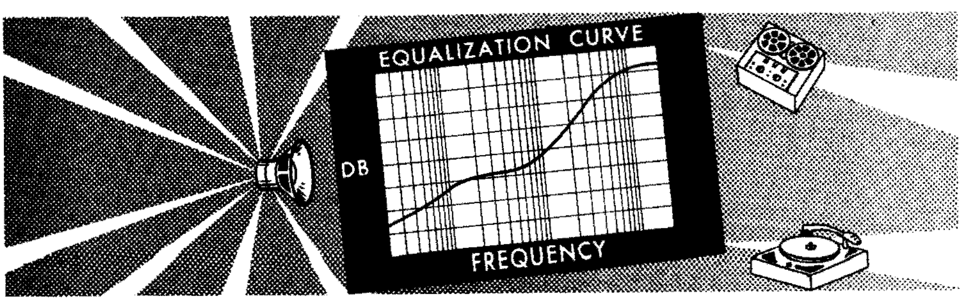
When input is	1	.1 millivolt
	2	1 millivolt
	3	20 millivolts
gain is	1	80 db
	2	80 db
	3	60 db
output is	1	1 volt
	2	10 volts
	3	20 volts
bias is	1	0
	2	0
	3	10 volts

Tone Controls

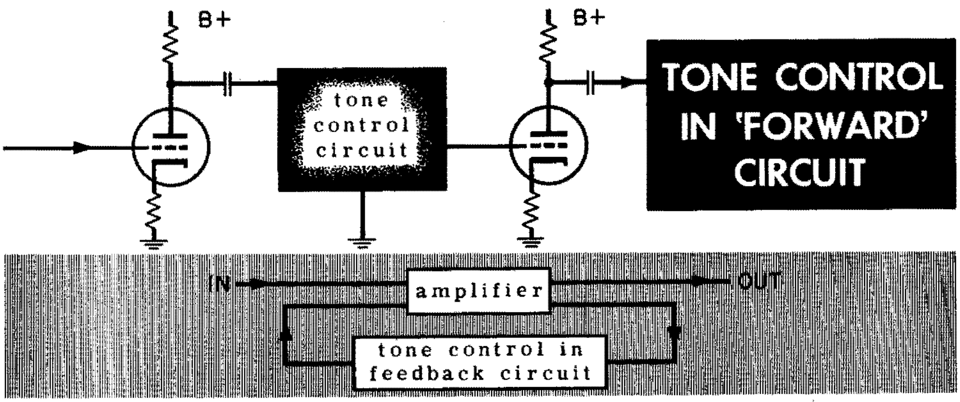
Another kind of control often required in amplifiers is the *tone* control. This name is used to describe an arrangement that will continuously adjust frequency response, increasing or reducing the high- and the low-frequency output. A tone control usually acts to boost or reduce the frequencies toward one end of the range amplified. (The control for the high-frequency end is called a *treble* control, for audio purposes, and that for the low-frequency end is called a *bass* control.)



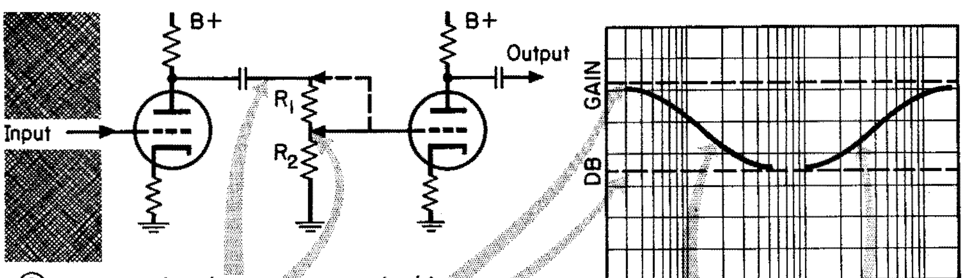
A fixed circuit that adjusts frequency response to a curve previously decided upon is called an *equalizer*. (Sometimes equalizers may also be required to remove undesired resonances, which requires another kind of circuit.) These circuits are called equalizers, because they are usually needed to equalize for the characteristics of something else in the system: to correct for the response deficiencies of a microphone, recorder, pickup, playback head on a tape machine, a loudspeaker, for the studio or listening-room acoustical characteristics, or in some instances, for personal preference in the kind of sound desired.



Tone Controls (contd.)



Tone control and equalizer circuits can work in one of two basic ways. One system adjusts the frequency response of an amplifier on the way through, while the other adjusts it by varying the amount of different frequencies fed back. There are also two kinds of adjustment to the response. One is a reduction of some frequencies in comparison with the rest. The other is an accentuation or boost of some frequencies compared with the rest. Unless we add another stage the amplifier has only a certain amount of total gain and a tone-control circuit containing only resistances, capacitances, and possibly inductances, cannot give us any *more* gain. For this reason the only way to accentuate some frequencies is to cut down frequencies in the rest of the range.

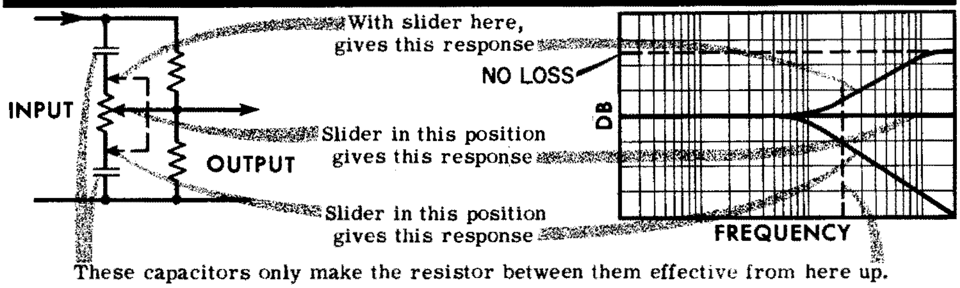


① Connected to here, response is this.
② Connected to here, response is this.
③ By arranging transition with capacitors, response can be this or this.

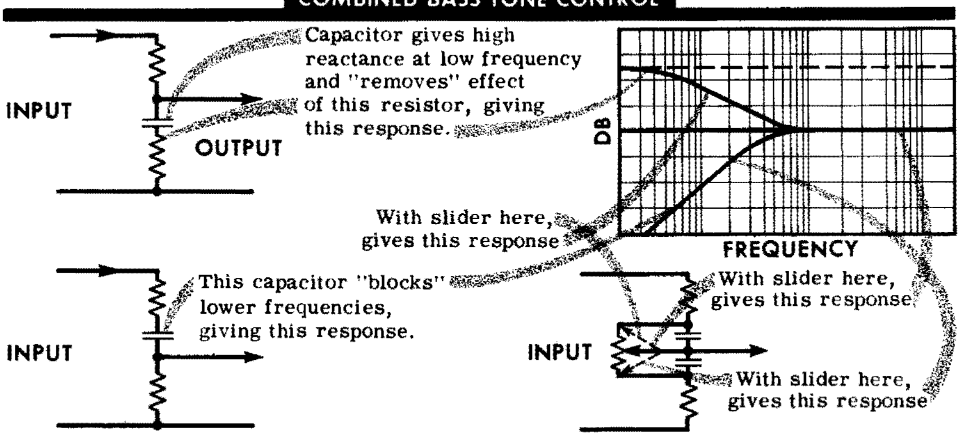
Tone Contols (contd.)

A simple way to achieve tone control is to use a voltage divider between two stages, rather like the method used for volume control. (Here, however, the voltage divider is fixed rather than variable.) A variable voltage divider connected in parallel with the fixed divider through two small capacitors will affect the high frequencies only. (This occurs because the small capacitors block current to the resistors in the variable voltage divider at lower frequencies.)

TREBLE TONE CONTROL

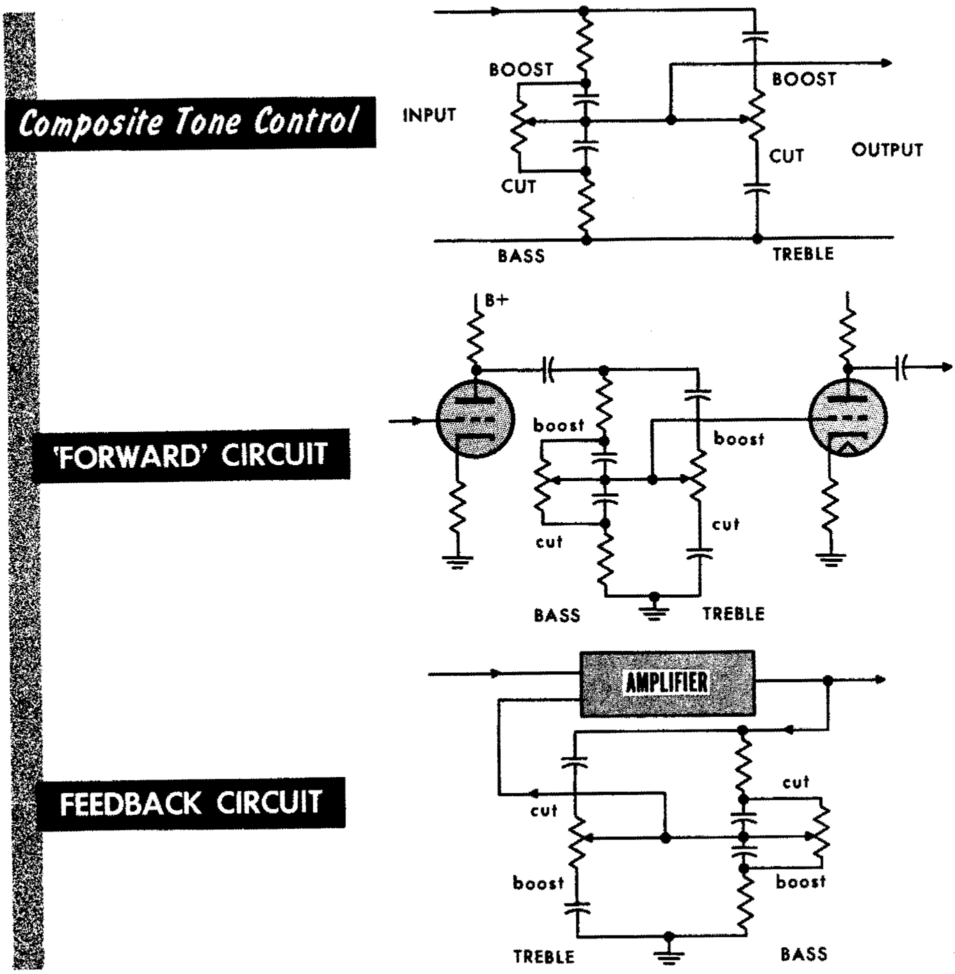


COMBINED BASS TONE CONTROL



Control of the low frequencies can be achieved by inserting capacitors in series with the resistor used for dividing the audio voltage. A capacitance in series with the lower resistor in the drawing will develop a considerable audio voltage, particularly at the low frequencies, resulting in a larger voltage being passed on to the next stage at the extremely low frequencies than over the rest of the audio range. Putting a capacitor in series with the upper resistor develops the greater part of the voltage at the low frequencies and reduces the amount developed across the lower resistor. This produces an attenuation or loss of the extremely low frequencies.

Tone Controls (contd.)



By combining the arrangement and using a potentiometer across the two capacitors, we provide a continuous adjustment that will go from bass boost to bass cut. Most modern tone control circuits combine the two arrangements with two controls, one for the treble boost and cut, the other for bass boost and cut.

An alternative system of tone control uses feedback. In this case, a similar control arrangement is placed in the feedback network. The action of this control is the reverse of that just discussed. If more of the high frequencies are fed back, the amplification at the high frequencies is reduced, resulting in treble cut. If less of the high frequencies are fed back, then the amplification of the high frequencies is increased, resulting in a treble boost.

QUESTIONS AND PROBLEMS

1. How do coupling capacitors and circuit self-capacitance affect the performance of feedback circuits? At a certain frequency, the phase shift due to coupling capacitors, of which there are four, is 45°, 55°, and 30° for the first three. How much phase shift can be allowed in the fourth capacitor at this frequency, for what is nominally negative feedback to become positive?

2. What is a Nyquist diagram? If the basic quantity plotted on the polar diagram is loop gain, how can the diagram be used to show the response of the feedback factor and the overall response shaping?

3. What is gain margin? Explain how an amplifier that is stable with feedback connected can become unstable if its gain is increased, or if the fed-back amount is increased.

4. What is a cathode follower? Explain its action in terms of feedback action. A certain tube, using a plate load resistance of 100,000 ohms, has an a-c resistance of 60,000 ohms and gives an amplification of 48. Rearranging this as a cathode follower; calculate (a) the input impedance with a grid return resistor of 220,000 ohms and (b) the output source resistance (a-c).

5. How can positive feedback be used with negative feedback to get certain improvements? State the specific precautions necessary in adopting this method.

6. Positive feedback is increased just to the point where oscillation commences, without negative feedback. Overall negative feedback that reduces gain by 14 db when the positive feedback is not connected is also used. By how much will the gain change when the positive feedback is now connected?

7. Discuss the relative merits of different positions for a gain control or volume control that lead to choice of an ideal, stating clearly the objections to other positions.

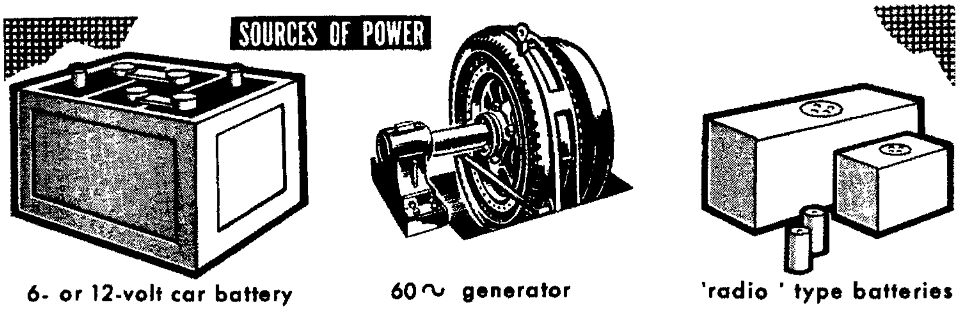
8. Describe the action of a special type of tube that is used to provide automatic volume control electronically. Show how the same basic arrangement can be modified to act as a volume limiter, rather than as a compressor.

9. What is the function of a tone conrtol? Discuss simple circuits that will provide the basic variations required.

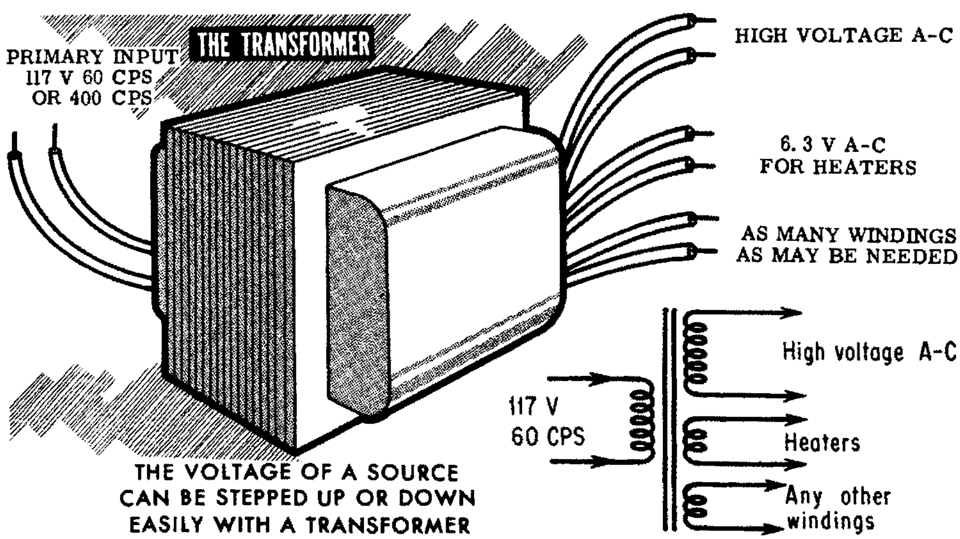
10. Why is equalization necessary? Name various locations where an equalizer is required. In what way does an equalizer differ from a tone control?

Souces of Power

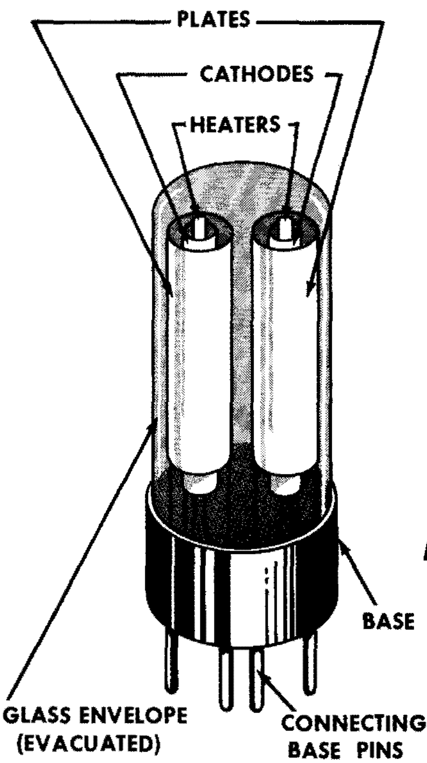
In all our discussions thus far, we have described amplification in terms of tubes and other components that need voltages applied to them to make them work. We have not considered where these voltages come from, but they all need to have the right voltage supplies, whatever they may be.



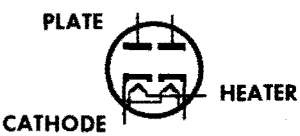
In most modern equipment these supplies come from an electrical power company, most of which deliver a voltage of 117 volts at 60 cycles. Other supplies are sometimes used; for example, in aircraft, where the frequency is not 60 cycles, but 400 cycles, and in automobiles, where the supply is not a-c but d-c, at 6 or 12 volts from a battery. We are principally concerned with 117-volt 60-cycle a-c sources. If the supply is a-c of a different frequency, as in aircraft, the details of the design will be altered but the principles will be the same. Where the supply is d-c, as in automobiles, and in a few isolated locations where the power company supplies it at a higher voltage, such as 110 volts, a different kind of power supply circuit is required. The convenience of a-c as a source of supply is that the voltage can easily be stepped up or down by means of a power transformer.

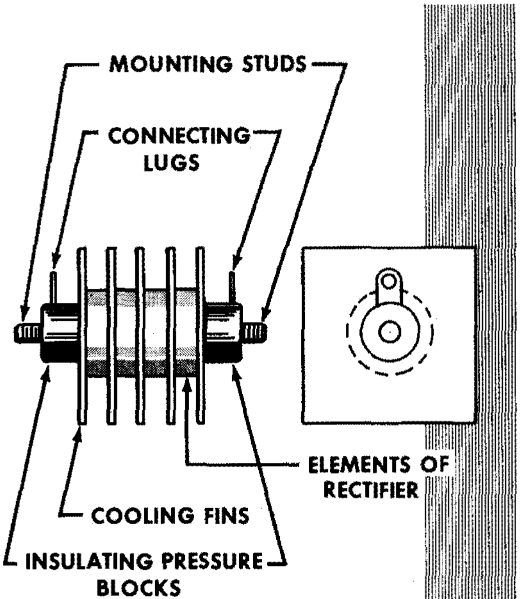


Rectification

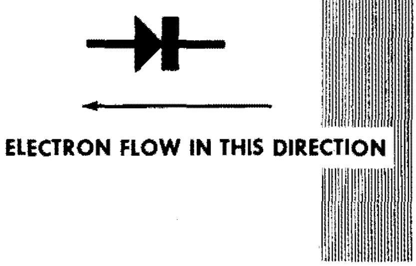


Tube Rectifier



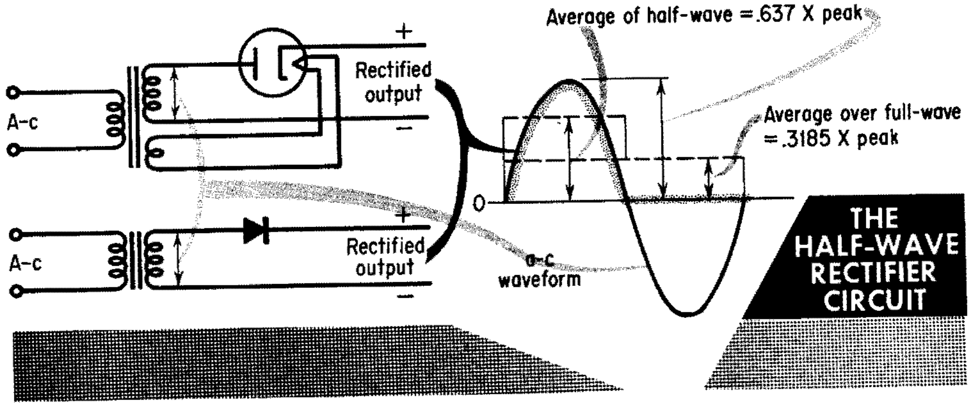


Barrier Layer Rectifier



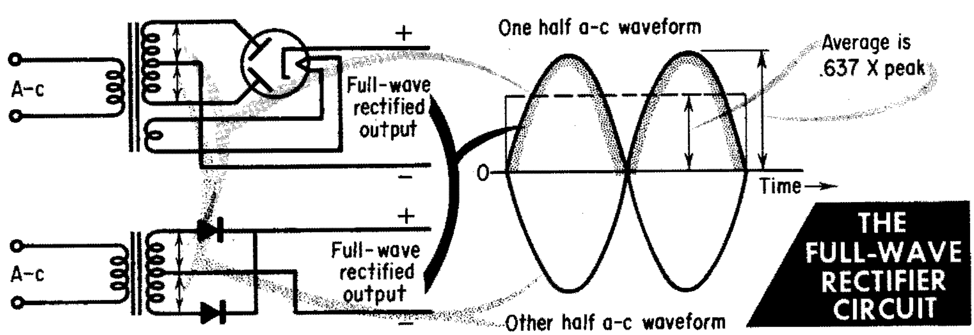
We can get the desired d-c voltage for plate supplies by transforming a-c and rectifying it. Two kinds of rectifier are used for audio amplifier circuits: the thermionic rectifier, which employs one or possibly two electronic diodes in one envelope; and the barrier-layer rectifier, which uses the rectifying properties of copper oxide, selenium, germanium, or silicon. (These materials are listed in the order of improving efficiency and also their sequence of development. Choice of which kind of rectifiers to use usually depends on the relative cost or the particular space requirements, whichever happens to be the most important.

Rectification (contd.)



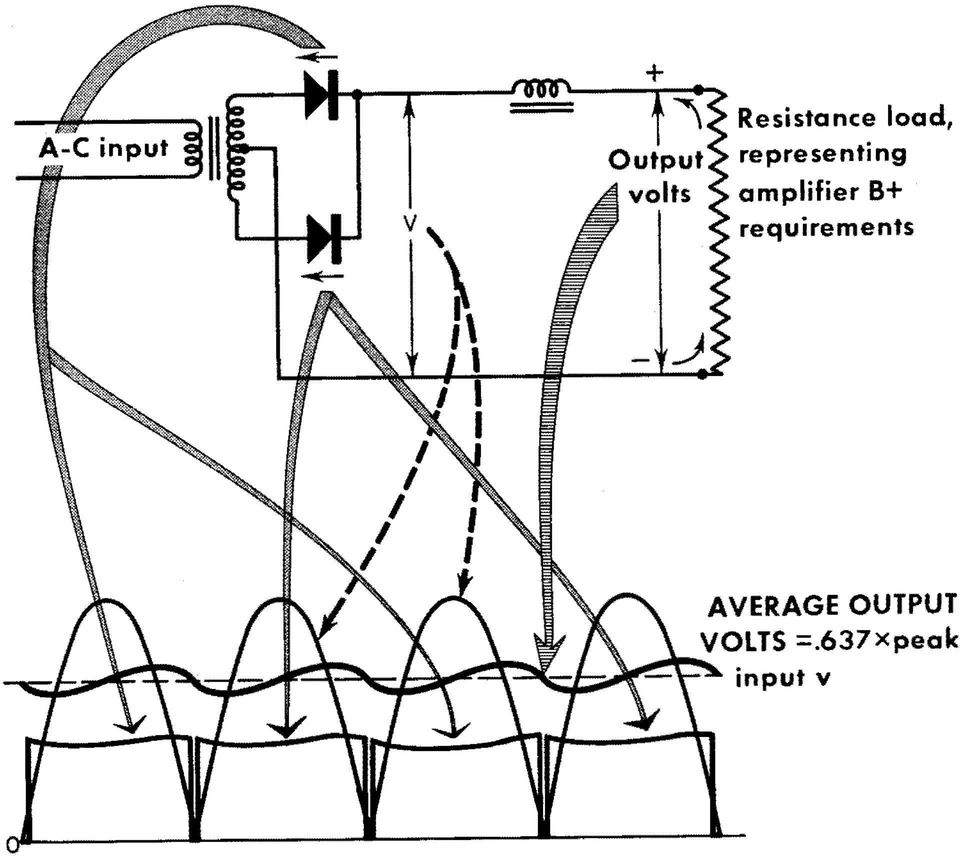
We can use either half-wave or full-wave rectification. Half-wave rectification saves one rectifier component; we only need one diode or one rectifier element. For small plate current supplies, where the current drain is very limited, this circuit is quite convenient. It has, however, the disadvantage that the whole of the current drain has to be passed through the rectifier in a small fraction of *one* half of the cycle. It also needs more elaborate attention when smoothing out the voltage. Where larger plate currents are needed, above say 50 milliamperes, full-wave rectification is almost always used. This enables two current pulses to be taken in every cycle instead of only one, and makes it much easier to smooth out the ripple.

The exact way in which the rectifier works to give the required d-c output voltage depends on the kind of circuit used to smooth out the ripples. If the rectifier is fed straight into a resistance load, the output waveform will be either a succession of half-waves, with a gap for each alternate half-wave, or if full-wave rectification is used, a succession of half-waves end to end. This kind of voltage supply is not suitable for audio amplifiers, because it would result in very considerable hum. For this reason we need to do something to smooth out the ripple component.



Filtering

CHOKE IS USED TO SMOOTH OUT RECTIFIER OUTPUT

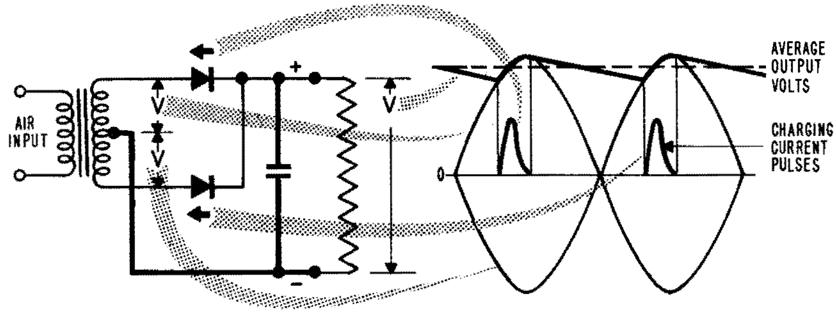


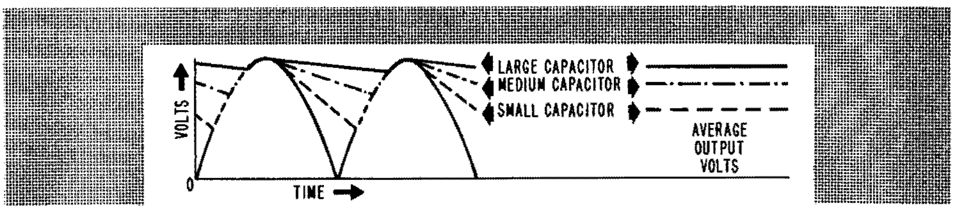
One method of smoothing is to put a choke in series with the feed to the amplifier. If we regard the amplifier as a resistance taking a constant current at a constant voltage, a choke of sufficiently large inductance will pass *almost* constant current and allow a considerable voltage fluctuation across itself. (If an inductance is large, a very large voltage is necessary to produce only a small change in current.) In this case, the voltage at the input to the choke is the same fluctuating voltage that comes out from the rectifier, whereas the output of the choke is an almost smooth d-c because of the almost constant current in the resistance "load."

Filtering (contd.)

The other element of a smoothing arrangement is a capacitor. If we connect a capacitor in parallel with the resistance load that represents the amplifier, the rectifier will charge this capacitor up to the peak value of the alternating voltage. As the wave dips back toward zero, the capacitor will maintain the output current by discharging into the load, thereby keeping the voltage nearly constant between peaks. If only a small current is taken by the load relative to the charge contained in the capacitor at this voltage, the output of a rectifier, using a capacitor in this way, will come very close to the peak voltage of the a-c waveform.

An Input Capacitor May Also Be Used To Smooth Out Output



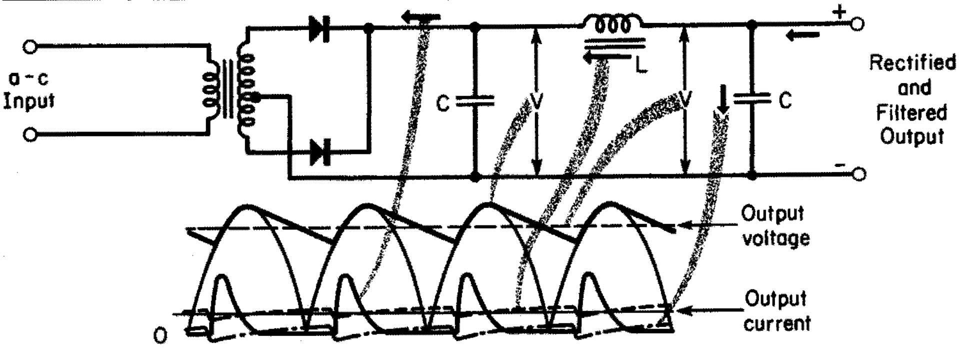


If the capacitor is not large enough to maintain this high a charge over the interval between consecutive half-cycle pulses, the voltage will drop away more during this interval, and the average output voltage will not be quite so high as the peak of the alternating waveform.

A single choke or a single capacitor does not smooth out the ripple completely. The choke has to have a fluctuating current, *however small*, to produce the fluctuating voltage across its terminals. The voltage across the capacitor drops by some amount, *however small*, between charges, before the next pulse comes along to restore the charge. Hence we need further smoothing action to get an adequately smooth or steady d-c and to avoid producing hum in the amplifier.

The Capacitor-Input Filter

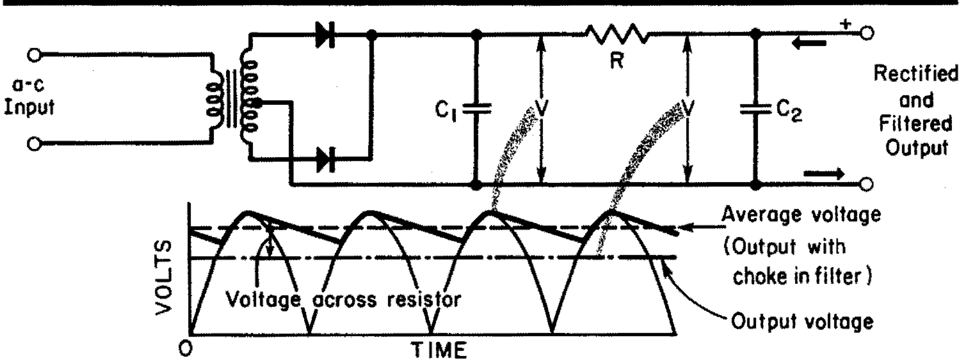
CAPACITOR-INPUT FILTER USING CHOKE



The most common circuit used for smoothing is the *capacitor-input* filter. It produces a starting voltage in the same way as a capacitor connected directly across the load. The load current is then passed through a choke and another capacitor is connected after the choke. Because the fluctuation at the input end of the choke is now quite small, the choke can do much more toward stabilizing the current passing through it. Any residual fluctuations in voltage that still might appear at its output end are "soaked up" by the second capacitor.

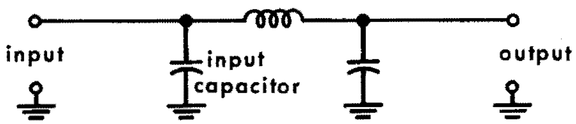
For low-current supplies, or even moderately larger current supplies (up to 100 or 200 milliamps in modern amplifiers), a resistor is sometimes used to replace the choke; this is an economy measure. (Resistors are considerably cheaper than chokes, and modern electrolytic-type capacitors can get very large values of capacitance into quite a small space at low cost.) The disadvantage of the resistor is that it produces a voltage drop so that the *rectified* voltage needed is appreciably higher than the required *output* voltage.

CAPACITOR-INPUT FILTER USING A RESISTOR

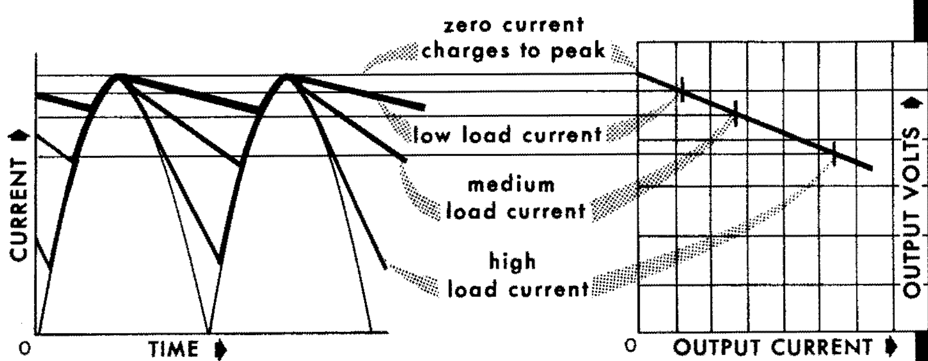


The Capacitor-Input Filter (contd.)

IN A CAPACITOR-INPUT FILTER



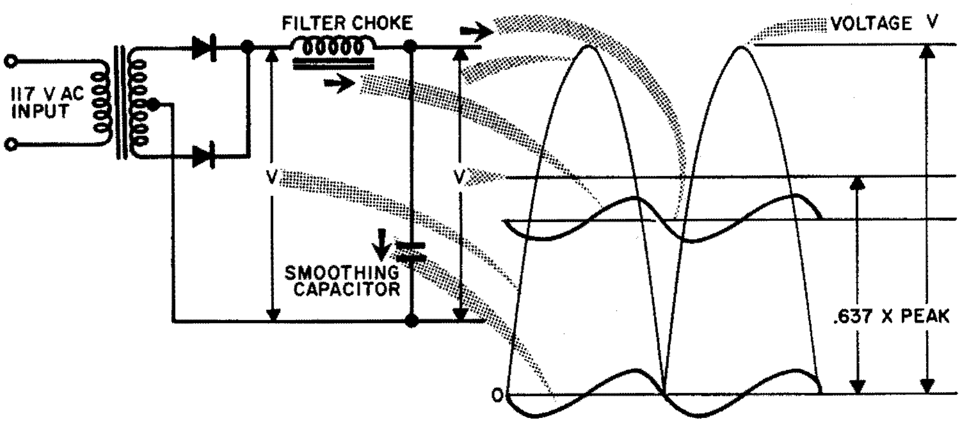
THE RECTIFIED OUTPUT VOLTAGE ALWAYS CHANGES WITH THE LOAD CURRENT



A problem that arises with a capacitor-input filter is that the rectified output voltage always changes with load current. The reason for this is that the load current determines how much the voltage drops between charging pulses. The output voltage is averaged between these peaks and the amount that the voltage drops between them. A larger load current produces a bigger drop, and the average output voltage drops as well. Some kinds of audio circuits require considerable fluctuation in plate current of the output tubes. If the capacitor-input filter is used, the supply voltage also fluctuates with the current. This is where the choke-input filter has an advantage.

The Choke-Input Filter

The Choke-Input Filter



THE OUTPUT OF A CHOKE FILTER IS ALWAYS .637 TIMES THE PEAK ALTERNATING VOLTAGE OR .9 TIMES ITS RMS VALUE.

The voltage at the output end of the choke (provided its resistance is reasonably low) is constant because the choke averages out a voltage fluctuation at its input end that is always the same — from zero to the peak of the alternating voltage. According to each half-wave of the rectifier waveform, the output of a choke input filter is always 0.637 times the peak alternating voltage which is (0.637/0.707) or 0.9 times its rms value.

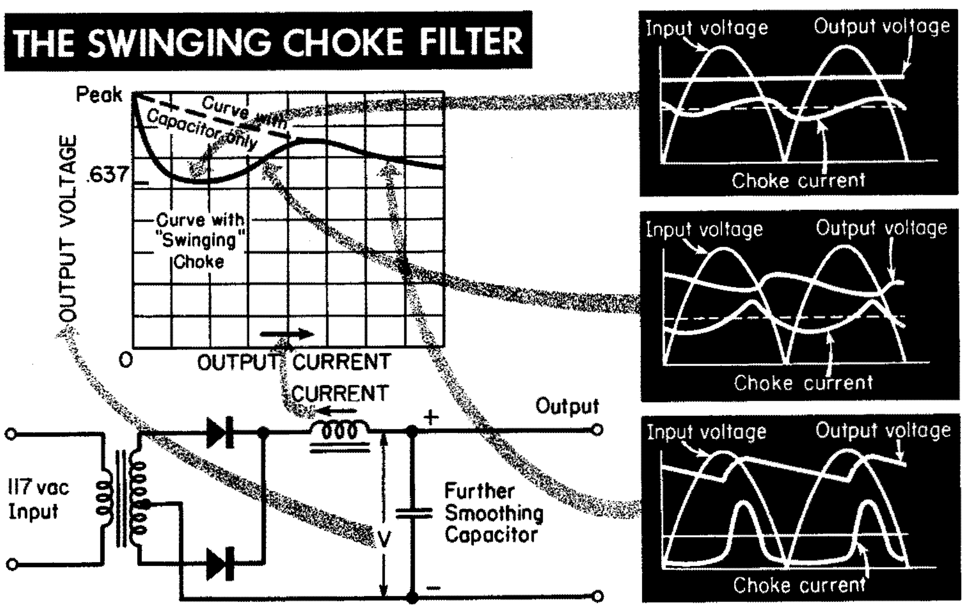
In practice, the output will be slightly lower than this figure, owing to the voltage drop in the resistance of the choke. Further smoothing is achieved by means of a capacitor connected at the output end of the choke. It does not act as a reservoir capacitor as in the case of capacitor input, but merely serves to minimize voltage fluctuation by soaking up the slight *current* fluctuation in the choke.

The Swinging Choke

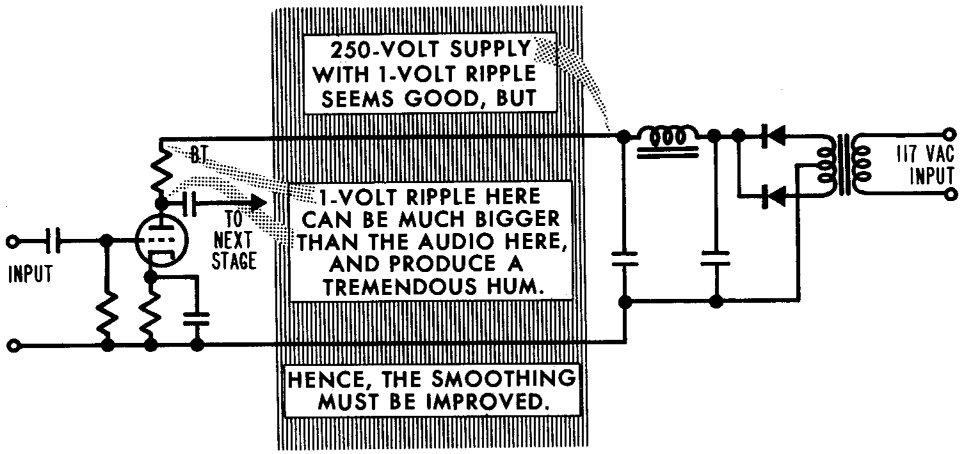
Another kind of filter circuit employs the so-called "swinging" choke. All smoothing chokes employ iron cores with air gaps that prevent saturation. By properly choosing the size of the air gap, a special action is produced. At low load currents, the core is not saturated, but for higher current it progressively approaches saturation, which makes the circuit act as a capacitor-input filter. Capacitor-input filters produce higher output voltages; hence, the output at the filter can be made to *rise* with increased load current.

At small load currents, the inductance of the choke is sufficient to make the filter behave as a choke-input arrangement, and the output voltage is not more than 0.637 of the alternating peak voltage. As the current drain increases, the choke begins to saturate, and the rectifier starts pulse-feeding the capacitor at the output end of the choke. The circuit then begins to act as a capacitor-input filter and the output voltage rises.

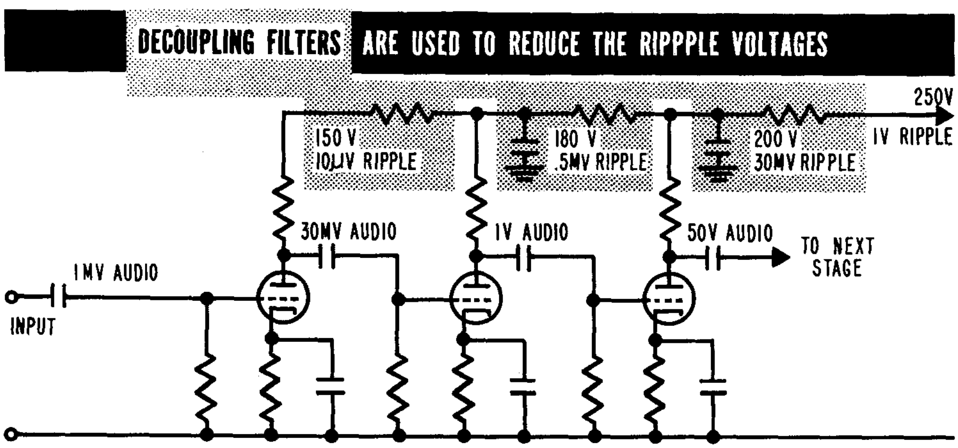
Because the current is increasing at the same time, the output cannot possibly *reach* the peak value of the applied a-c because the drain effect will cause dips between the peaks, but the average voltage *can* rise with a carefully designed filter of this kind. This is useful because it will serve to offset the voltage drop in the supply circuit that always tends to reduce the output voltage with increased load current. If the rise produced by the swinging choke just offsets the losses produced by increased current through the rectifier, the power transformer, and possibly a further smoothing choke, the output voltage of this kind of filter will be almost perfectly constant as the load current is changed.



Decoupling



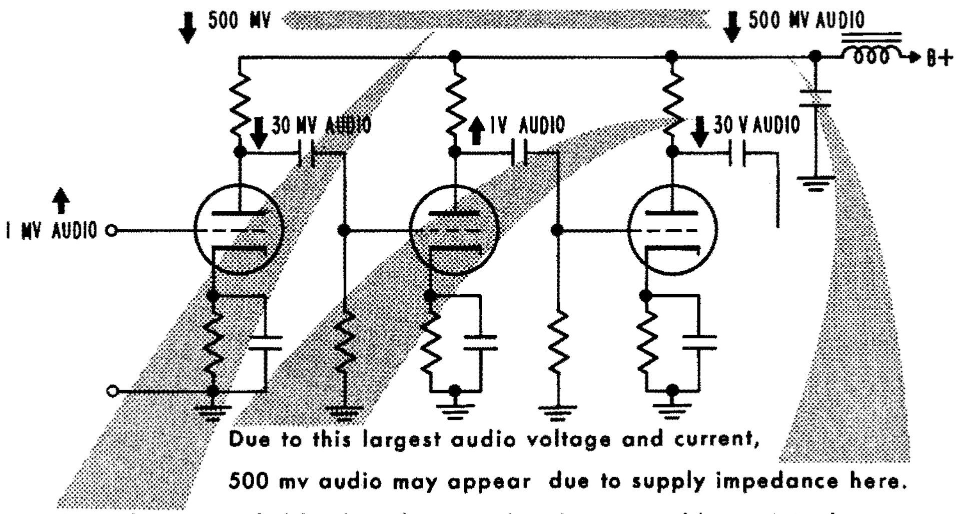
The filters that we have discussed will reduce the ripple to well below 1 volt in a 250-volt supply. This would seem to be quite good enough, until we consider that this plate supply may be needed for the first stage of a high-gain amplifier. The audio voltage at the plate of the first stage may not be more than, say, 10 millivolts. If there should be as much as 10 millivolts of hum in the plate supply voltage fed to the top end of the plate load resistor, the ripple would be equal to the audio voltage at this point. The hum, of course, has to be kept well below the audio voltage to avoid its becoming audible. This means that extra smoothing is needed to cut the hum down to a much lower value. (This attention is not necessary at the output stage, where there may be 100 volts or more audio, so we can use the simple smoothing circuit there.) The additional smoothing required in the supply for the early stages may be provided by additional resistors and capacitors.



Decoupling (contd.)

Fortunately, high-gain low-level stages do not take much plate current. This means that relatively large-value resistors and capacitors that provide a high degree of ripple reduction can be used without dropping the plate potential appreciably. These additional components are necessary for another reason. We discussed feedback (at the beginning of this volume) as something desirable that we introduce intentionally. Feedback can also be undesirable and be introduced unintentionally. This is one place where this occurs if care is not taken.

UNDESIRABLE FEEDBACK CAN CAUSE OSCILLATION

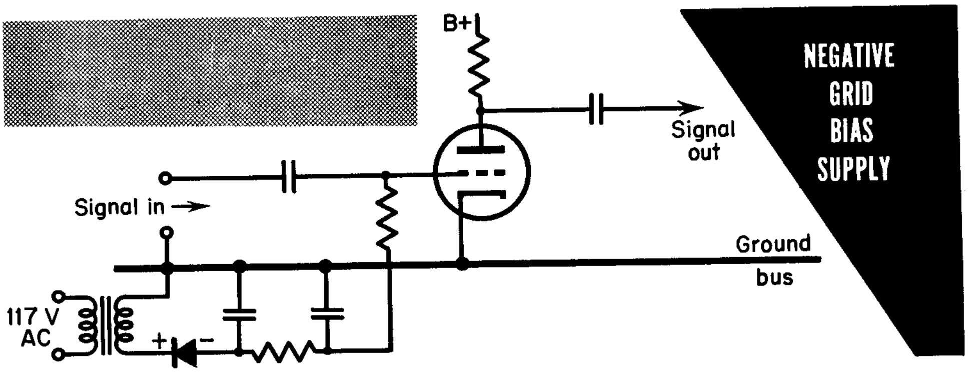


This voltage gets fed back and is in right 'phase' to add to original 30 mv at this plate, probably causing oscillation. This also may be avoided through the use of decoupling resistors and capacitors.

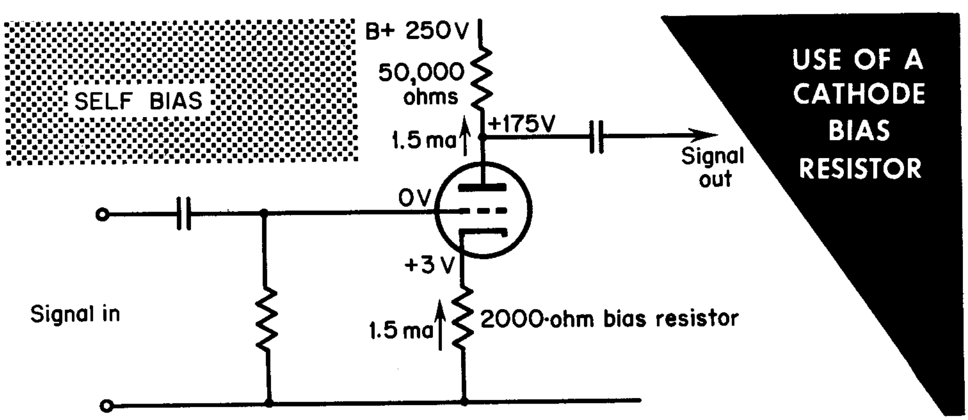
The normal high-voltage supply has an impedance due to the reactance of the final smoothing capacitor and resistors in the circuit that varies from a few ohms to perhaps several hundred ohms, according to design. Even if this impedance is only a few ohms, the output-stage audio current will probably be a fluctuation of 50 milliamperes or more and the power-supply impedance will produce half a volt or more audio across it. This half-volt of audio superimposed on the high voltage supply, will be injected into the front end stage, unless we provide some means of getting rid of it. The further resistor and capacitor, used to reduce the ripple or hum voltages as well as smoothing, bypass *(decouple)* this audio voltage. For this reason these extra resistors and capacitors are called decoupling elements.

Grid Biasing Methods

We have seen how to achieve the correct supply voltage for the plate circuits, but to work in an amplifier correctly, each tube must also have the correct grid bias voltage and, if it is a pentode, the correct screen voltage. There are two ways of providing the grid bias voltage: one method uses a a separate supply of fixed voltage. This can be the same as any of the plate supply arrangements, except that the polarity is reversed and very little current is needed.



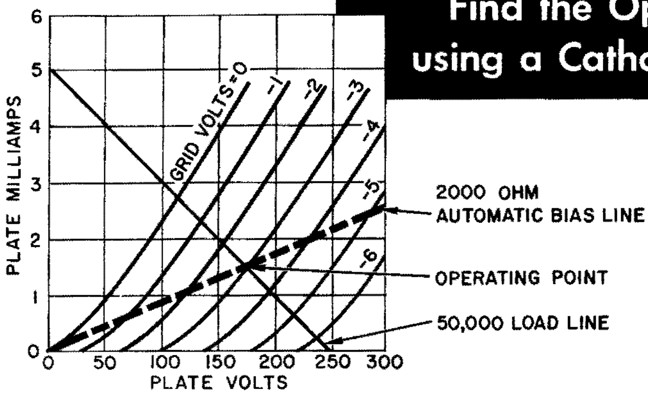
However, the most commonly used method works by making the cathode positive rather than the grid negative. As far as the tube is concerned, the effect is the same. A small resistor, from 100 to 5000 ohms, according to plate current and the bias voltage required, is connected between cathode and ground. This makes the cathode positive with respect to ground by an amount that varies with the plate current. Because one end of the grid resistor is connected to ground, the d-c or bias voltage on the grid is that of ground. For this reason, making the cathode positive with respect to ground will be the same as making the grid negative from cathode.



Self Bias

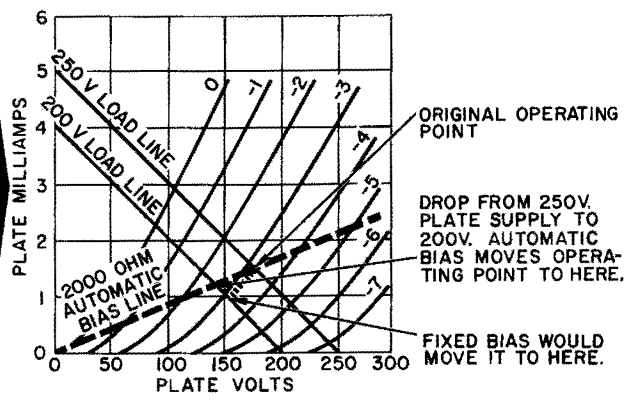
Assume that the cathode resistor is 2000 ohms. If the plate current is 1 milliampere, the grid bias will be 2 volts, and so on. We can now find the operating point of this value of cathode resistor by plotting a curve on the tube characteristics. We mark the points where the 2-volt bias curve crosses 1 milliampere, where the 4-volt bias curve crosses 2 milliamperes, and so on. The operating point of the tube is the point at which this curve crosses the load line for the chosen value of the plate resistor.

Find the Operating Point, using a Cathode-Bias Resistor



OPERATING POINT IS AT INTERSECTION OF LOAD LINE AND AUTOMATIC BIAS LINE

Fluctuations of Plate Voltage affect a Self-Biased Circuit



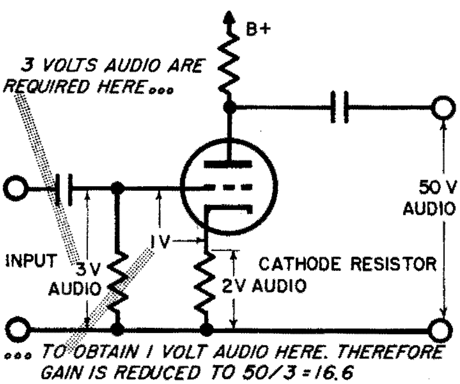
An advantage of this method of biasing is that the bias automatically adjusts to any variations in the circuit. Suppose, for example, that the plate potential drops from 250 volts to 200 volts. If the bias voltage were fixed, this might well over-bias the tube. Using this bias system, however, the shift in load line corresponding with the drop in plate supply voltage produces a new operating point, which will still be optimum. For this reason, this method of biasing is called *automatic* or *self* bias.

Self Bias (contd.)

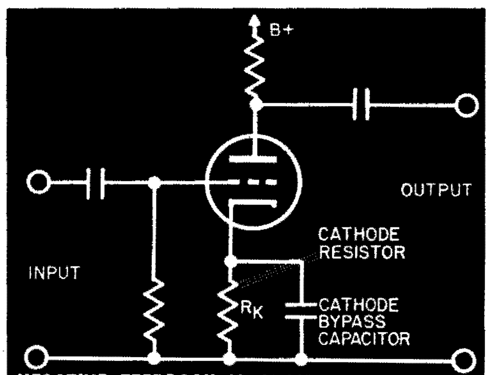
The cathode resistor is not quite all that is needed for providing bias in some instances. Suppose that the plate load resistor is 50,000 ohms and that, with this value of load line, the tube gives a gain of 50. A 1-volt audio signal between grid and cathode will produce 50 volts audio at the plate. The same audio current fluctuation passes through both the plate and the cathode resistors; consequently there will be a proportionate audio voltage at the cathode.

Because the plate resistor is 50,000 ohms and the cathode resistor is 2000 ohms, a 50-volt fluctuation at the plate will be accompanied by a 2-volt fluctuation at the cathode. This fluctuation effectively takes the grid positive from its bias point, increasing plate current, which makes the plate swing negative and the cathode swing positive. Thus a positive fluctuation from grid to cathode will be accompanied by a positive fluctuation from cathode to ground. The total input voltage from grid to ground must be the total of these fluctuations: 1 volt from grid to cathode and 2 volts from cathode to ground, or 3 volts from grid to ground. Thus a 3-volt input is required to produce a 50-volt output and instead of the tube giving a gain of 50, the gain is only about 17. The cathode resistor is providing negative feedback.

To get the full gain of the tube we must avoid this feedback effect. This is accomplished by shunting the cathode resistor by a large-value low-voltage electrolytic capacitor that bypasses the audio voltages. A 50-microfarad capacitor has a reactance of only 63 ohms at 50 cycles and much lower reactances at higher frequencies. With a 1-volt audio input from grid to cathode (at 50 cycles), there will be 50 volts audio output at the plate and about 60 millivolts at the cathode, which is not sufficient to make an appreciable difference in the gain of the stage.

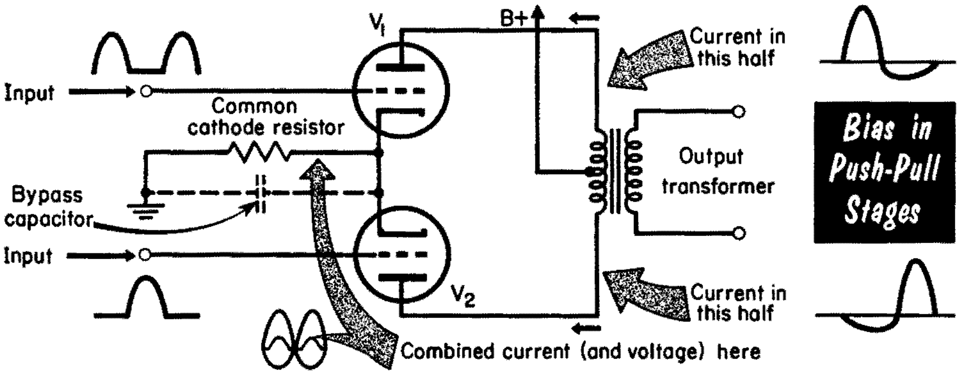


THE CATHODE RESISTOR PROVIDES NEGATIVE FEEDBACK



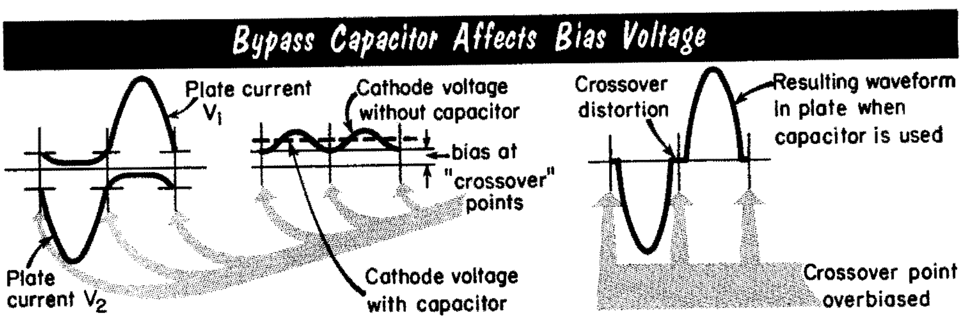
NEGATIVE FEEDBACK IS AVOIDED THROUGH THE USE OF BYPASS CAPACITOR WHICH BYPASSES AUDIO CURRENT. ONLY THE STEADY D-C CURRENT FLOWS THROUGH R_K

Bias in Push-Pull Stages



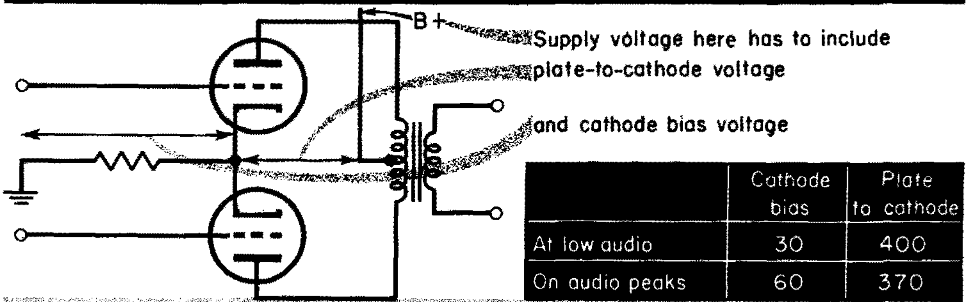
Self bias using a resistor in the cathode circuit is found in push-pull stages as well as in "single-ended" stages. When the push-pull stage is operated class AB, the plate currents in the tubes change appreciably during different parts of the waveform. These currents are added in the common cathode resistor, not subtracted as in the output transformer primary. This results in a double-frequency current in the cathode resistor, because of the asymmetrical current waveforms in the tubes.

The usual operating point of the tubes is arranged so that when no signal is passing, the plate current is appreciably less than the maximum signal-current fluctuation in each direction. This current provides a voltage drop across the cathode resistor that establishes a bias that is quite close to cutoff. When signal current flows, this voltage drop increases, increasing the bias, and reducing the gain of the tube as in a single-ended stage. If a capacitor is connected across the cathode resistor, the fluctuation due to signal currents is smoothed out, resulting in an almost steady bias that is always higher than that present when no signal passes. This means that when the current waveform in the tubes falls toward zero (for which the bias should be at the no-signal level), the higher bias provided by the capacitor may cause premature cutoff and *crossover distortion.*



Bias in Push-Pull Stages (contd.)

PUSH-PULL STAGE USING AN UNBYPASSED CATHODE RESISTOR

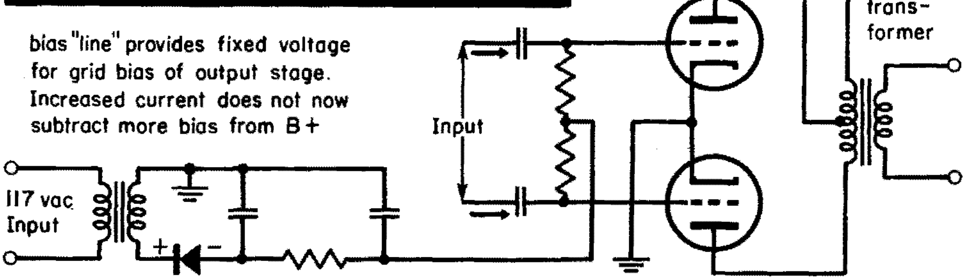


	Cathode bias	Plate to cathode
At low audio	30	400
On audio peaks	60	370

NOTE: This table assumes that B+ does not drop with increased current demand. If it does, reduction in plate-to-cathode voltage on peaks is even worse.

Unfortunately, we cannot solve this problem by using a smaller cathode resistor because it would result in too high a current in the tubes when no signal was passing. The only solution is to omit the decoupling capacitor across the cathode resistor. If we wish to use self bias at all, we must be willing to sacrifice some gain.

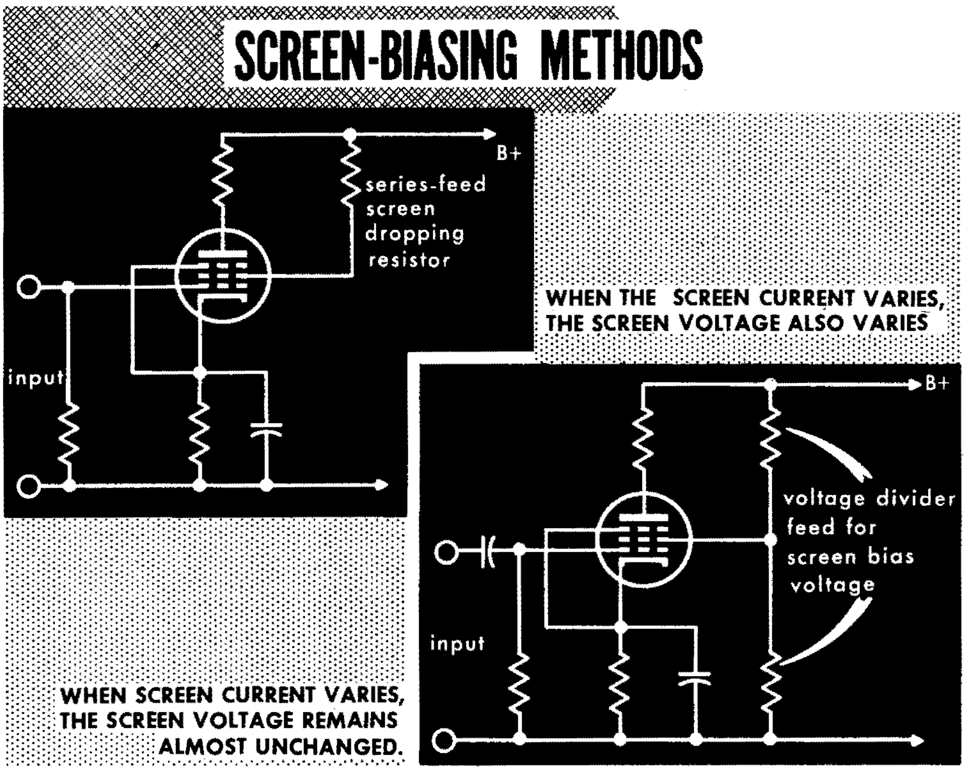
A SEPARATE BIAS SUPPLY IS USED TO OBTAIN AN INCREASED OUTPUT FROM THE PUSH-PULL STAGE



For this reason, increased output can often be obtained from a push-pull stage by using a separate fixed-voltage bias supply. This supply is usually a simple rectifier that takes an alternating voltage from a suitable point (such as from a voltage divider connected across the high-voltage secondary of the power transformer) and rectifying it with a single diode. A simple resistor-capacitor combination will provide sufficient smoothing because there is no grid *current* requirement. This means the capacitors charge up to the peak alternating voltage, and the resistor merely provides additional filtering to prevent the small leakage current pulses from being passed on to the grid circuits.

Screen-Biasing Methods

For a pentode tube, a supply is also required for the second (screen) grid. This supply usually has to provide a fixed voltage not greater than the average plate voltage. It would be possible to design a completely separate supply for the screen grid, but because this electrode only requires a small current (a fraction of a milliampere for small tubes of the voltage-amplifying type and a few milliamperes for large output tubes), a separate supply would involve unnecessary expense.



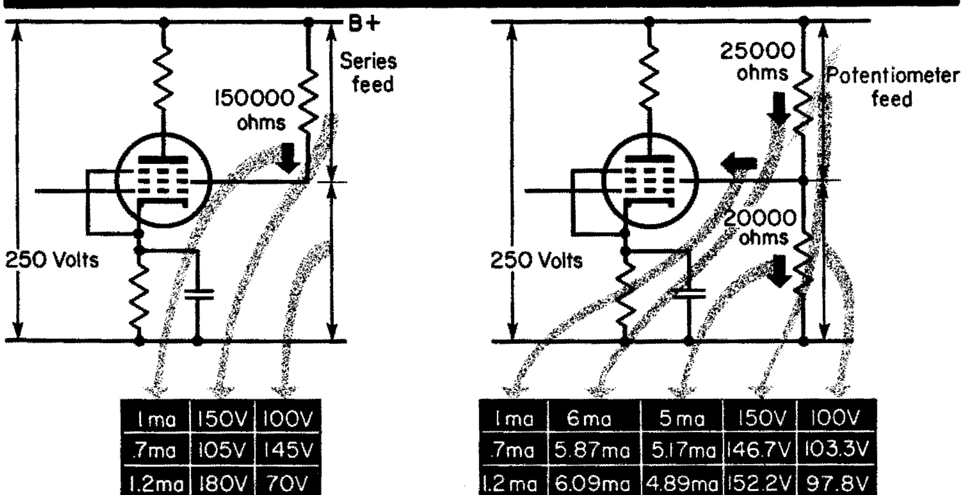
Instead, the screen supply is usually derived from the plate supply, using either a series resistor or a voltage divider. The series-resistor method has the advantage of economy on supply current, because it passes only the current necessary to feed the screen. The disadvantage is that, if the screen *current* should vary (which it does under certain conditions), the screen *potential* will also vary, because of the change in voltage drop across the feed resistor. This difficulty may be overcome by using the voltage-divider method *(potentiometer feed)*. The two resistors pass a current that is larger than the average screen current and thereby keep the screen potential more steady.

Potentiometer Feed

Suppose that the supply potential is 250 volts, the average screen current 1 milliampere, and the desired screen potential 100 volts. A single 150,000-ohm resistor would provide the necessary 150-volt drop at this current. If the current rose to 1.2 milliamperes, however, the drop across the resistor would increase to 180 volts, leaving only 70 volts at the screen. If the current should drop to 0.7 milliampere, the resulting drop in the resistor would be only 105 volts, leaking 145 volts at the screen. It is clear that comparatively small changes in screen current result in quite large changes in screen potential, which means that the operating conditions of the tube are not very steady.

If we arrange a voltage divider that draws 5 milliamperes in addition to the screen current, this situation will be improved. The 5-milliampere current from screen to ground will require a resistor of 20,000 ohms to drop the required 100 volts; the 6 milliamperes from the supply to the screen, a difference of 150 volts, will require a resistor of 25,000 ohms. Now an increase in screen current from 1 to 1.2 milliamperes will only change screen potential to about 97.8 volts. If the screen current should drop to 0.7 milliamperes, instead of rising to 145 volts the screen potential will only rise to about 103.5 volts. In this way the extra 5 milliamperes flowing in the voltage divider helps considerably in maintaining a steady screen potential.

Advantages of the Screen Potentiometer Feed



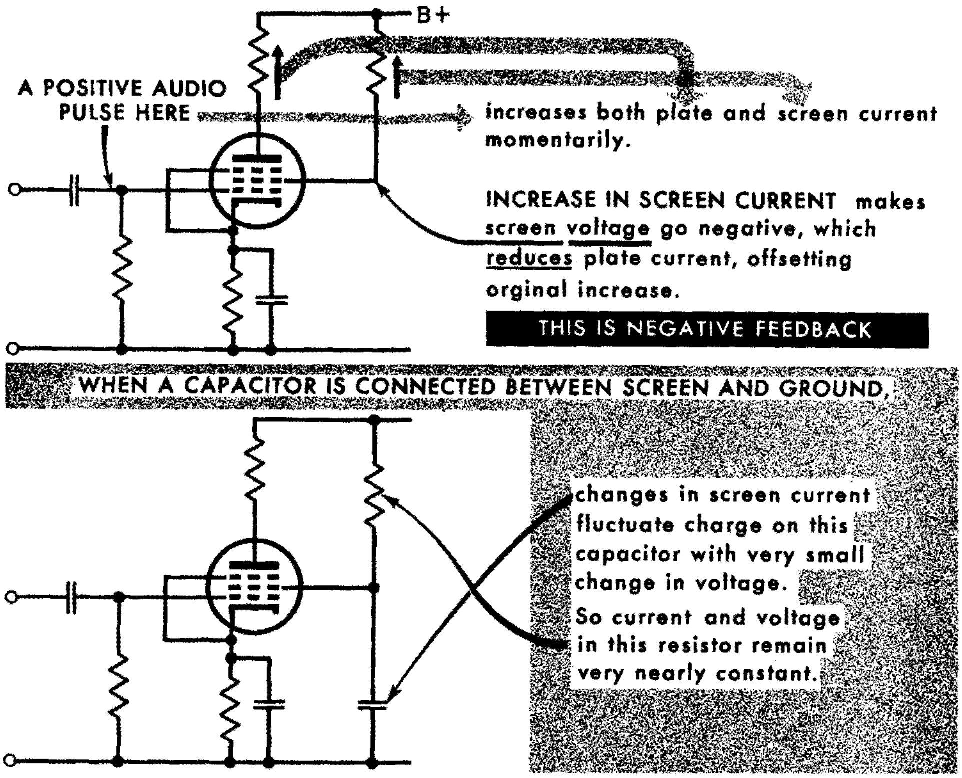
1 ma	150V	100V
.7ma	105V	145V
1.2ma	180V	70V

1 ma	6 ma	5 ma	150V	100V
.7ma	5.87ma	5.17ma	146.7V	103.3V
1.2 ma	6.09ma	4.89ma	152.2V	97.8V

The potentiometer causes much reduced variation of screen voltage in this column for the same change of screen current in this column.

Maintaining Constant Screen Potential

AN AUDIO SIGNAL APPLIED TO THE GRID WILL CAUSE THE SCREEN CURRENT TO VARY



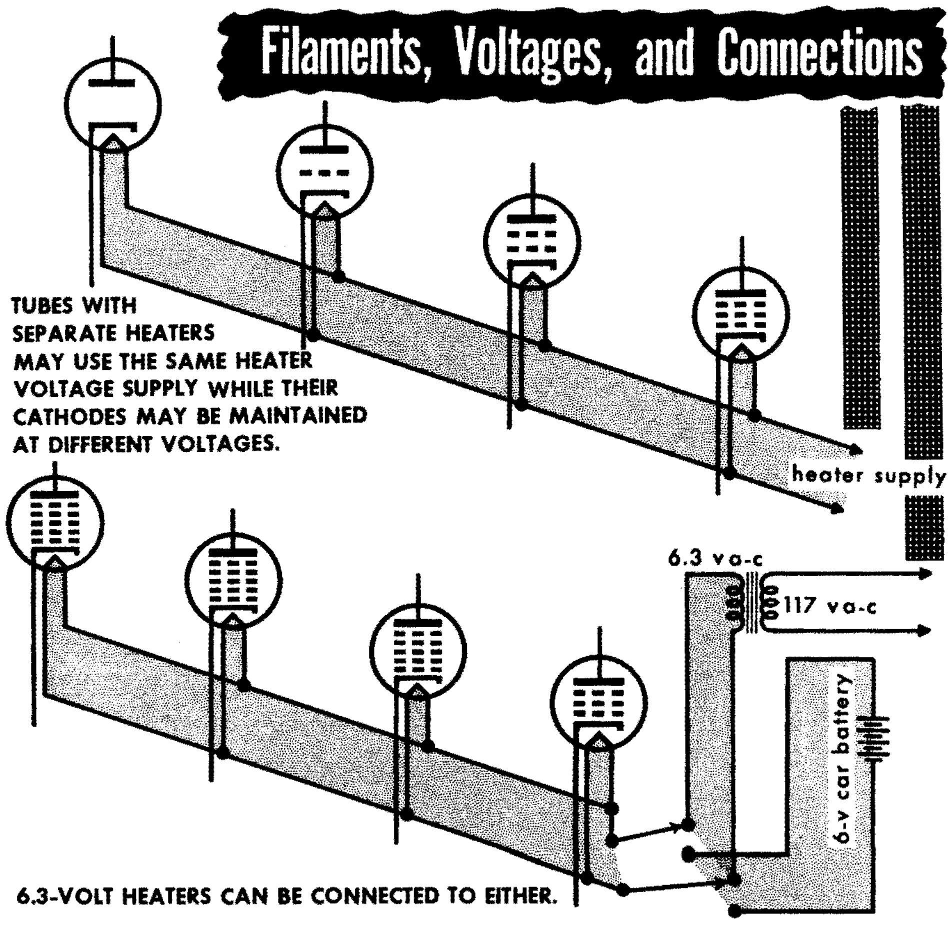
When audio voltages are applied to the grid of a pentode tube, screen current fluctuates as well as the plate current. The screen potential, however, must be held constant if we are to achieve the best operation from the tube, because a constant screen potential enables the plate current and plate voltage fluctuations to be almost independent of each other. If there were only a resistor between the supply voltage and the screen, screen *current* fluctuations would also be accompanied by some screen *voltage* fluctuations, and the plate voltage fluctuation would be affected by the screen voltage fluctuation.

To avoid this, a fairly large capacitor whose charge does not have time to change during the audio fluctuation is connected between the screen and ground. A 0.1-microfarad or smaller capacitor is usually quite large enough for this purpose, because screen current and its fluctuations are small.

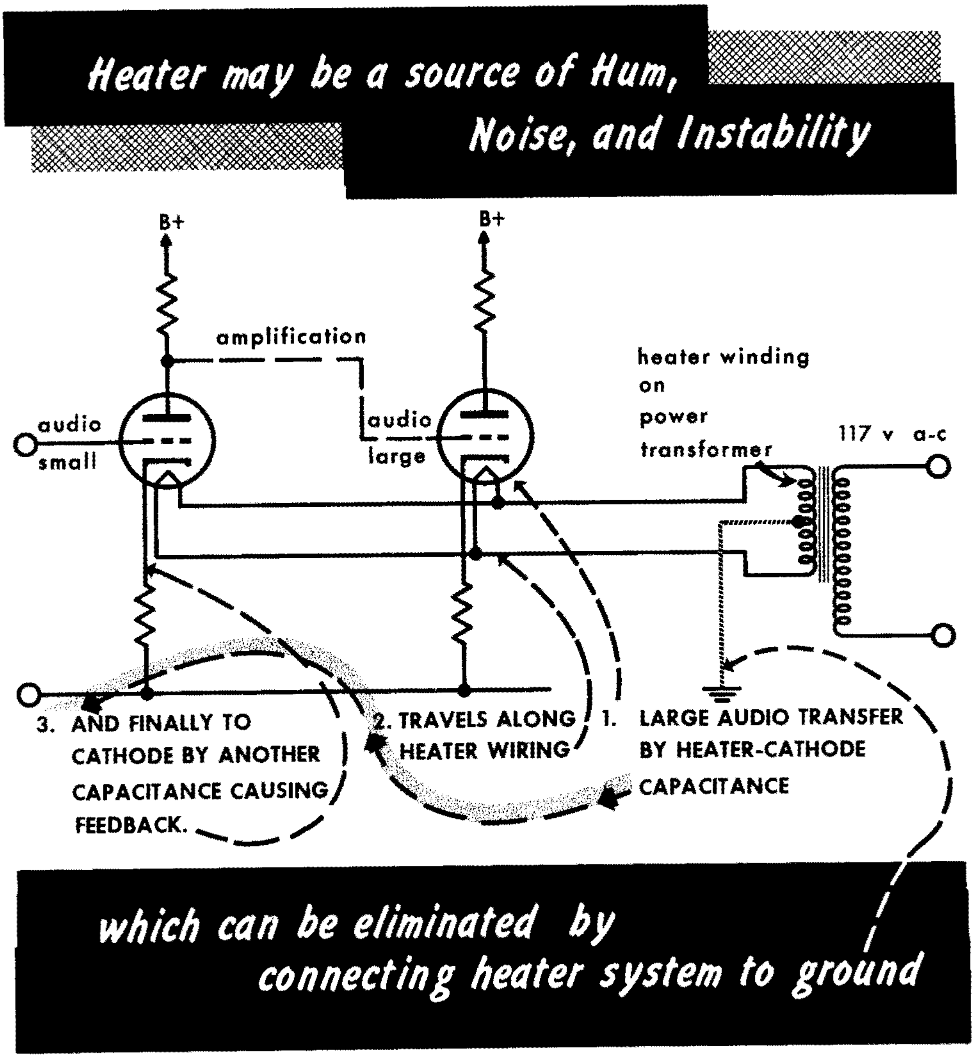
Filament Connections and Voltages

Thus far, we have not discussed how we get electrons into the tube. This requires heating of either filaments or heaters. Most modern tubes employ a cathode with a separate heater. This is a considerable help in the construction of a complete amplifier, because it allows the cathode to be biased to any suitable voltage, while permitting all of the heaters in the different tubes to run from the same supply in any convenient manner.

Usually, with a-c-operated amplifiers, all of the heaters are connected in parallel to a winding on the power transformer. The potential across this winding is chosen to maintain the correct temperature in the tube. This potential is usually 6.3 volts, because this particular voltage happens to coincide with the battery voltage at one time used on almost all automobiles. This heater voltage became standard, because it was then possible to operate the tubes alternately from a battery or a 6.3-volt winding on a transformer.



Eliminating Hum

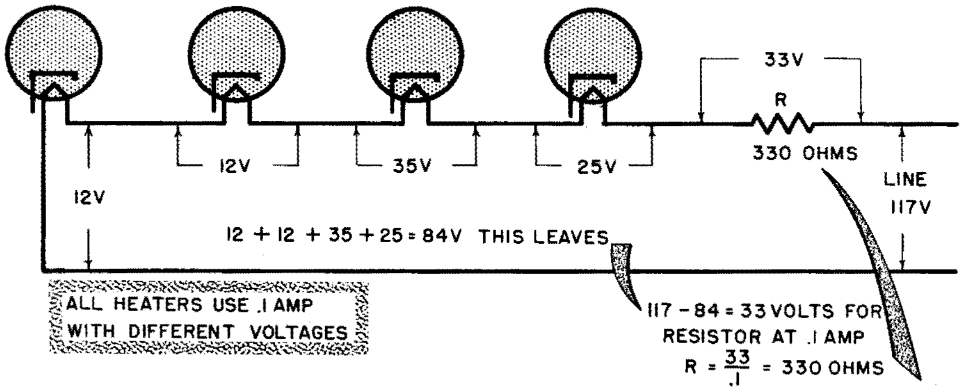


The heaters are a possible source of hum, noise, and instability in the amplifier, particularly at high frequencies. Owing to the capacitance between cathode and heater, audio can be transferred from the cathode in an output tube to the heater wiring and from there to the cathode of the input stage, which would cause oscillation. The remedy for these troubles is to connect the heater wiring to a ground point, so that any capacitive transfer from the output-stage cathode to the heater wiring is immediately conducted to ground. Then, although minute currents may flow between the cathode and heater wiring, there will be no *audio* voltage corresponding to them.

Eliminating the Power Transformer

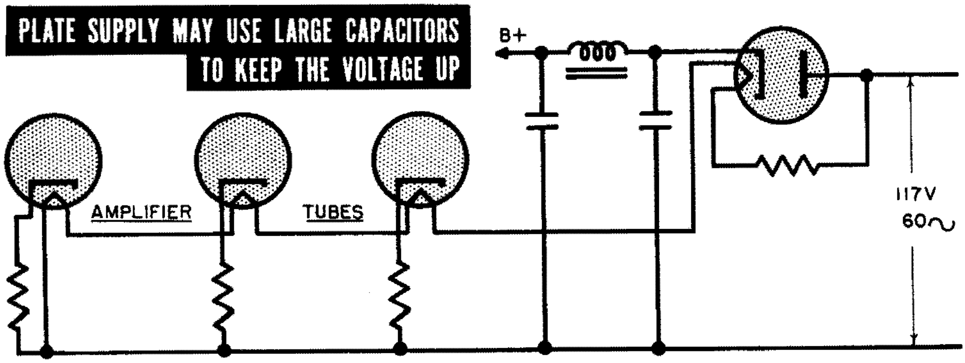
Another system of supply was developed for use where the power source might be either a-c or d-c. (There are parts of the country where d-c power is still provided.) This method has the advantage that a power transformer is not required, which results in some saving in cost. (For this reason, similar circuits are also applied where there is no intention of using the equipment on d-c. In this arrangement, the heaters are wired in series. The tubes are designed to take the same heater *current* instead of the same heater voltage. (Typical heater *current* lines operate at 0.1 ampere or 0.15 ampere.) If the total voltage drop across all of the heaters in series adds up to, say, 84 volts, a series resistor will be used to give the required 117-volt drop total.

HEATERS ARE CONNECTED IN SERIES



In this case, the plate supply is usually half-wave rectified, using very large capacitors to keep the voltage up. In addition, special tubes are used that operate satisfactorily at relatively low plate voltages, such as 150 volts.

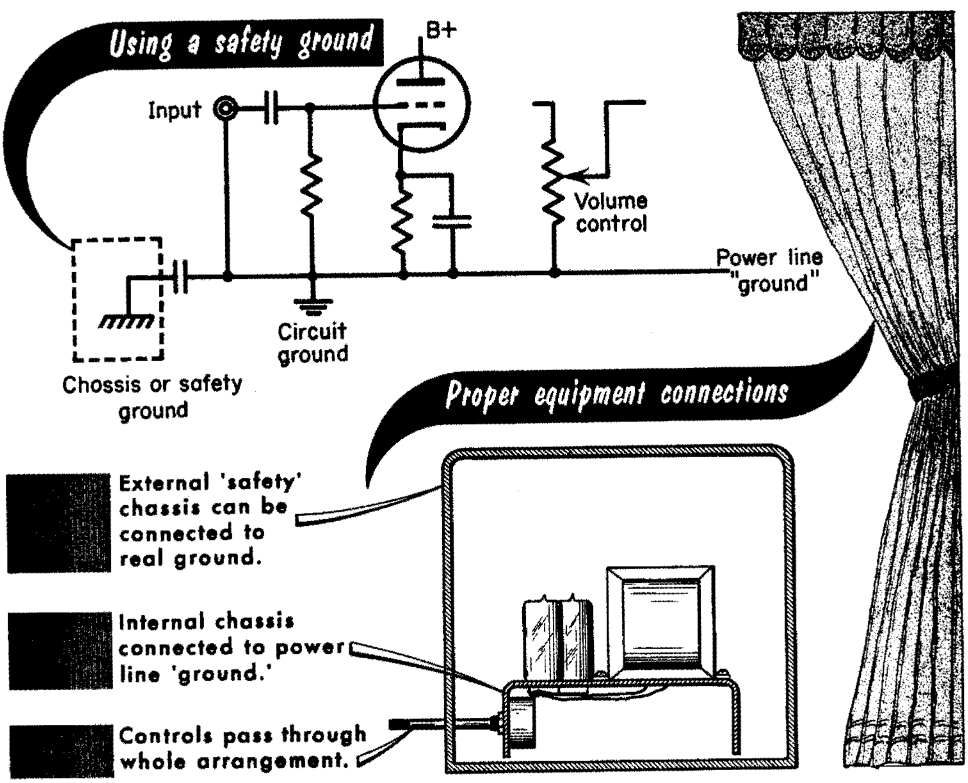
PLATE SUPPLY MAY USE LARGE CAPACITORS TO KEEP THE VOLTAGE UP



Ground Problems

A special ground connection is needed to make transformerless equipment safe. The supply-circuit "ground" is connected to one side of the supply, which may be the "live" side, if the power plug is not put in the right way around. This is a hazard because metal parts of the amplifier become capable of giving shocks. For this reason, two ground points must be provided. The circuit ground is wired so that it is not accessible at any point. A very small capacitor is connected between this high-voltage negative line and the metal parts that are accessible and thus liable to be touched. (This capacitor has to have a very high breakdown rating if the metal parts are to remain isolated from the line voltage.) The use of this capacitor insures the user that touching a metal part of the amplifier can give him only a slight tingle, (due to the microscopic currents that pass through the capacitor). To avoid introducing hum, the chassis has to be carefully arranged so that any accessible metal parts are not close to the low-level audio wiring.

This means the chassis construction must be double. All of the amplifier circuit is arranged to be within a chassis that is connected to supply negative, while a second chassis, insulated from the first, encloses it, eliminating the risk of shock.

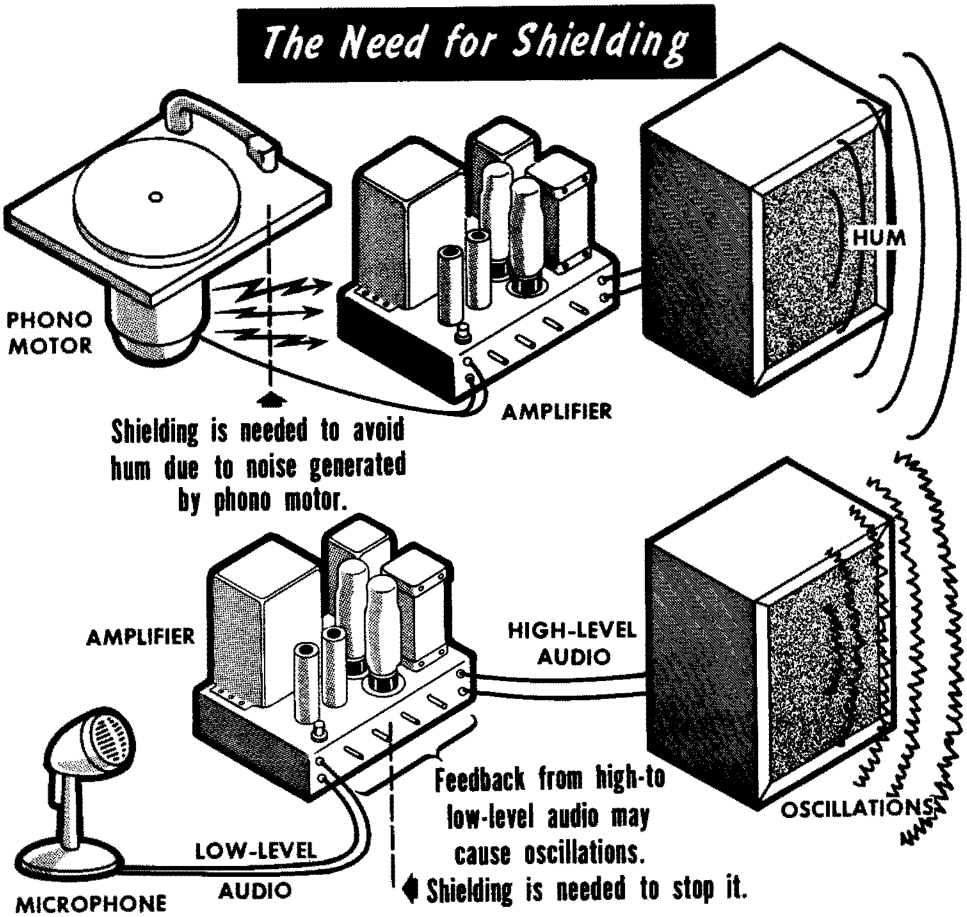


QUESTIONS AND PROBLEMS

1. Why is an a-c power supply more convenient for many purposes than a d-c? What kinds of rectifier can be used for converting a-c to d-c for operating tubes or transistors?
2. Explain how (a) a choke and (b) a capacitor may be used to smooth out the large fluctuations in rectified a-c.
3. Distinguish between the inherent characteristics of smoothing filters utilizing (a) capacitor input and (b) choke input. Also, show how a "swinging choke" combines features of both and for what purpose.
4. Additional filtering is often provided for various supply points in an amplifier, but is usually called "decoupling." Explain the significance of this term.
5. Distinguish between self or automatic bias and fixed bias for tube operation. In what way is the former automatic?
6. Why should a bypass capacitor be used with an automatic bias circuit? Under what circumstances does automatic bias work better without this bypass, and why?
7. What is crossover distortion, and what causes it? How can it be avoided?
8. What two methods of supplying screens are used? If a screen is to be operated at 150 volts from a 250-volt supply, what resistors should be used so that change of screen current from 0.5 milliampere to 1.5 milliamperes only changes the voltage from 155 to 145 volts?
9. Why is a capacitor necessary with a screen supply circuit? Explain why a cathode bypass capacitor may be as much as 50 microfarads or more, while a screen capacitor can be only about 0.1 microfarad.
10. What two principal methods of connection are used for tube heaters?
11. For what reasons may equipment be designed to work without a transformer in the power supply? What particular precautions are necessary in this kind of equipment?
12. What remedies would you try if an amplifier or other piece of audio equipment developed hum?

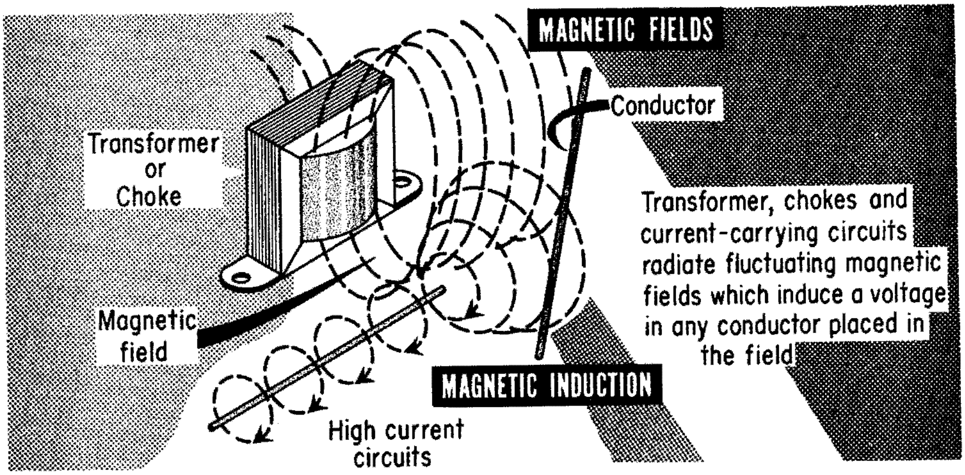
The Need for Shielding

One of the problems of audio amplifiers is the fact that audio voltages and currents are often so small that the circuits carrying them have to be protected against unwanted voltages and currents induced from other circuits. There are three kinds of unwanted induction: power, crosstalk, and feedback.

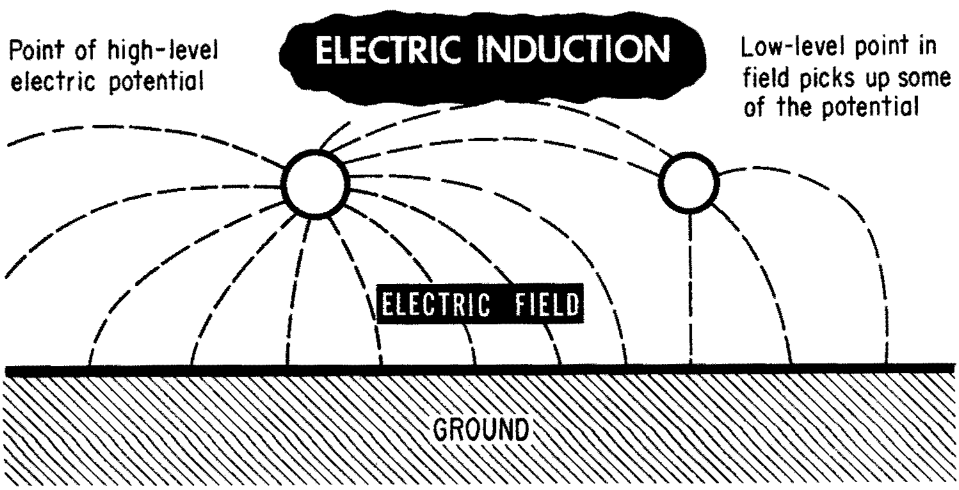


All kinds of power components and circuits radiate at power frequencies (mostly 60 and 120 cycles). If this radiation is induced into a low-level audio circuit, hum will result. The second kind of induction is that of high-level circuit, if the two are in close proximity to each other. (This interference is known as *crosstalk.*) If the *same* audio at high level is induced in low-level circuit, it is the same as undesired feedback, which can cause instability or oscillation.

The Need for Shielding (contd.)

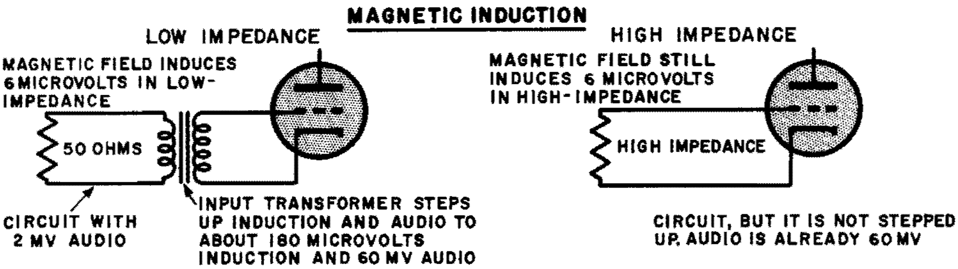


The solution for all of these troubles is shielding to protect the low-level circuit against pickup due to magnetic or electric fields. Any transformer will radiate some *magnetic* field. Audio transformers radiate a magnetic field that contains audio, and power transformers radiate at power frequencies. (Chokes can also radiate at these frequencies.) These magnetic fields can induce voltage in low-level circuits by magnetic induction. This works by causing a voltage to be induced (in the same way that a transfer of energy occurs in a transformer) whenever the field changes or fluctuates. An electric field is caused by the presence of high voltages. Any high-level circuit carrying high audio or power voltages will radiate an electric field that can induce charges in a low-level audio circuit, producing pickup.



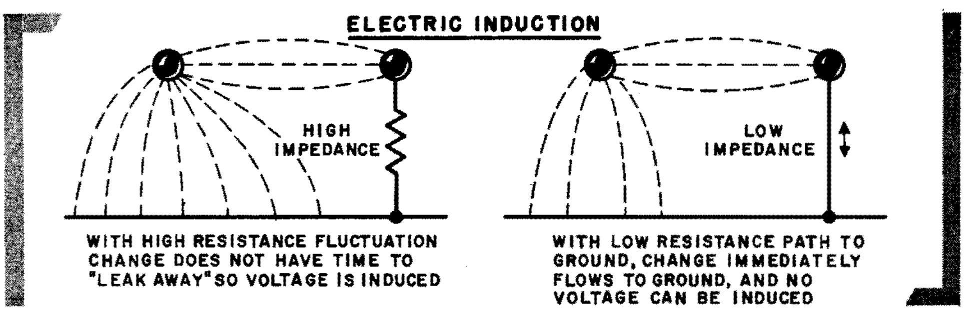
The Effect of Circuit Impedance

WITHOUT SHIELDING, MAGNETIC INDUCTION IS MORE TROUBLESOME IN LOW-IMPEDANCE THAN IN HIGH-IMPEDANCE CIRCUITS



Without the use of shielding, magnetic induction is usually more troublesome to low-impedance circuits than to high-impedance circuits, because a magnetic field will induce the same voltage in a loop regardless of its impedance. Suppose that the magnetic field induces 6 microvolts in a given circuit loop. If the circuit has a high-impedance, the audio voltage level may be 60 millivolts, which gives a margin of 80 db (a voltage ratio of 10,000:1). If the circuit has an impedance of only 50 ohms, however, the audio is likely to be about 2 millivolts. A transformer with a step-up ratio of 30:1 (an impedance ratio of almost 1000:1) could step up this voltage to the same 60 millivolts as that in the high-impedance circuit. As well as stepping up the 2 millivolts of audio, however, the transformer would step up the undesired 6 microvolts to about 180 microvolts. Hence, *with the same magnetic field*, the high-impedance circuit gives a margin of 80 db (10,000:1), while the low-impedance circuit only gives a margin of 50 db. (312:1).

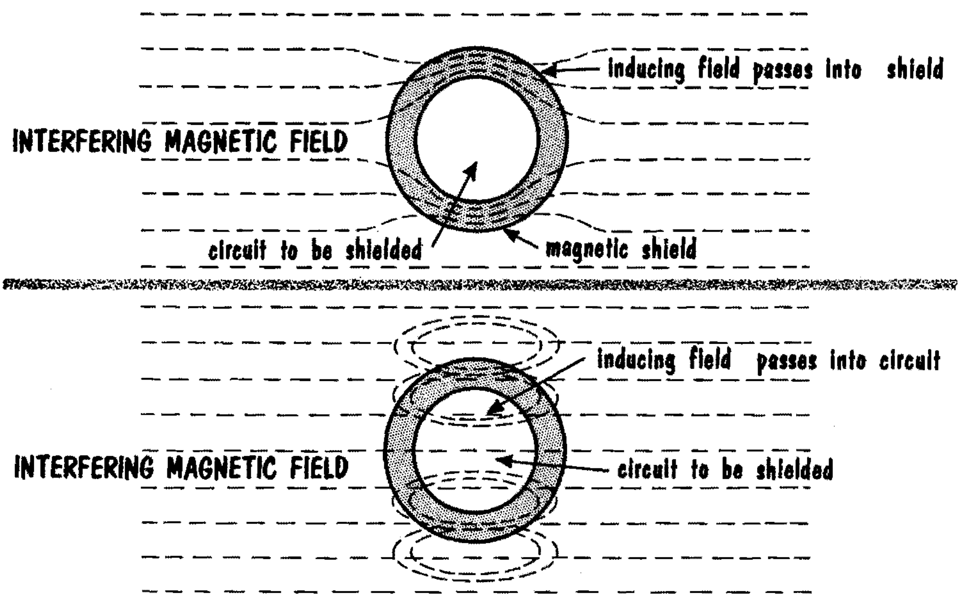
The opposite is true with electric induction. If the surface on which charge is induced by a nearby voltage has a high-impedance connection to ground, the charges will not have time to leak away and voltages will appear with them, causing interference. If, on the other hand, the surface is connected by a low-resistance path to ground, the charges will leak away rapidly without causing this interference. This means that low-impedance circuits are less susceptible to electric induction than high-impedance circuits.



Magnetic Shielding

There are three kinds of shielding: magnetic, electromagnetic, and electric. Magnetic shielding prevents the magnetic field that causes the induction from reaching the low-level circuit. The entire circuit or transformer is surrounded by the shield. The inducing field passes into the shield, around the circuit to be shielded and out at the other side. This reduces the induction inside the shield by a factor dependent upon the *permeability* of the magnetic material of which the shield is made, the thickness of this material that provides a magnetic conducting path, and the frequency of the magnetic fluctuation.

Magnetic Shielding

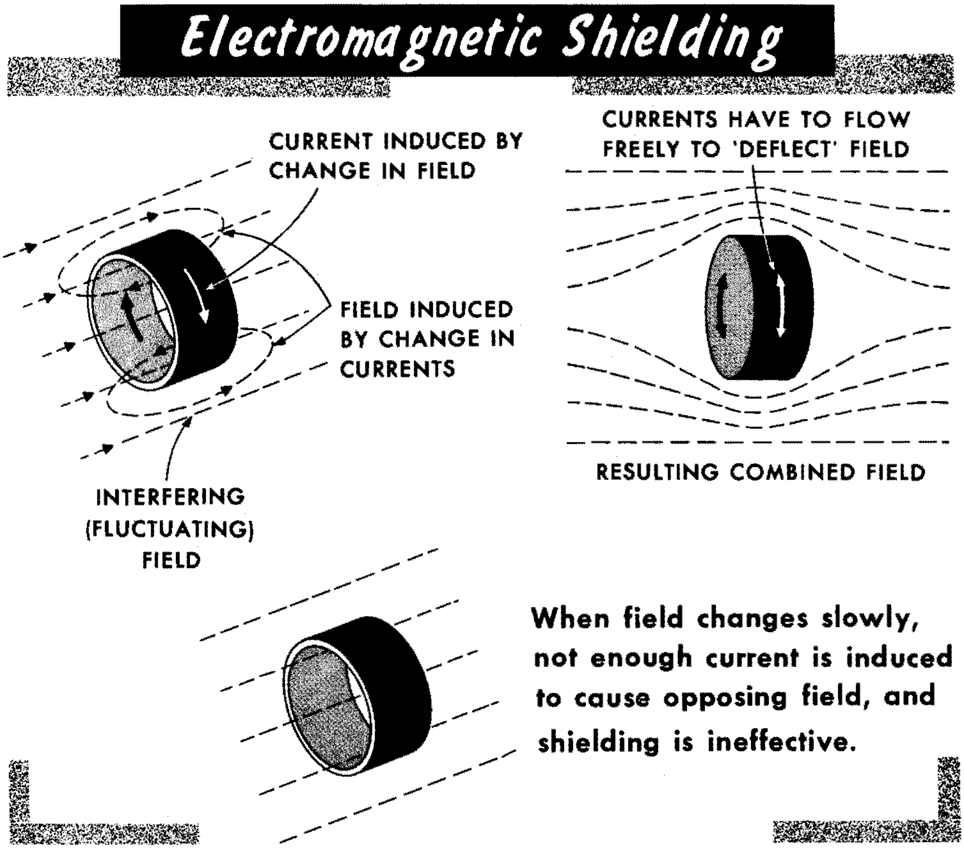


AT HIGHER FREQUENCIES, FIELD CHANGES TOO FAST FOR SHIELD TO WORK. SHIELD PRODUCES DELAYED REACTIONARY FIELD, WHICH DOES NOT ELIMINATE INDUCTION ANY MORE.

Magnetic shielding is most effective against steady magnetic fields and low-frequency induction, such as hum. At higher frequencies, it becomes less and less effective because the magnetism takes time to be induced in the material of the shield. The faster the inducing magnetic field fluctuates, the less effective the shield is in conveying the magnetic field around the shielded circuit.

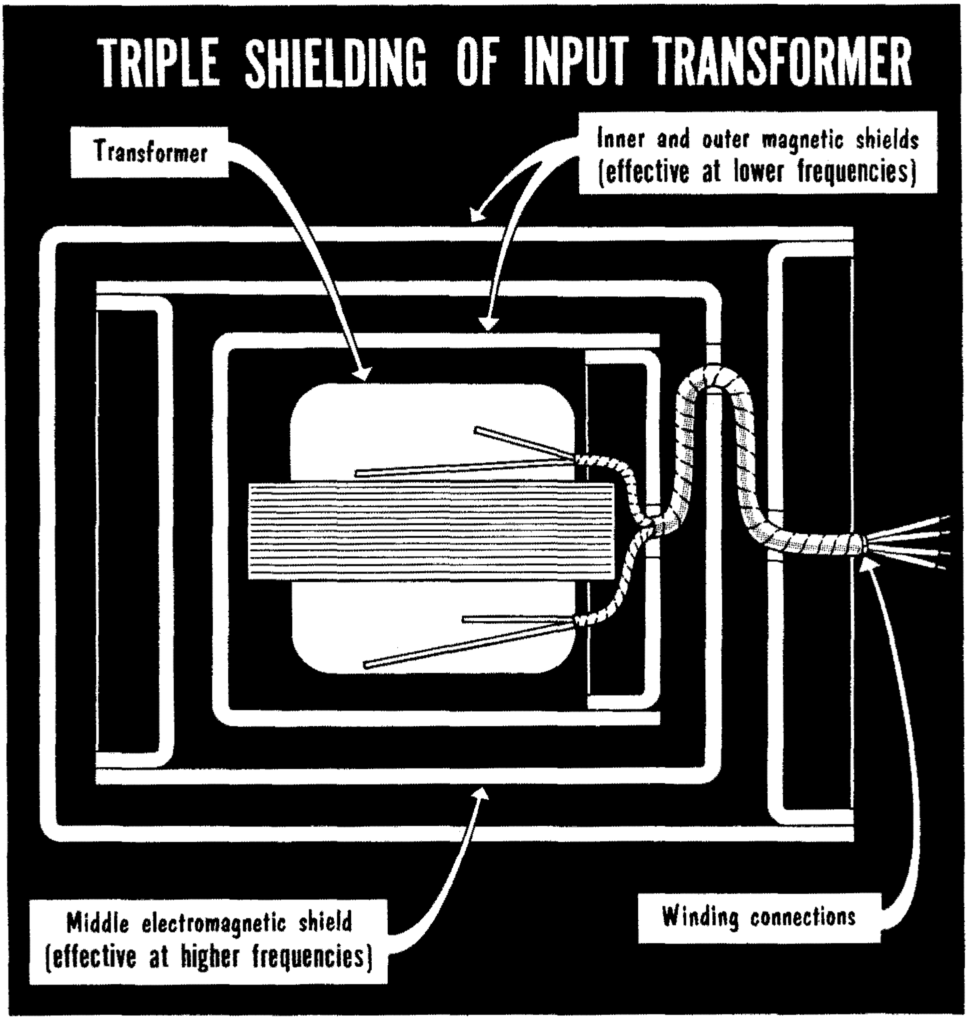
Electromagnetic Shielding

The second type of shielding uses a different principle, but is also effective against magnetic fields. Electromagnetic shielding does not use a magnetic material, but rather one that conducts electric current well, such as copper or aluminum. When the magnetic field causing the induction fluctuates, it causes current to flow in the shield. This current in the shield sets up its own magnetic field that opposes the original inducing magnetism, and the two fields tend to cancel inside the shield.



Outside the shield, the two fields are in the same direction and the combined magnetic field is increased. Therefore, this kind of shield has the effect of "pushing" the inducing field outward, instead of permitting it to reach the shielded circuits. Because it depends on fluctuating fields, this kind of shield is completely ineffective against steady magnetism. It is also comparatively ineffective against low-frequency fluctuation, and only becomes really effective at higher frequencies.

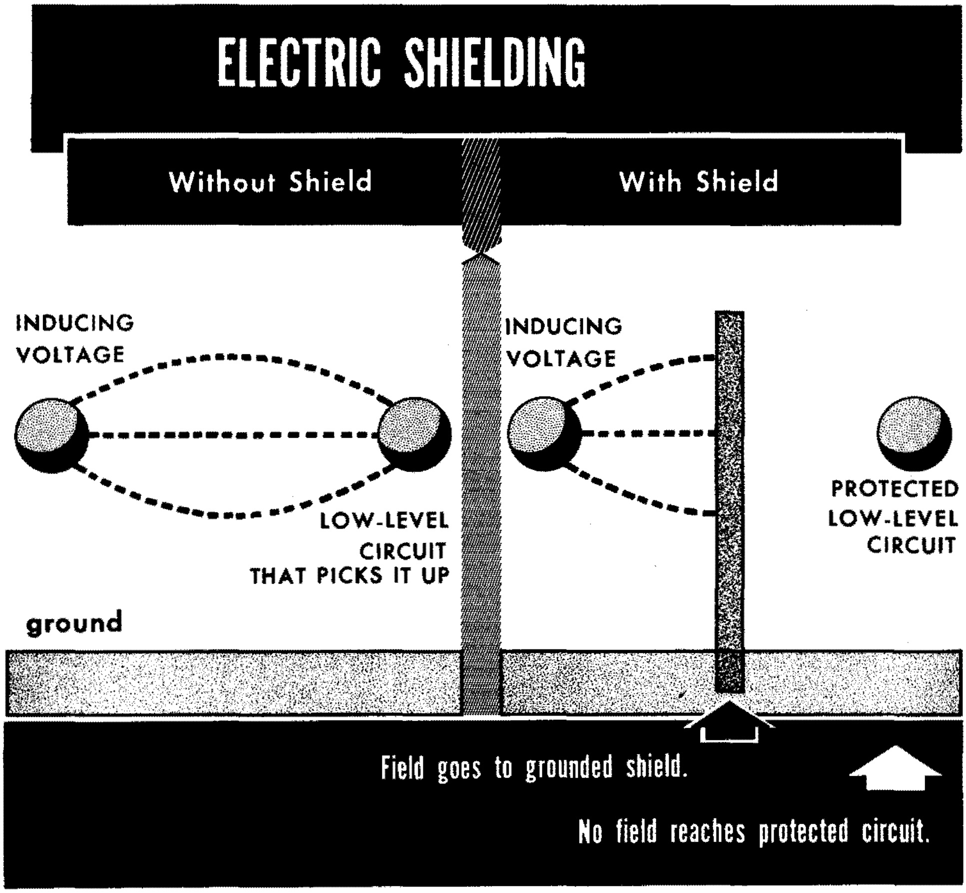
Electromagnetic Shielding (contd.)



With input transformers, which are particularly susceptible to magnetic fields because they have a magnetic core, the shielding often consists of an arrangement combining magnetic and electromagnetic shielding into one composite assembly. As well as providing increased protection, the combination makes the arrangement more effective over the entire frequency range. The magnetic shielding takes care of frequencies down to zero or d-c and begins to become ineffective at frequencies between 60 to 300 cycles. The electromagnetic shield, on the other hand, begins to become effective between 60 and 300 cycles, and the combined protection is effective for all frequencies.

Electric Shielding

The third kind of shielding is a protection against electric induction, commonly called electro*static* induction. (Electro*static* induction is really a misnomer; this induction is better called *electric* induction, because it depends upon the continuous fluctuation of the charges induced.)



Electric shielding consists of interposing a grounded shield between the interfering voltage and the low-level circuit that might pick up the electric field. The voltage induced is immediately carried off to ground, and the electric field is prevented from reaching the circuit that is shielded. Unlike either form of protection against magnetic fields, shielding against electric fields can be almost 100% effective. All that is necessary is to insure that no path is left through which the electric field can pass. For this reason, much more attention is given to the prevention of induction by *magnetic* fields.

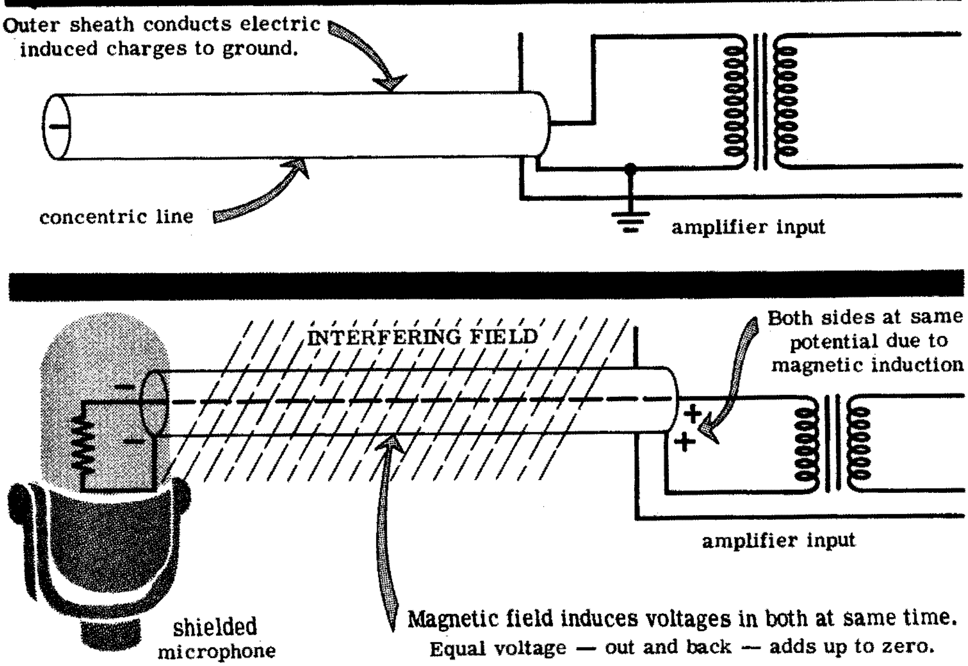
Coaxial Lines

Sometimes a connection of considerable length will be required between a microphone and its amplifier. Because the signal is at a very low level, we have to take precautions against possible unwanted pickup due to induction from magnetic or electric fields. In addition, we must transfer to the amplifier as much as possible of the audio voltage picked up by the microphone, to maintain the signal well above the noise level.

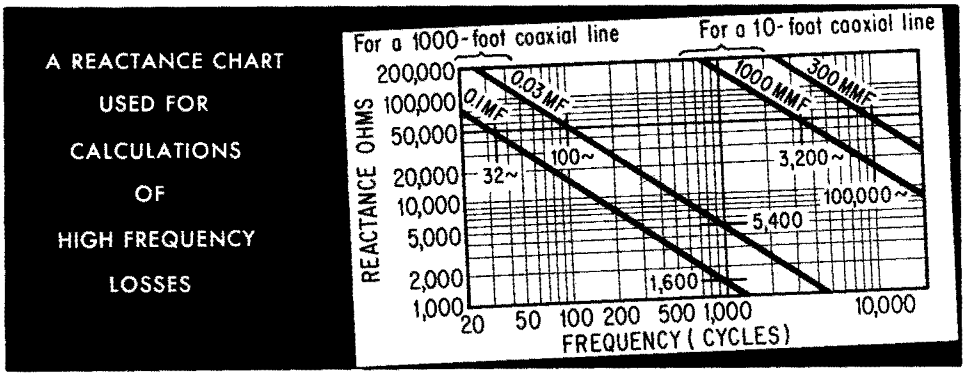
As a protection against electric pickup, a high-impedance connection starting from a 50,000-ohm or higher-impedance microphone would have to use a concentric or coaxial arrangement of the wires. The outside conductor (sleeve) would be connected to ground, so that any electric field reaching it would be conducted along this outer sleeve and go to ground at the amplifier without inducing any voltage on the inside wire. This effectively protects against electric pickup.

This type of line also protects against magnetic pickup, because any magnetic field will induce exactly the same voltages in both the inside wire and the outer sleeve. As the complete circuit from ground to the live side of the input *at the amplifier* consists of the entire line out to the microphone and back, the total voltage induced by the magnetic field will cancel out giving zero resultant induction.

Shielding a Line against Electric and Magnetic Fields

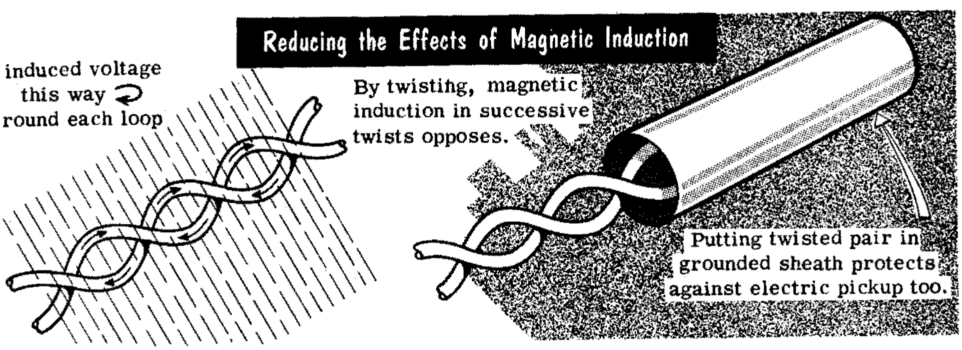


Coaxial Lines (contd.)

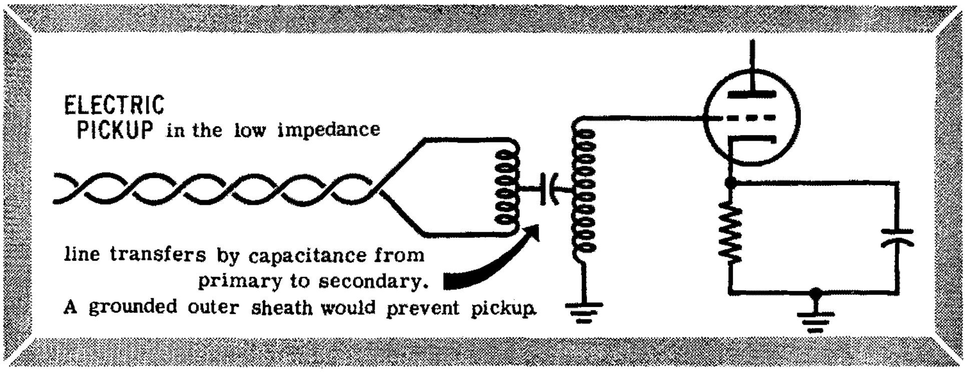


The reason why this type of line is not satisfactory is that the capacitance between the center wire and the sleeve is very considerable — usually between 30 and 100 micromicrofarads per foot. (This capacitance shunts the source and load impedances between which the line is usually connected.) A 1000-foot length of coaxial line will have a capacitance of 30,000 to 100,000 micromicrofarads, which on a 50,000 ohm circuit would have a reactance equal to the source impedance at a frequency between 32 and 100 cycles. This would cut the signal voltage fed to the lead by 70% (3 db). At 1000 cycles, the reactance would be about 1600 to 5400 ohms, resulting in a loss of 20 or 30 db, and the loss gets progressively more severe with higher frequencies. Even a 10-foot length of coaxial line, with a capacitance of from 300 to 1000 micromicrofarads, will have a reactance of 50,000 ohms at from 3200 to 10,000 cycles, which will cause an appreciable loss of signal at frequencies above this point.

Low-impedance circuits, as already mentioned, are more susceptible to the pickup of magnetic fields than electric fields. Use of a concentric line helps to shield them, and a twisted line improves matters by making the field induce voltages opposite directions in successive twists. The twisted line may also be shielded by a separate ground sleeve to minimize any electric pickup.



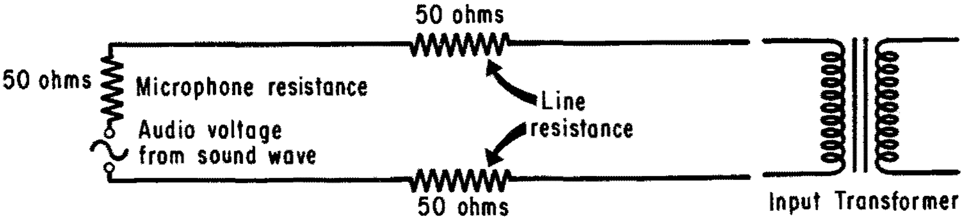
Low-Impedance Connection



Although electric pickup is not important to the low-impedance circuit itself, it *can* be transferred from the primary to the secondary of the transformer by the capacitance between windings. Use of an outside sheath as an electric shield will help prevent this.

The principal problem with a low-impedance (50-ohm or lower) line for long-distance connection is the resistance of the wire itself. Unless a very heavy gage is used, the line will have a resistance of about one ohm for every 50 feet. (A 5000-foot length would have a resistance of 100 ohms.)

RESISTANCE LOSSES DETERIORATE PERFORMANCE

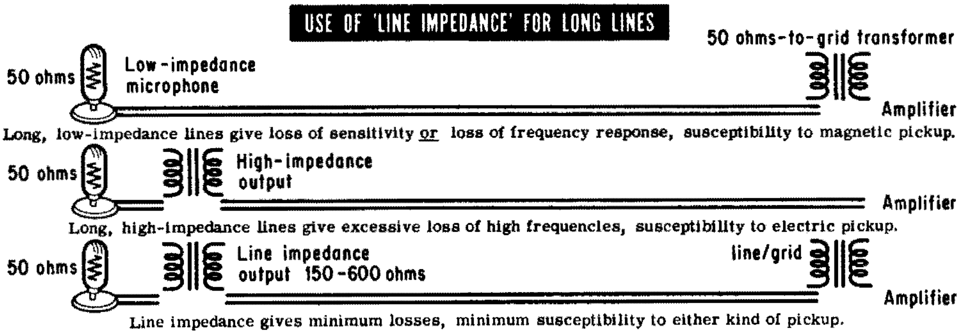


1. If input transformer primary is for 150 ohms, step-up will be 5 db less than for 50 ohms.

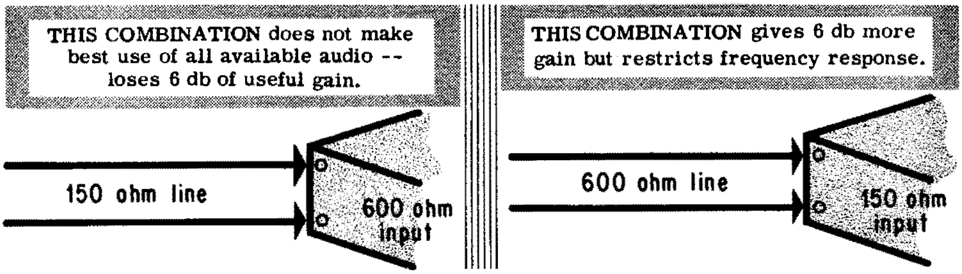
2. If input transformer primary is for 50 ohms, frequency response will be seriously depleted.

If the microphone impedance is 50 ohms and the resistance of the connecting wire is 100 ohms, the total input resistance to the amplifier will be 150 ohms. This means the input transformer can step up only from 150 ohms instead of from 50 ohms, as it could if the line were quite short. If all of this 150-ohm impedance were due to the microphone itself, the microphone would produce a correspondingly higher audio voltage; however, this is not the case. What we actually have is a 150-ohm source providing only the audio voltage that would be provided by a 50-ohm source.

Line Impedance

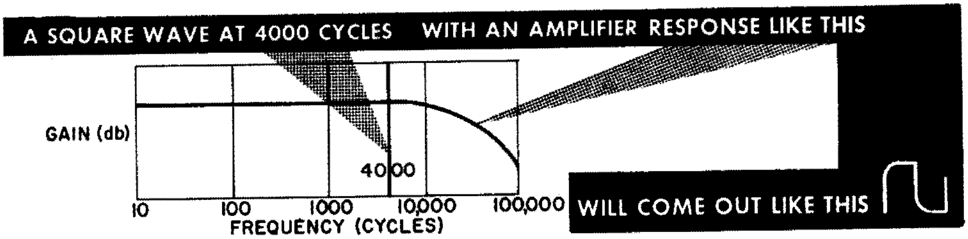


There is clearly a disadvantage to both high and low impedance for running long lines. For this reason, an intermediate impedance, in the region of 500 or 600 ohms, is usually chosen for making long-distance connections. It minimizes the possible effect of magnetic and electric induction, and avoids high-frequency losses that occur at high impedance and the *attenuation* due to line resistance that occurs in using low impedance. In input circuits, for example, a transformer is used so that the impedance of the microphone or pickup looks like 500 or 600 ohms at the transformer secondary. The amplifier has an input transformer that works correctly with a 500- or 600-ohm source. The impedance measured across the line between the two transformers is 500 or 600 ohms, and the line is said to work at an impedance of 500 or 600 ohms. Similar techniques are used with high-level output circuits as well. The impedance at which a line is being used is a characteristic of its termination, not of the line itself.



There is nothing particularly magical about one particular line impedance. The use of any middle-value impedance (150, 250, 500, or 600 ohms) merely minimizes the defects of either high or low impedances. It is, of course, good to use a consistent impedance in any particular system. Using a 150-ohm impedance and connecting it to a transformer designed for a 600-ohm impedance at the amplifier (or vice versa) will not make the best use of the available audio.

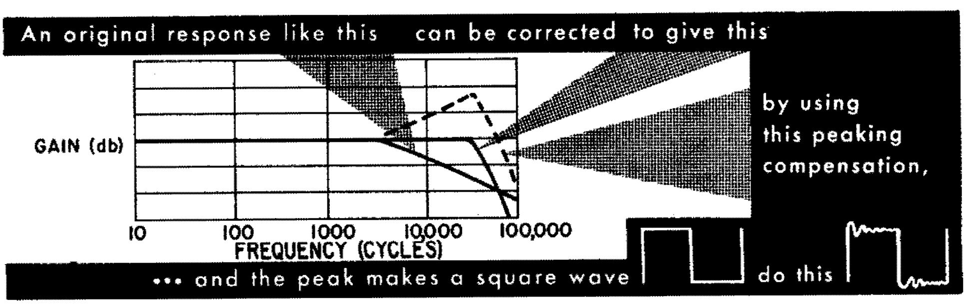
Ringing



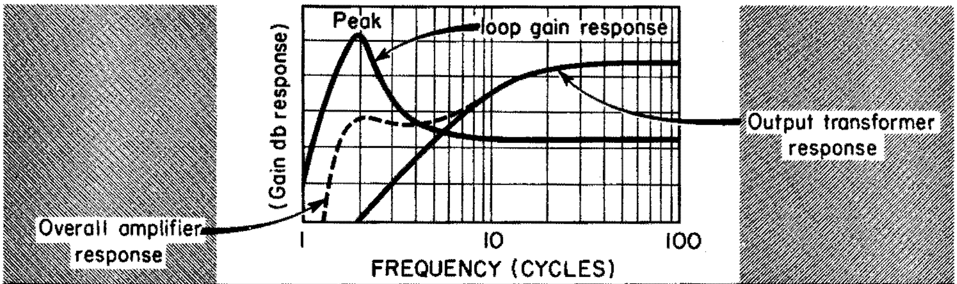
In Volume II, we briefly discussed the problem of *transient distortion.* The usual cause of distortion to square waves is the way in which amplification varies at high frequencies. If the amplifier's frequency response rolls off slowly, the corners of the square wave will be rounded. If, however, the response is uniform up to the highest frequency in which we are interested, say 10,000 or 20,000 cycles, and then at some frequency *after that* rises up to a peak, *ringing* occurs. The sudden shock given by the corner of a square wave excites this peak or resonance in the frequency response and the square wave loses its square corner, developing an oscillatory waveform every time this arrives.

It is not necessary to have a detectable peak at the high-frequency end of the response to get a similar effect. If the range of frequencies over which uniform response is maintained is extended by offsetting the high-frequency loss by peaking, the overall response curve appears quite flat, and then drops off sharply, without showing a peak. The fact that a peaking circuit has been used to extend the response produces the same effect as a peaked overall response curve — every time the corner of the square wave hits it, ringing results. For this reason, an amplifier with a sharp rolloff also produces ringing on square waves. This is true whether the sharp rolloff is produced by this kind of synthesis or, even more important, by feedback adjustments.

Ringing can cause other troubles in an amplifier in addition to unwanted oscillation on the corners of a square wave. For example, the oscillation may cause grid current to flow, when otherwise the amplifier would be well within its safe limit, and this, in turn, can initiate other troubles.

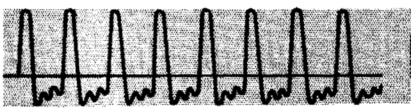


Bounce



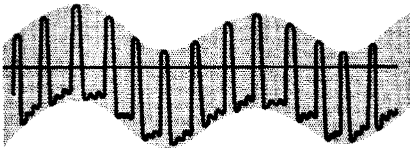
Asymmetrical waveforms can cause low-frequency ringing

Another variation of transient distortion occurs with feedback amplifiers. The low-frequency stability, as shown in the Nyquist diagram, may not have sufficient margin. This may not result in oscillation at the low frequency, but gives a peak in the *loop gain* response ($A\beta$) at a low frequency, such as 1 or 2 cycles. This peak may not show in the *overall* response of the amplifier, because the output transformer may produce sufficient loss to offset it. Nevertheless, asymmetrical waveforms (which have the effect of a sudden application of d-c) can cause this peak to give a low-frequency ringing *(bounce)* of 1 or 2 cycles per second. The bounce itself may not appear in the output, but it can result in quite serious intermodulation because some stages will not amplify uniformly at different parts of this high-amplitude very-low-frequency waveform.

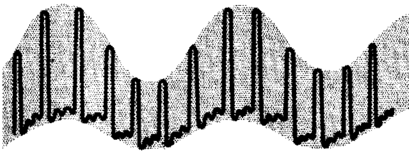


asymmetrical wave before amplification

normal way bias readjusts for asymmetry

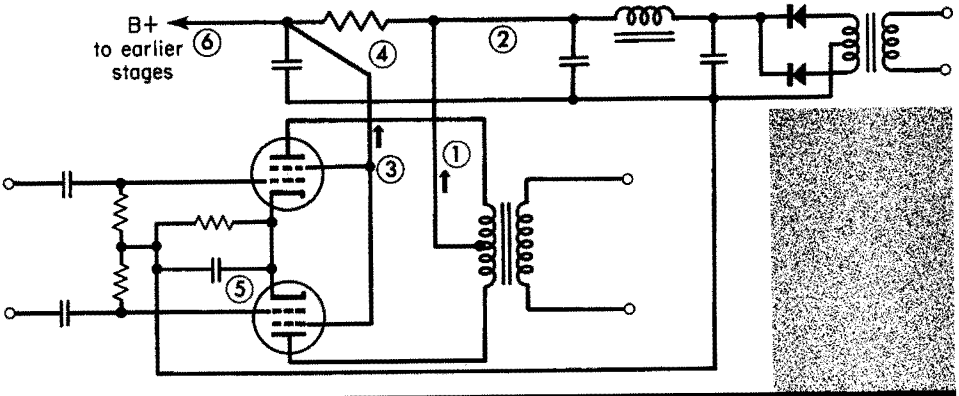


'bounce' due to low-frequency feedback ringing



form of distortion caused by 'bounce'

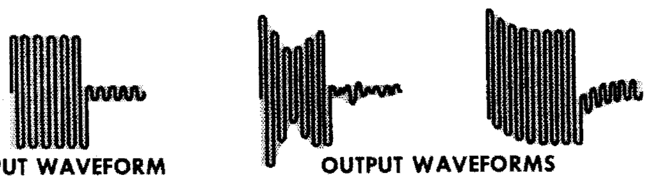
Due to Power Supply



Sudden changes of audio level in Output Stage cause

① plate current rises
② plate voltage drops
③ screen current rises
④ screen voltage drops
⑤ bias voltage increases
⑥ supply voltage to earlier stages drops

When audio level rises.



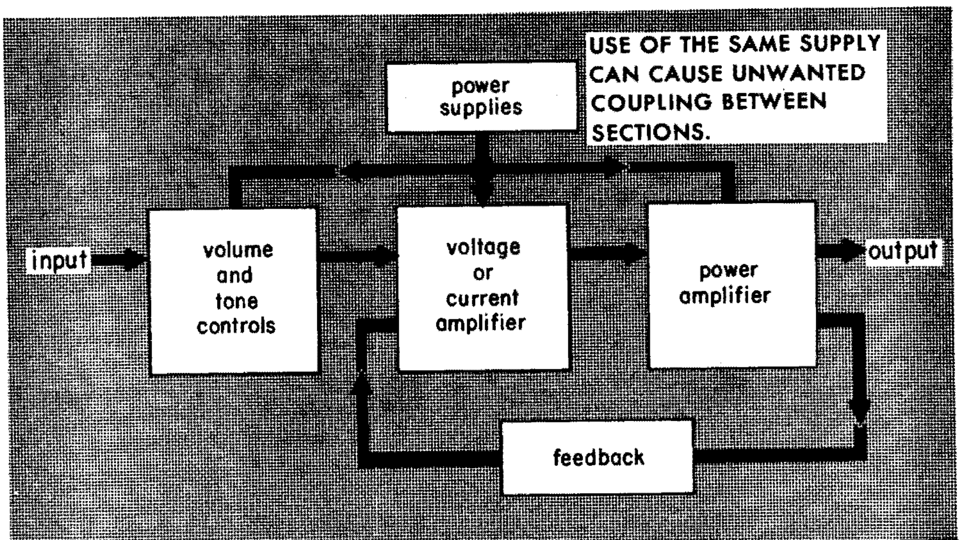
Two Kinds of Distortion Caused by Delayed Changes Voltage

The effect of a transient on the power supply may result in one type of transient distortion. A sudden burst of audio usually causes the current demand from the high-voltage supply to change, and with it the grid bias for the output stage and possibly the voltage supplied to the screens of any pentode tubes. This change takes time. When the change in audio is suddenly applied to the amplifier (as might happen when amplifying piano notes, for example), all the operating voltages and currents in the tubes are at their condition for minimum audio. When audio arrives at a high level, the amplifier starts to handle it, with the operating voltages and current as a moment before, but the current increases rapidly and the voltages begin to change. This change sometimes alters what happens in the amplifier and can introduce some kind of distortion before the amplifier gets to its new condition. The gain might also be altered in such a way that the quality of the applied transient (the piano note, or whatever it may be) is changed considerably.

The Complete Amplifier

The complete audio amplifier requires a lot of "putting together." From the designer's viewpoint, first is needed a power output stage to deliver whatever power is required, with satisfactory performance. To make this work, it has two requirements: adequate audio input to control it and the various power supplies to operate the tubes or transistors. If the current or voltage requirement of the output circuit varies according to the audio input level, the power supply circuit must accommodate this variation properly, without changing the voltage or current more than can be permitted.

To get the audio input, voltage or current amplification (according to whether you use tubes or transistors) is needed so that the small input from a pickup or microphone delivers enough to drive the power stage. Then feedback will be added to get the best possible performance from this combination. In considering feedback, do not forget that supplying the different stages from a common power supply can result in output circuit audio fluctuations being fed back to input stages, which are more sensitive. This can cause instability if not properly controlled.

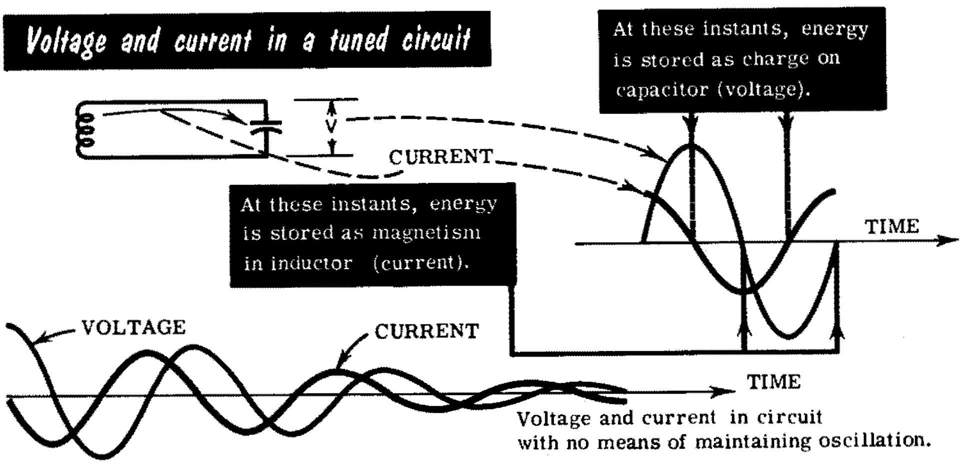


These are the Main Building Blocks for a Typical Audio Amplifier

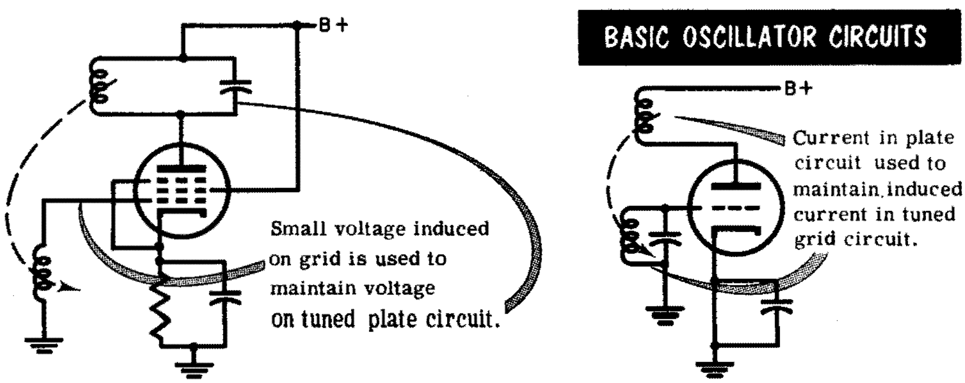
Finally, you need certain controls, volume, and tone, as well as equalization. These have to be put outside the main feedback loop, although they may use some feedback of their own. But if you make the mistake of putting feedback around a volume or tone control, the feedback will carefully undo all the effect of the control! It holds the gain steady, and has no means of knowing that you *meant* it to change.

L-C Oscillators

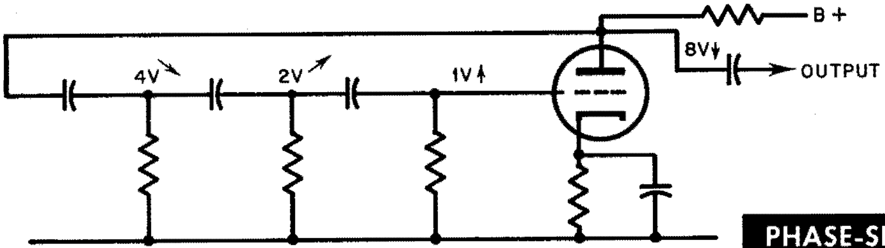
As well as amplifiers, to make audio waveforms larger, we also need oscillators to produce audio waveforms. They may be needed for testing, for electronic musical instruments, or for the high-frequency bias and erase in tape recording. In addition, they supply control frequencies needed for some kinds of automatic operation. In audio work, the exact waveform of the oscillation is important.



There are three basic kinds of oscillators. One uses an inductor and capacitor in parallel to produce a tuned or resonant circuit in which energy passes between the inductor and capacitor at the resonant frequency of the circuit. All that is needed to use this circuit in an oscillator is to feed back enough energy to make up for that lost in every transfer due to the losses in the circuit. This is usually achieved by having the tuned circuit in either the plate or grid circuit of a tube and using the amplification of the tube to feed back some of the energy in the correct phase to maintain the oscillation.



Phase-Shift and Positive Feedback Oscillators

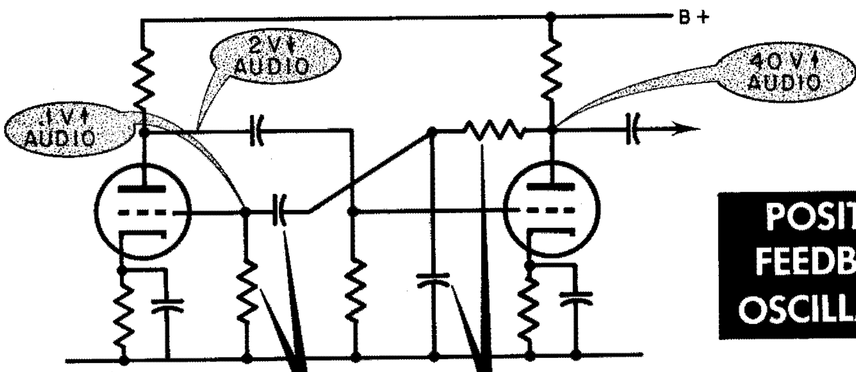


FREQUENCY OF OSCILLATION IS DETERMINED BY RESISTANCE AND CAPACITANCE VALUES

PHASE-SHIFT OSCILLATOR

The other two types of oscillators do not use inductors because inductors can lead to various difficulties. They are, for example, likely to pick up magnetic induction, which can cause hum, or to pick up unwanted audio. The first of these oscillators uses a progressive phase shift to take advantage of the oscillatory condition of a feedback amplifier. The combination of successive resistors and capacitors are arranged to produce the phase reversal with a gain of 1 at a frequency that can be specifically controlled. This frequency can be adjusted by varying either the capacitors or the resistors.

The third type of oscillator uses positive feedback. The positive feedback is sufficient to cause oscillation at just one frequency. An arrangement of resistors with capacitors in series produces a loss of low frequencies. A capacitor in parallel produces a loss of high frequencies. The correct combination produces a maximum transfer of the positive feedback at one frequency with no phase shift. The amplifier oscillates at this frequency because the feedback is smaller at all other frequencies.



POSITIVE FEEDBACK OSCILLATOR

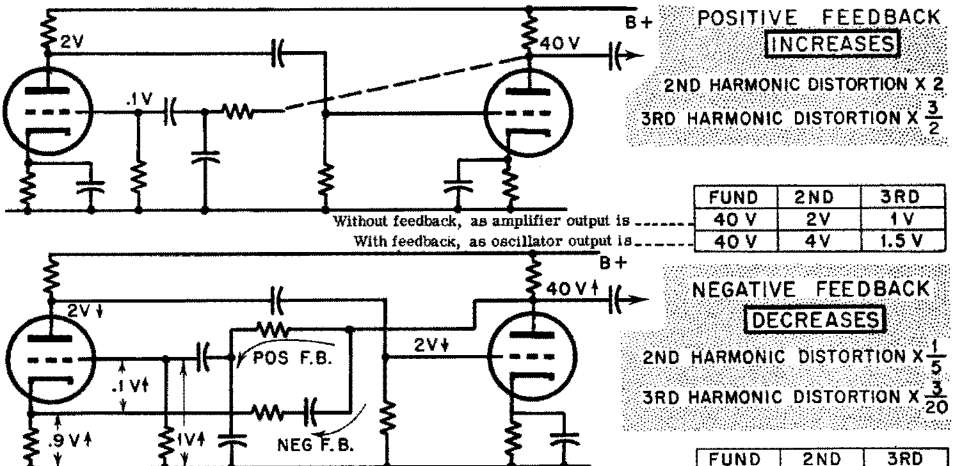
This resistor and capacitor prevent effective feedback at lower frequencies.

This resistor and capacitor prevent effective feedback at higher frequencies.

These four components control oscillation frequency because there is maximum transfer at a given frequency.

Impoving the Output

The waveform of the positive-feedback oscillator can be improved by adding negative feedback. Positive feedback that provides 100% of the required input at the one frequency will provide at least half as much voltage at double this frequency. If the amplifier produces any non-linearity that would cause second harmonic distortion, there will be at least a 50% positive feedback of the second harmonic as well as 100% feedback of the fundamental. If the tube causes third-harmonic distortion, there will be at least 33% positive feedback for it as well. For this reason, the positive-feedback oscillator produces considerably more harmonic distortion than either of the other types.



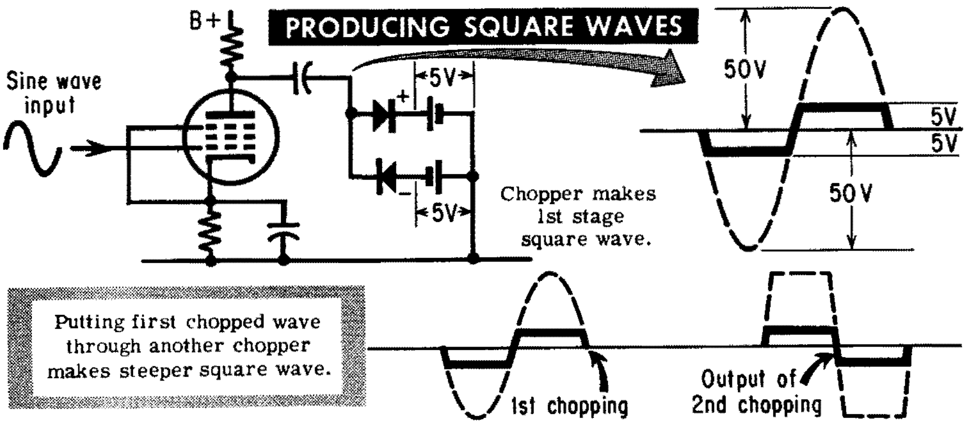
	FUND	2ND	3RD
Without feedback, as amplifier output is	40 V	2V	1 V
With feedback, as oscillator output is	40 V	4V	1.5 V

	FUND	2ND	3RD
Without positive or negative feedback, as amplifier output is	40 V	2 V	1 V
With both, as shown, output is	40 V	.4 V	.15 V

Negative feedback with positive gives lower distortion

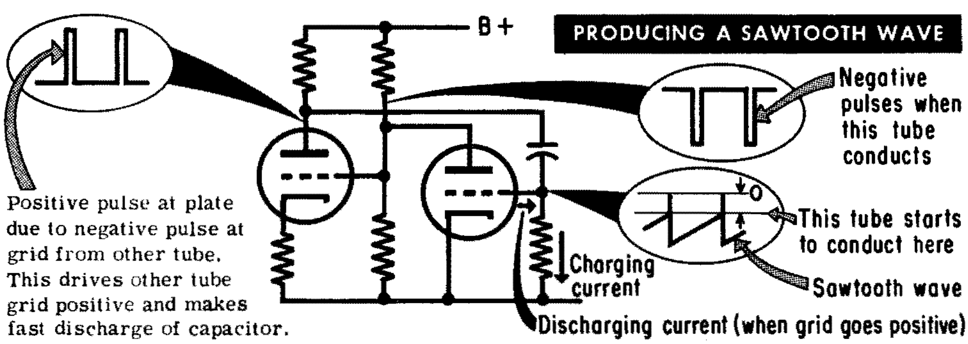
By combining frequency-selective positive feedback with *negative* feedback that is not frequency-selective, distortion can be reduced. Suppose that the negative feedback produces an Aβ of 9, while the positive feedback produces an Aβ of 10. The net positive feedback is (10 − 9) or 1 — still 100%, to allow oscillation. At double the fundamental frequency the positive Aβ is only 5, while the negative feedback will still be 9. This means that the second harmonic will have a predominant negative feedback of (9 − 5) or 4. Without the negative feedback, there is 50% positive feedback of the second harmonic, resulting in a 6-db increase. The combination of negative with positive feedback results in a 14-db negative feedback, reducing the second harmonic distortion of the oscillator to 1/5, and higher harmonics even more. Thus, the positive and negative feedback arrangement inherently produces better sine waves than any other kind of audio oscillator.

Oscillator Waveforms



Audio requires waveforms other than sine waves. Square waves used in transient tests can be produced simply by amplifying sine waves in a circuit using diodes that start to conduct when the voltage reaches a specified point. This kind of circuit chops off the top and bottom of the sine wave, making something very much like a square waveform. If 9/10 of the sine wave is chopped off, the slope of the line joining the horizontal sections is very nearly vertical. The sides can be made even steeper by amplifying this wave again, and again chopping off 9/10 of it.

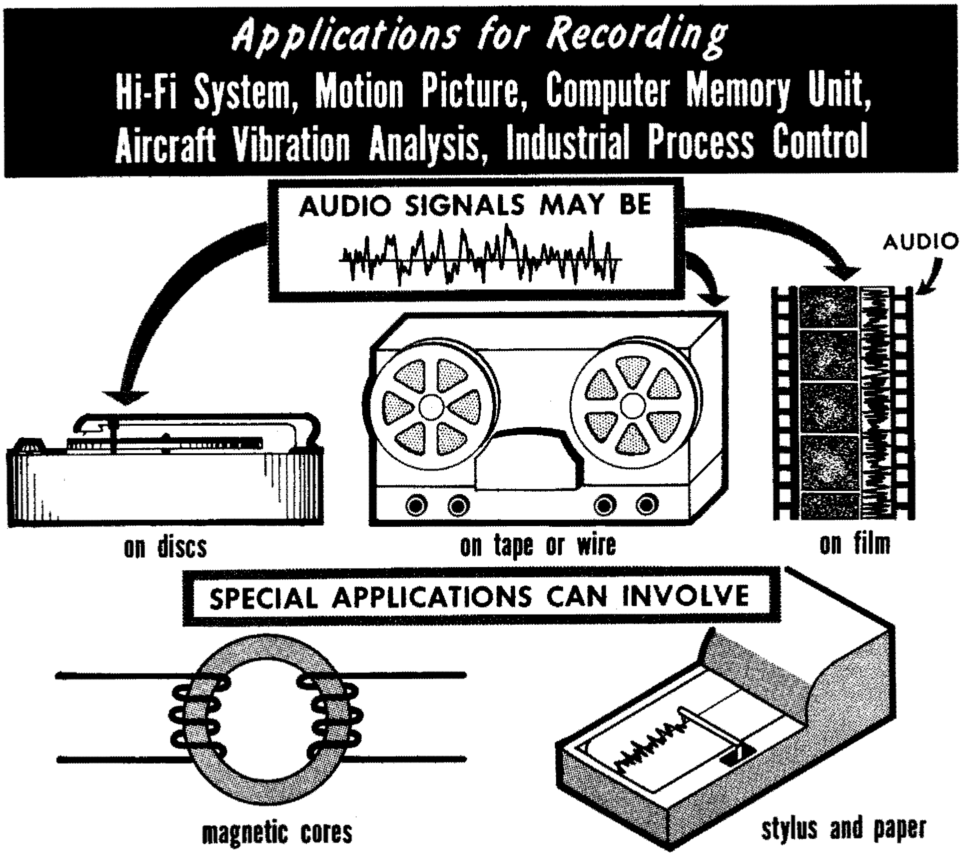
Sawtooth waves are made by a variety of means. A simple one is a circuit in which a steady current flows into a capacitor until the voltage across it reaches a specified point. This starts a quick feedback action through a couple of tubes that rapidly discharges the capacitor to its original level, from which it starts the charging sequence all over again. The sawtooth waveform is used for the sweep voltage in oscilloscopes and in electronic musical instruments as a basis for the variety of musical tones. A sawtooth wave has a very useful combination of fundamental with *all* of its harmonics, whereas the square wave possesses only the odd harmonics.



QUESTIONS AND PROBLEMS

1. Shielding prevents what three kinds of induction? In what ways can this induction occur?
2. What effect does circuit impedance have on the relative importance of the different ways induction occurs? Explain.
3. What are the essential features and characteristics of (a) magnetic, (b) electromagnetic, and (c) electric shielding?
4. Explain the advantages and limitations of coaxial lines.
5. Give the deficiencies associated with using lines at high and low impedance that lead to choice of a medium line impedance. What are common values of line impedance?
6. How can a condition near to instability (oscillation) cause distortion due to transient effects? Explain this possibility with reference to both low- and high-frequency near-instability.
7. Describe a group of transient effects in amplifier performance that is essentially caused by power supply features.
8. Name four major types of oscillator designed to produce sine waves. Explain the principle of operation of each. Which give the best waveform? Show why.
9. What other types of waveform are required from oscillators, and for what purposes?
10. A tube used in an oscillator circuit amplifies with the production of 5% harmonic. What positive and negative loop gain figures must be used so the output from this oscillator is 0.1%?
11. What is the essential difference between the phase-shift and positive feedback types of oscillator?
12. What difference is there, from a musical viewpoint, between a square wave and a sawtooth?

Uses for Recordings

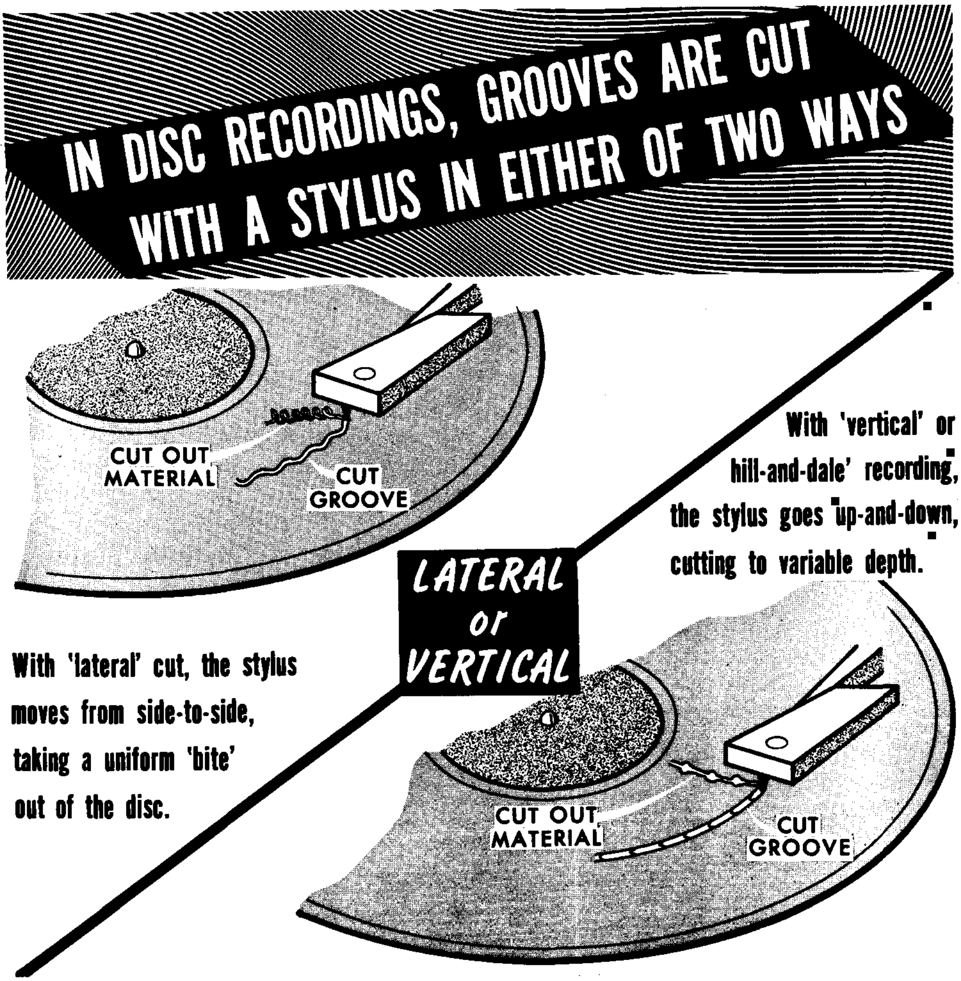


Recording plays an important part, not only in audio for high-fidelity purposes, but also in the many other applications for which modern audio equipment is used. It can be regarded as a storage medium in which long sequences of audio waveforms can be indefinitely preserved for reproducing at a later date.

In the case of high fidelity, this audio waveform sequence may be a complete musical performance. For computer and other industrial applications, the sequence may be any audio waveform combination representing information. It may be a record of the vibrations that occur in different parts of a supersonic missile in flight or a record of the progress of a certain industrial process that requires careful comparison of the results of the process in successive hour periods. By recording the data during a complete hour and then comparing the result with the corresponding data exactly one hour later, the process could be continually controlled. A variety of media are used for storing audio material: disc, tape, wire, and film.

Disc Recording Techniques

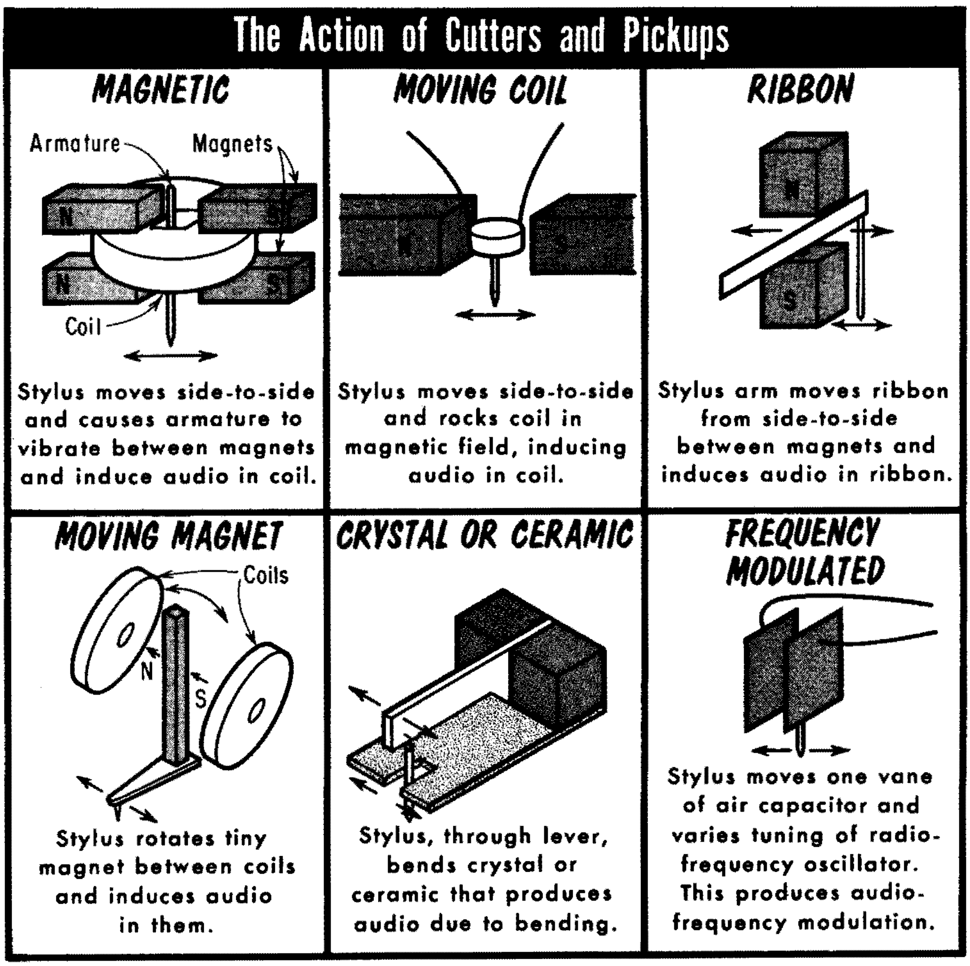
In disc recording, grooves are cut on a smooth surface by means of a cutting stylus. Later on the pickup stylus will follow the same groove and reproduce the mechanical vibration that cut the original groove. The grooves can be cut two ways: from side-to-side *(lateral)* or up-and-down *(vertical)*.



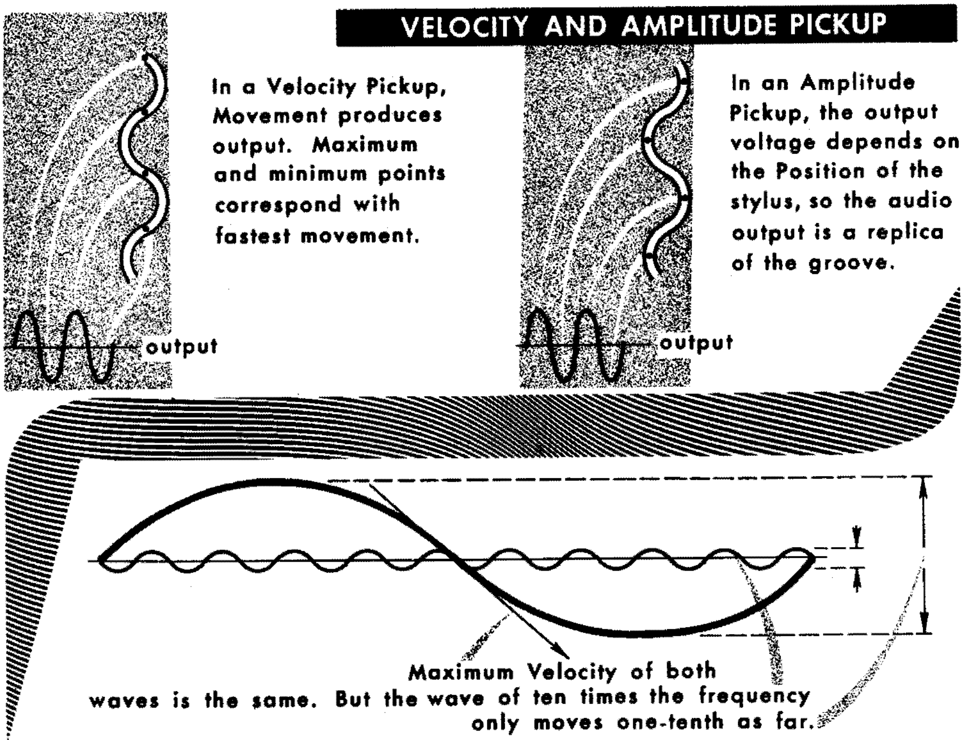
Lateral cuts are invariably used in modern recordings, because this method involves continuous removal of the same amount of material (called *swarf*) from the surface by the cutter. Vertical (hill-and-dale) cutting involves variation in the amount of material removed at the different points on the waveform, which places a varying load on the cutter. This variation in itself can cause distortion.

Cutters and Pickups

The action of pickups and cutters used on discs can be regarded as similar to the action of microphones and loudspeakers with sound waves. The principal difference is that pickups and cutters only move back and forth in a specific direction, whereas sound waves distribute themselves in a more complicated manner. In addition, the material in the disc is heavier than air. Consequently, the mismatch problems met in making the diaphragm of the microphone or loudspeaker act uniformly over a wide range of frequencies are not present. The loading effect of the disc material better controls the action of the cutter or pickup stylus to produce uniform results, than does the loading action of the air on the microphone or loudspeaker diaphragm. Common varieties of cutters and pickups include the magnetic, moving coil, ribbon, crystal or ceramic, and capacitor types.



Velocity and Amplitude Systems



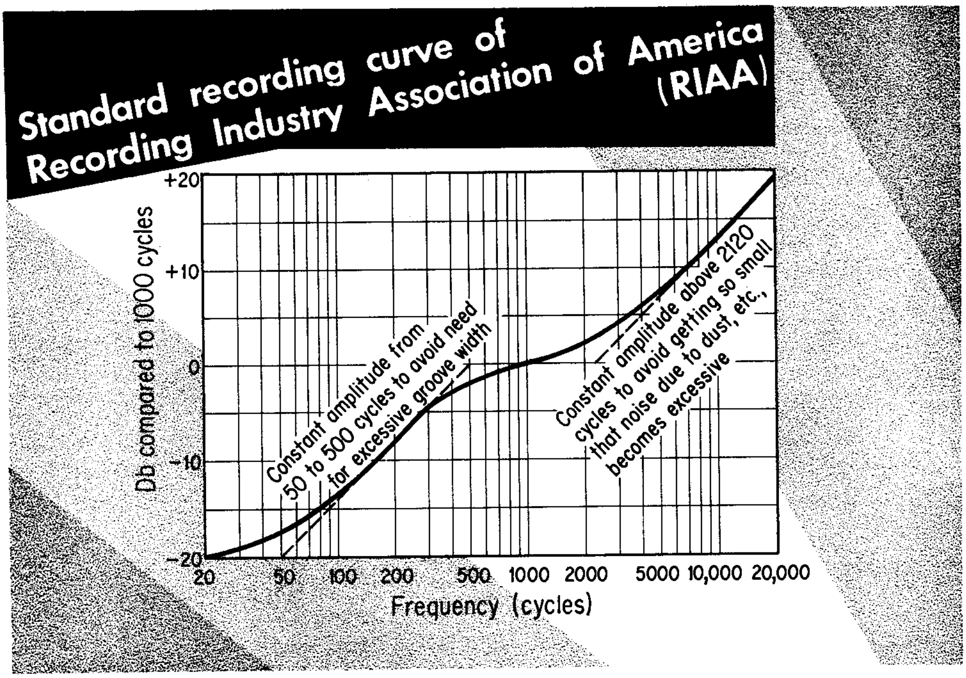
As well as different kinds of transducers for the cutter or pickup head, there are different principles of operation; some work on a *velocity* principle in which the output voltage is proportional to the *rate* at which the stylus moves. Others work on an *amplitude* principle in which the output voltage is proportional to the *amount* that the stylus moves. The velocity-type transducers include the magnetic, moving coil, and ribbon. The amplitude-type transducers include the crystal or ceramic and the capacitor.

The important difference between these types is in the kind of frequency response that they produce. If the disc has grooves of equal magnitude from side to side at different frequencies, the amplitude-type pickup will give a flat response. The rapidity with which the stylus moves is, however, proportional to frequency; hence, this type of recording would produce an output proportional to frequency, if the pickup were of the velocity kind. (This would require special equalization on playback.)

To produce a recording of the velocity kind that gives uniform output with frequency, the amplitude must vary *inversely* with frequency. (Doubling the frequency must halve the amplitude, so that the maximum rate at which the stylus moves at the different frequencies will be constant.)

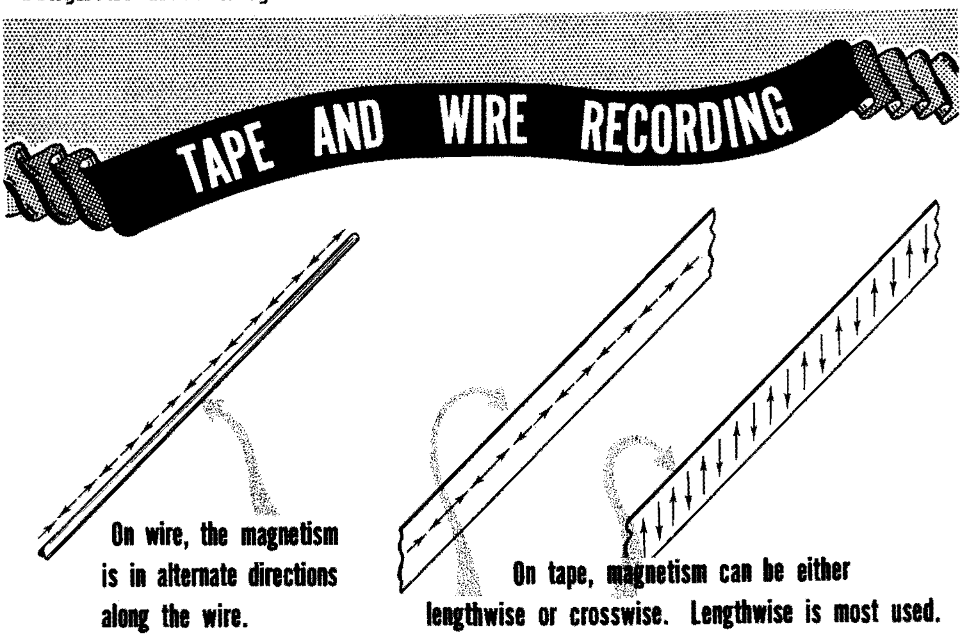
Velocity and Amplitude Systems (contd.)

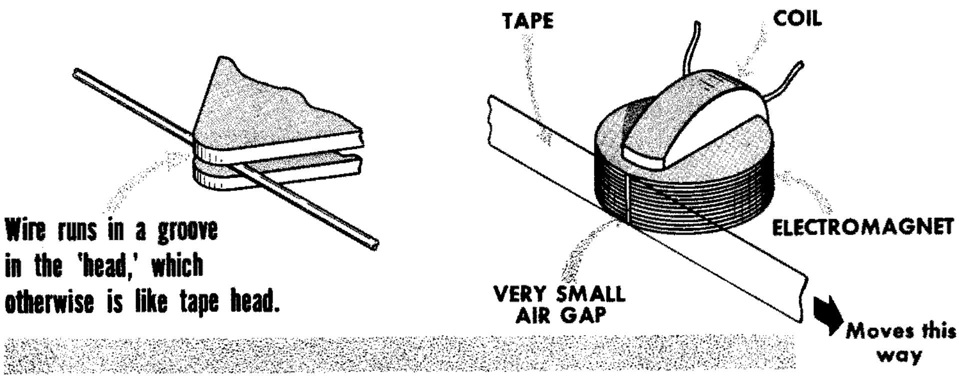
Standards of disc recording have been based on the *velocity* principle, using magnetic, moving coil, or ribbon type transducers for cutters or pickups. Assuming a recording from 20 cycles to 20,000 cycles, this would mean the *magnitude* of movement to give the same output at 20 cycles would be 1000 times what it is at 20,000 cycles. Obviously, this would be impractical. If the movement is 1/10,000 inch at 20,000 cycles, 1/10 inch movement will be needed at 20 cycles. 1/10,000 of an inch is so small that the roughness of the record will make more noise than the signal recorded; on the other hand, 1/10 of an inch is so wide that it would be impossible to get more than 10 grooves to the inch.



For this reason, there must be a change at some point to constant-amplitude recording. The Recording Industry Association of America (RIAA) puts this turnover at 500 cycles. To overcome the fact that the extremely small movement, above, say 10,000 cycles, would mean that the surface noise would be louder than the audio, constant amplitude is also applied above a frequency of 2,120 cycles. These two changes from the velocity characteristic avoid excessive stylus movement at low frequencies and prevent noise from overriding the audio at high frequencies. This varying frequency response is provided by an equalizer in the recording amplifier and must be compensated for by opposite equalization in the playback amplifier.

Magnetic Recording

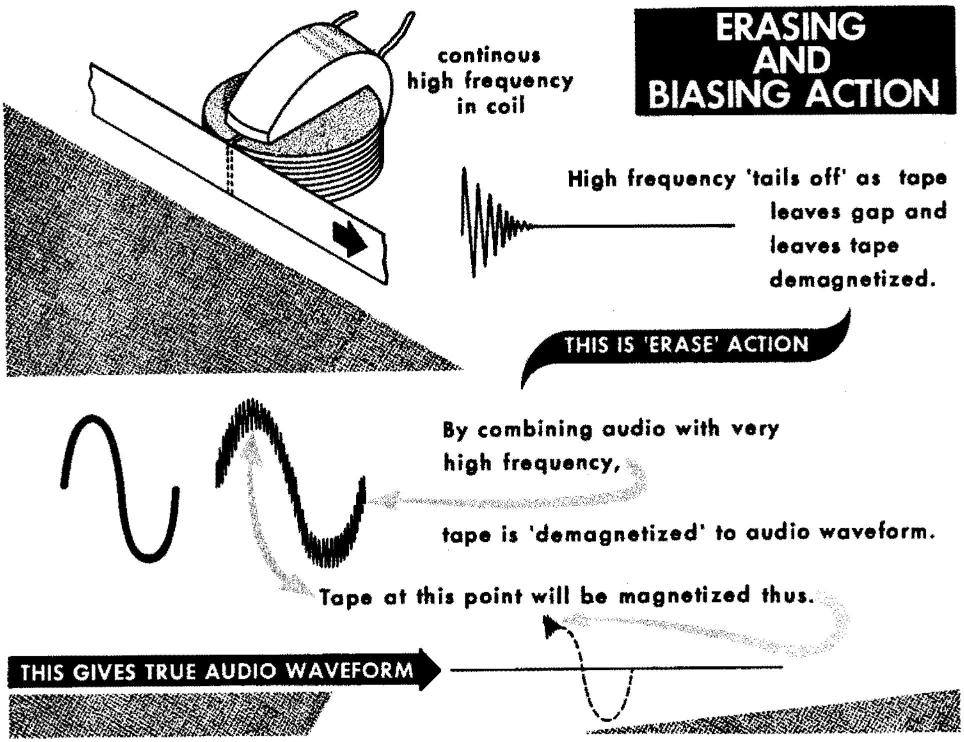




Tape and wire recording use a magnetic pattern corresponding to the audio, which is put onto a magnetic material that will hold it indefinitely. The additional track width on tape permits more magnetism to be put on it, giving a better discrimination of audio against background noise.

Part of getting a good recording depends on using a completely demagnetized tape. This is achieved by using an *erase* head, which the tape passes before it reaches the *record* head.

Ultrasonic Erase and Bias

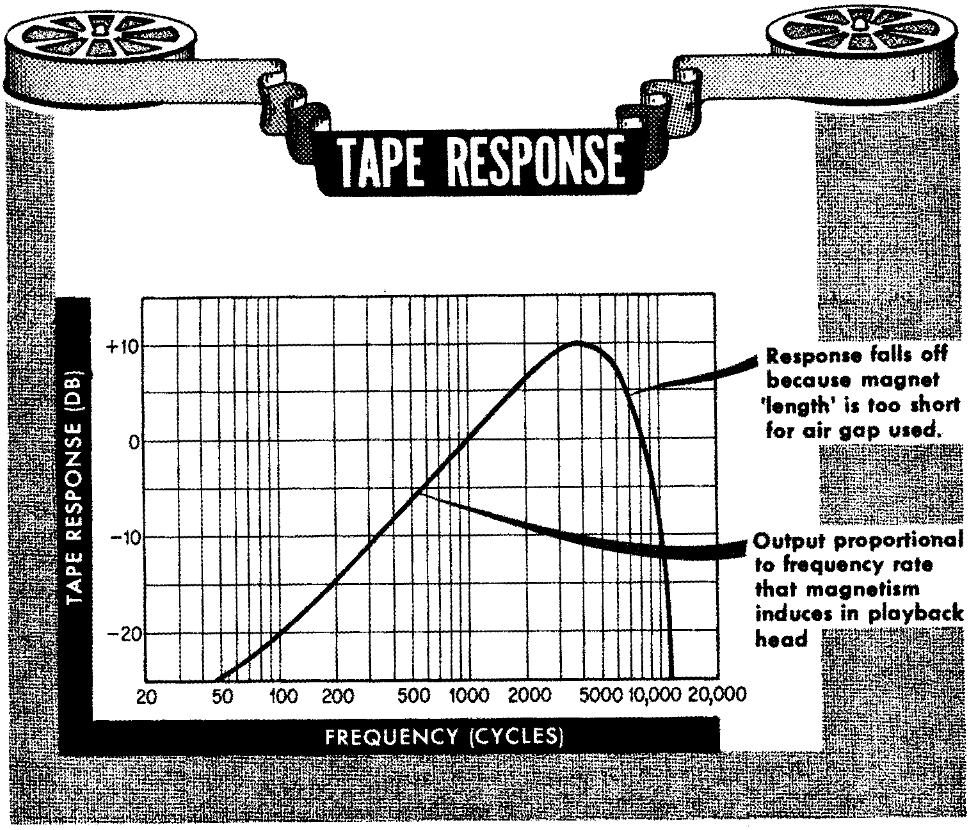


The best form of erase uses an ultrasonic oscillator. As the tape passes the magnetic gap, the magnetization due to the oscillator signal alternates several times because of its very high frequency. As a result, the magnetization cancels out as the tape leaves the gap, ultimately demagnetizing it. This type of erase is necessary because of the *hysteresis* effect in any magnetic material. Too little reverse magnetism does not demagnetize; too much remagnetizes in the opposite direction. Repeated magnetization at ever decreasing intensity reduces the magnetism until the tape is completely demagnetized.

This hysteresis action in magnetic material is responsible for very drastic distortion, unless something is done about it. The simplest way to eliminate this distortion is to use ultrasonic bias. The audio is combined with some of the same high-frequency signal used for erase. The supersonic bias is quite strong, compared to the smaller audio fluctuations at least (full-amplitude audio will be about equal to the bias). As the tape leaves the gap and the demagnetizing action of the ultrasonic frequency fades out, the magnetism left on the tape is the audio component, which does not fade out, because its fluctuation is much slower (only a fraction of a cycle, instead of many times, while the tape is passing the recording head).

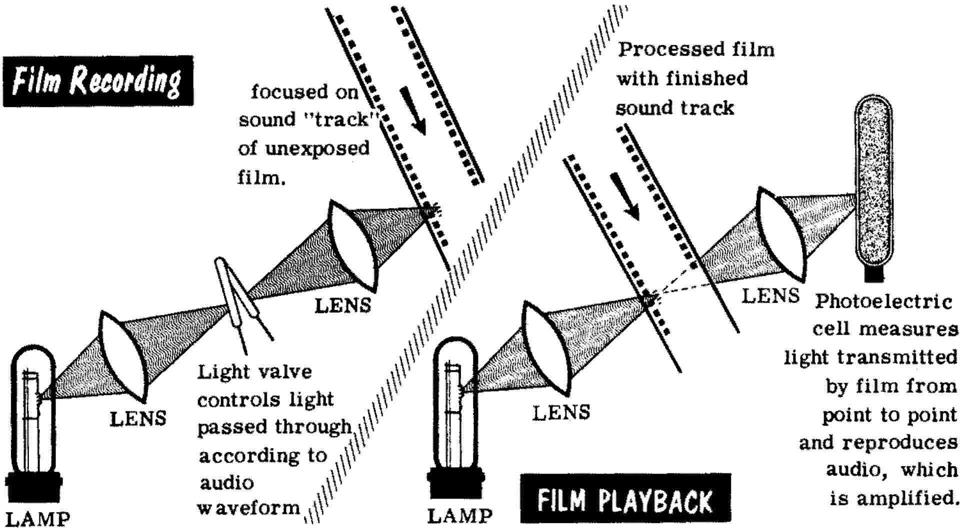
Tape Response

With magnetic recording, the output from the tape passing the gap is proportional to the *rate* at which the magnetism changes. Consequently, it will rise with frequency, at least up to a point where the length of the effective magnet on the tape is about the same as the air gap on the head that is used to pick up the magnetization. When the frequency gets higher than this, the air gap will start to compare the magnetization on the tape at half a wavelength or more, and cancellation begins to occur.



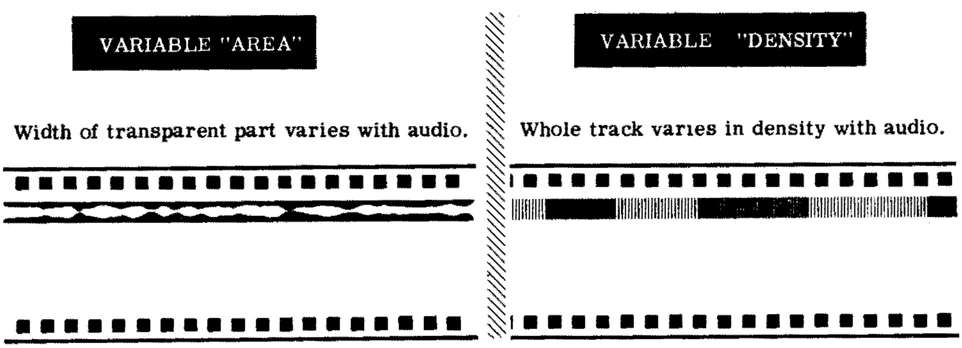
This means we get a rising response all the way up to a certain frequency (which is dependent on the width of the air gap and the speed of tape travel), which falls off very suddenly to zero. Equalization of the input signal, the output, or both, compensates for the characteristic. The magnetization on the tape theoretically is proportional to the current applied to the recording head. This does not allow for losses in the head, however, which complicate the response somewhat. Because of these complexities, only playback response is specified.

Film Recording



Film recording is usually achieved photographically by producing an exposure on the film proportional to the instantaneous audio. This is accomplished by means of a light "valve" that controls the amount of light falling on the film as it passes by a slot. In playback, the same film, with the recording photographically reproduced on it, is scanned by means of a photoelectric cell, the output from which is amplified by an audio amplifier.

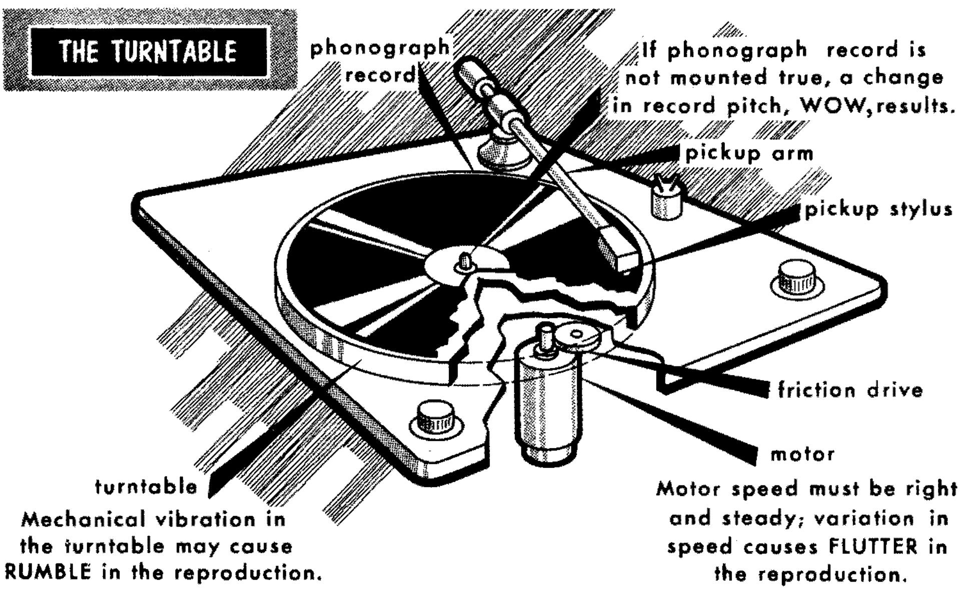
Because the photoelectric cell is sensitive to the quantity of light, the sound track on the film can vary either in the area through which light passes or in the transparency (density) of the film. Both methods have their uses, their advantages and disadvantages, which are concerned more with light and optics than with audio. As with the other methods of recording, special equalization is required on playback.



The Effects of Wow, Flutter, and Rumble

No matter which type of recording is used, constant speed drive is a vital necessity, both for making and for playing back the recording. Variation in speed not only changes the rate at which the program is recorded or reproduced, but also changes its pitch or frequency. Turning a phonograph record faster raises all the frequencies by the same ratio, and the pitch steps up by a constant tone interval. While it is important to have the right speed, it is more important for the right speed to be steady. It must not warble up and down, or flutter up and down.

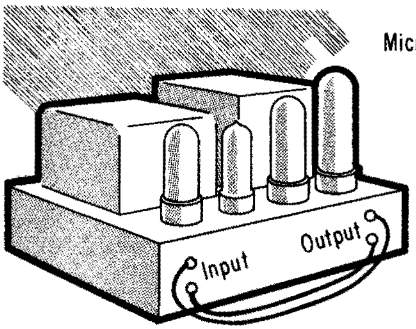
If a phonograph record is not mounted true, this will result in change in speed along the groove, which will cause pitch to vary once per revolution, an effect called *wow*. Wow can also be caused by variation in bearing friction as the turntable goes around. Variations in speed at a greater speed than once per revolution are called *flutter* because of the effect they produce on the reproduction. These effects can happen equally well to optical or magnetic recording, so careful mechanical design, to prevent any nonuniformity of speed, is needed with any recording drive.



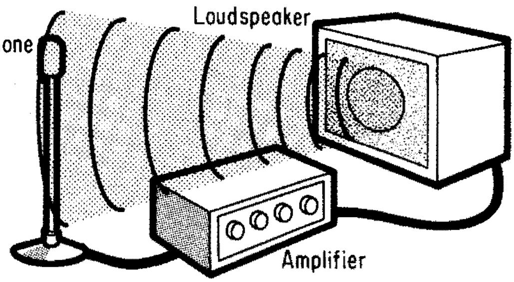
An effect most noticeable on phonograph recording, but entirely exclusive to it, is called *rumble*. This is caused by mechanical vibration in the turntable. The pickup stylus will be moved just as much by the whole groove vibrating as by vibrations that occur in the groove as it passes. Consequently any motor vibration that gets to the turntable will appear in the reproduction as rumble.

Acoustic Feedback

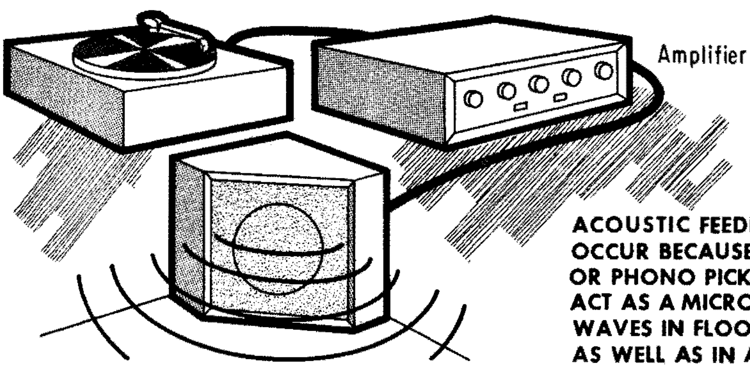
Complete audio systems involve some acoustics, by which we mean the way sound waves radiate, both in the air and through floors, ceilings, etc. The subject of concern in audio is called *electro-acoustics*, because it deals with the combination of electronic and acoustic effects.



IF AN AMPLIFIED OUTPUT FEEDS BACK TO THE INPUT, IT WILL OSCILLATE.



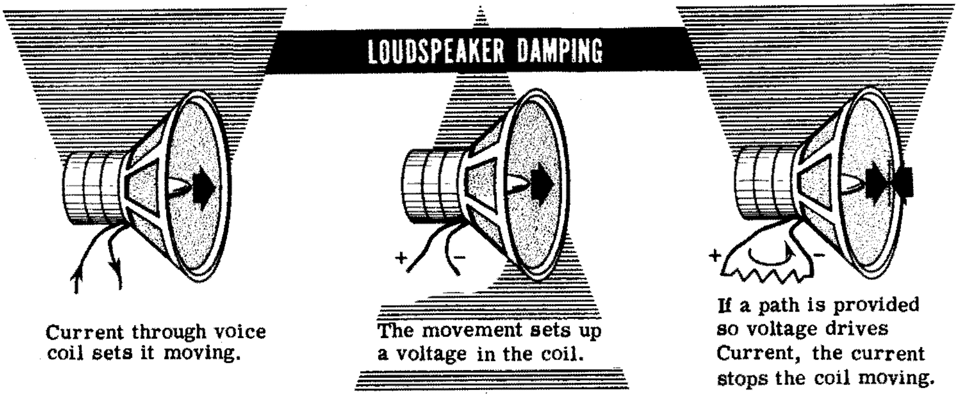
THE FEEDBACK CAN BE BY MEANS OF SOUND WAVES FROM A LOUDSPEAKER CONNECTED TO THE OUTPUT TO A MICROPHONE CONNECTED TO THE INPUT.



ACOUSTIC FEEDBACK MAY ALSO OCCUR BECAUSE AMPLIFIER TUBES OR PHONO PICKUPS SOMETIMES ACT AS A MICROPHONE TO SOUND WAVES IN FLOORS, WALLS, ETC., AS WELL AS IN AIR.

An amplifier can become unstable because high-level audio (electrical power) at the output is fed back to the low-level input circuits. It can also become unstable because high-level acoustic energy radiated from the loudspeaker is fed back to a microphone. This occurs in public address systems where the complete system consists of a microphone, an amplifier, and loudspeakers to reinforce the sound of the speaker on the platform, or the orchestra on stage. Acoustic feedback is not limited to systems possessing microphones; it can also occur in a home reproducing system due to a tube, or the phonograph pickup, *acting* as a microphone. It does this by picking up either sound waves from the air or acoustic vibrations coming through the floor or walls. For this reason, care has to be taken to insure that high-level vibration cannot reach the turntable, the pickup, or any of the amplifying tubes that handle low-level audio if they are at all microphonic.

Loudspeaker Damping



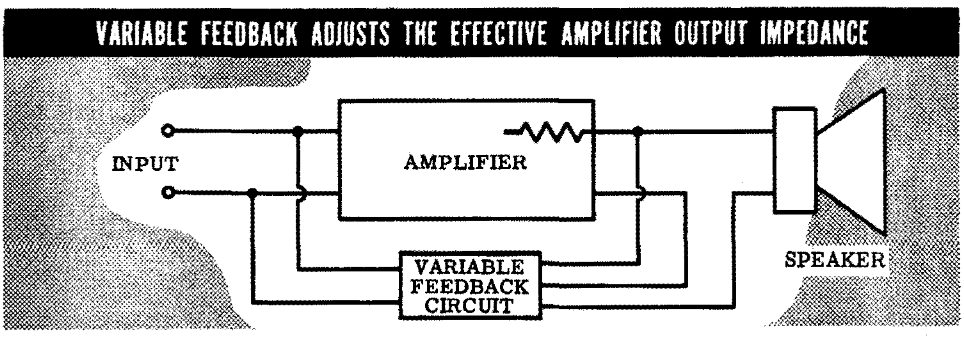
Current through voice coil sets it moving.

The movement sets up a voltage in the coil.

If a path is provided so voltage drives Current, the current stops the coil moving.

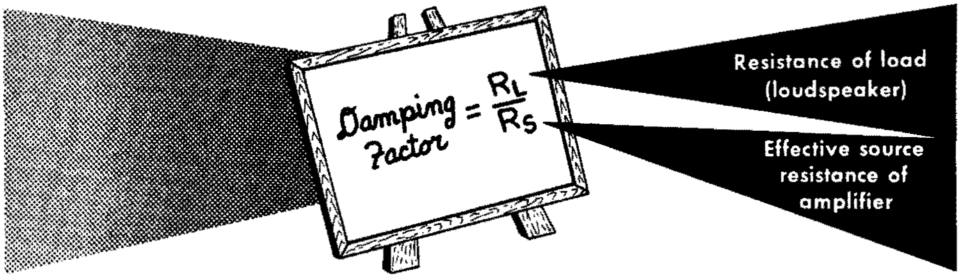
Another aspect of electro-acoustics concerns the proper coupling of the loudspeaker to the amplifier. Not only does the loudspeaker impedance have to be matched to the amplifier output impedance, but the output impedance of the amplifier affects the operation of the speaker. When the speaker diaphragm starts moving (due to an input signal from the amplifier), the momentum of the diaphragm will tend to keep it in motion even after the drive current has ceased. Diaphragm motion causes the loudspeaker to operate as a microphone, generating voltages in its voice coil. This overshoot movement will generate a further voltage. If the impedance of the amplifier is high, there will be a negligible current in the circuit due to this voltage. If the amplifier impedance is low, however, current flows producing a force that acts to stop overshoot.

This means the source resistance presented by the amplifier will influence the behavior of the loudspeaker. A low source resistance will prevent the voice coil overshooting, while a high source resistance will allow the voice coil to move erratically and affect the transient response of the loudspeaker. This means that the source resistance presented by the amplifier, due to the plate resistance of the output tubes, adjusted according to any positive or negative feedback, will influence the behavior of the loudspeaker.

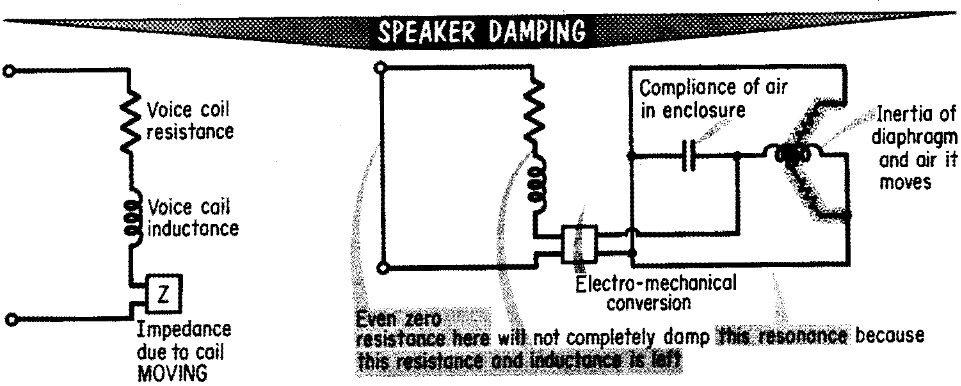


Loudspeaker Damping (contd.)

The effect of source resistance on loudspeaker behavior is called the *damping factor* of the amplifier. Because a higher source resistance damps or brakes overshoot less than a low source resistance, the damping factor is given by the load resistance *divided* by the effective source resistance.



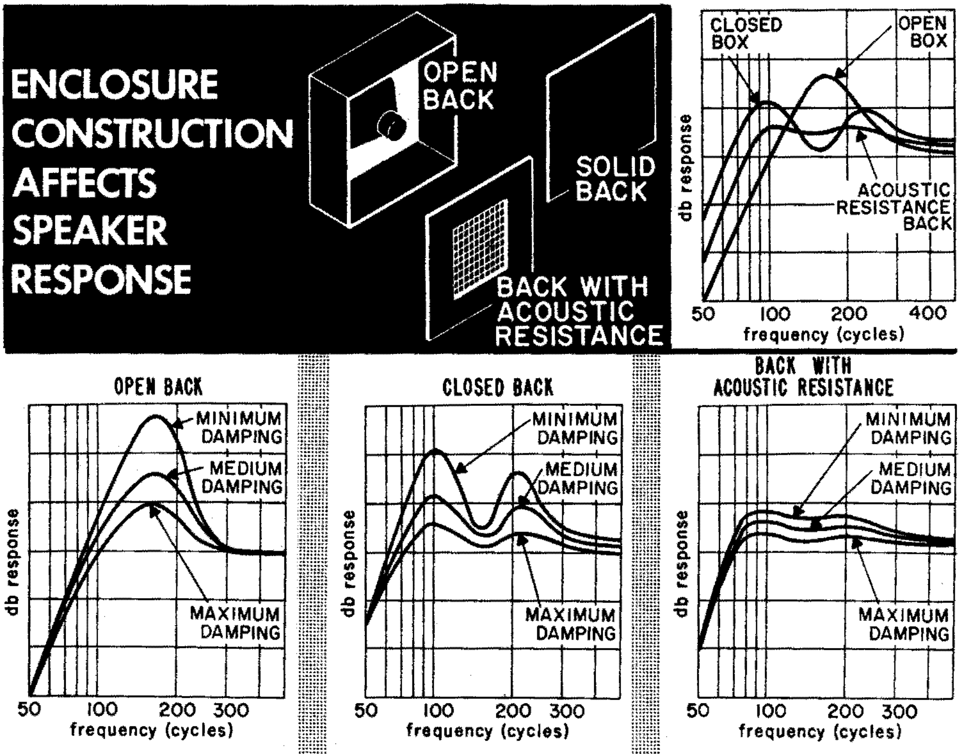
The use of a high damping factor (or a low source resistance) cannot do everything. The electrical damping has to act on the motional impedance of the loudspeaker. (Motional impedance is the impedance reflected back to the loudspeaker terminals due to the movement of the diaphragm.) This impedance is only a small fraction of the total impedance of the loudspeaker, most of which is due to the resistance and inductance of the voice coil.



For this reason, the loudspeaker is quite inefficient. The greater part of the power delivered to it by the amplifier is expended in the voice coil resistance and merely heats the voice coil, whereas only a relatively small proportion is radiated as sound, due to the motional impedance. Only a correspondingly small fraction of the speaker's output is available for damping. Hence, even an infinite damping factor would have a limited effect on mechanical or acoustic resonance. The resistance of the voice coil (which cannot be eliminated) limits the damping current.

Demonstration of Acoustic Damping

For this reason, the best way to achieve good operation of a loudspeaker is to damp its movement *acoustically* by attention to the construction of the loudspeaker enclosure. This can be illustrated quite effectively by mounting a loudspeaker in a simple enclosure with three alternatives for the back of the enclosure.

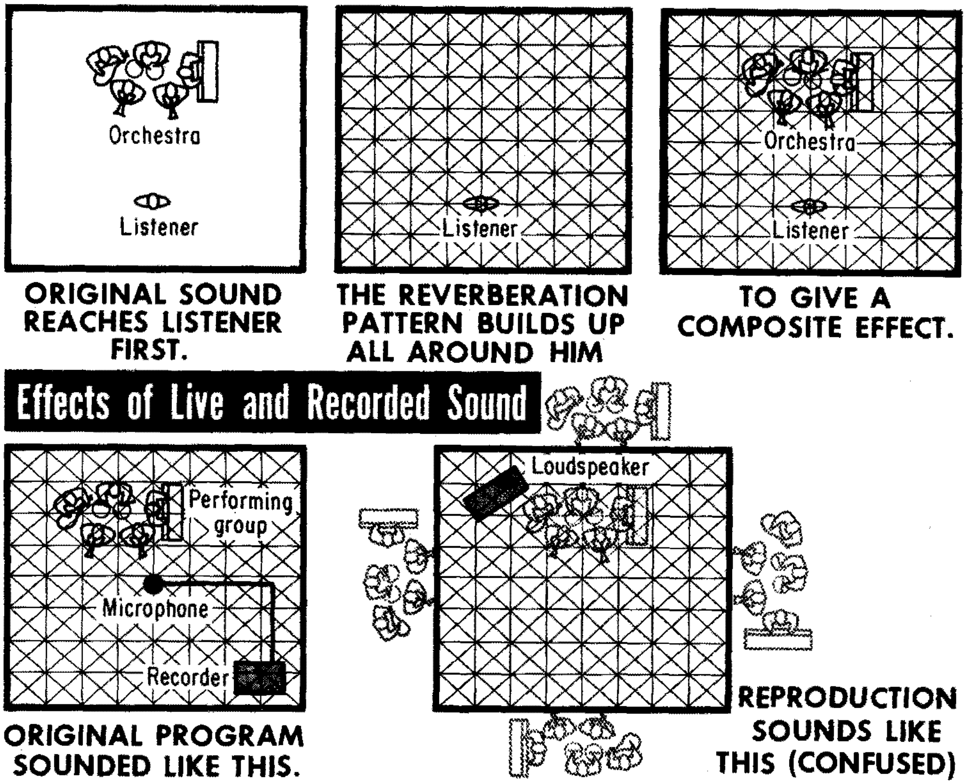


A completely open back results in a load on the diaphragm that acts as a mass or weight of air moving bodily. A completely closed back makes the air inside the cabinet act as a compressible cushion. Using an acoustic resistance consisting of a large number of holes that allow the air to pass through but offer resistance to its passage achieves an intermediate condition that damps the diaphragm properly.

Comparison of results shows that the response is much smoother with the acoustic damping and that under this condition, electrical damping becomes unimportant. With either the solid or the open back, electrical damping from the amplifier makes a considerable difference to the response of the loudspeaker, but neither performance is as good as that with the acoustic resistance.

Characteristics of Hearing

As discussed in Volume 1, the fact that we have two ears permits us to tell from which direction sound is coming. This permits us *subconsciously* to distinguish direct from reflected sound and to concentrate on what we wish to hear. At a live performance of a symphony, for example, we can concentrate on the sound coming to us from the orchestra and ignore the effects of reverberation. The degree to which this is true can be shown by placing a microphone in the position at which we were listening. The microphone cannot distinguish between direct and reflected sound, and a recording made in this way seems to have far too much reverberation. The sound reproduced through a loudspeaker seems quite confused because our hearing cannot separate direct and reflected sound now that they both come from one source.



ORIGINAL SOUND REACHES LISTENER FIRST.

THE REVERBERATION PATTERN BUILDS UP ALL AROUND HIM

TO GIVE A COMPOSITE EFFECT.

Effects of Live and Recorded Sound

ORIGINAL PROGRAM SOUNDED LIKE THIS.

REPRODUCTION SOUNDS LIKE THIS (CONFUSED)

The ability to concentrate in hearing permits us to carry on a conversation with one person in a crowded room. The sound of his voice is probably no louder than that of anyone else's (and certainly no louder than all of the others together), yet, by concentrating, you can hear what he is saying, almost to the exclusion of all other conversations going on around you.

The Physiology of Concentration

Concentration is aided by two things: your sense of direction (which helps you to concentrate on sound coming from where your friend is standing) and the individual characteristics of his voice (his "voice personality") which helps you to separate what he is saying from what people with different voice personalities are saying. How is this discrimination achieved, when all we can really tell apart (in audio terms) is the frequency content and the intensity of individual frequencies.

It is all possible due to the form that individual hearing perception takes in its transmission from the individual ears to the brain. Each ear has a number of resonators in the cochlea that are sensitive to individual frequencies, using from 20 to 29,000 to cover the entire audio frequency range. Each of these *receptors*, however, does not transmit its individual frequency to the brain. What it does transmit is a series of nerve pulses.

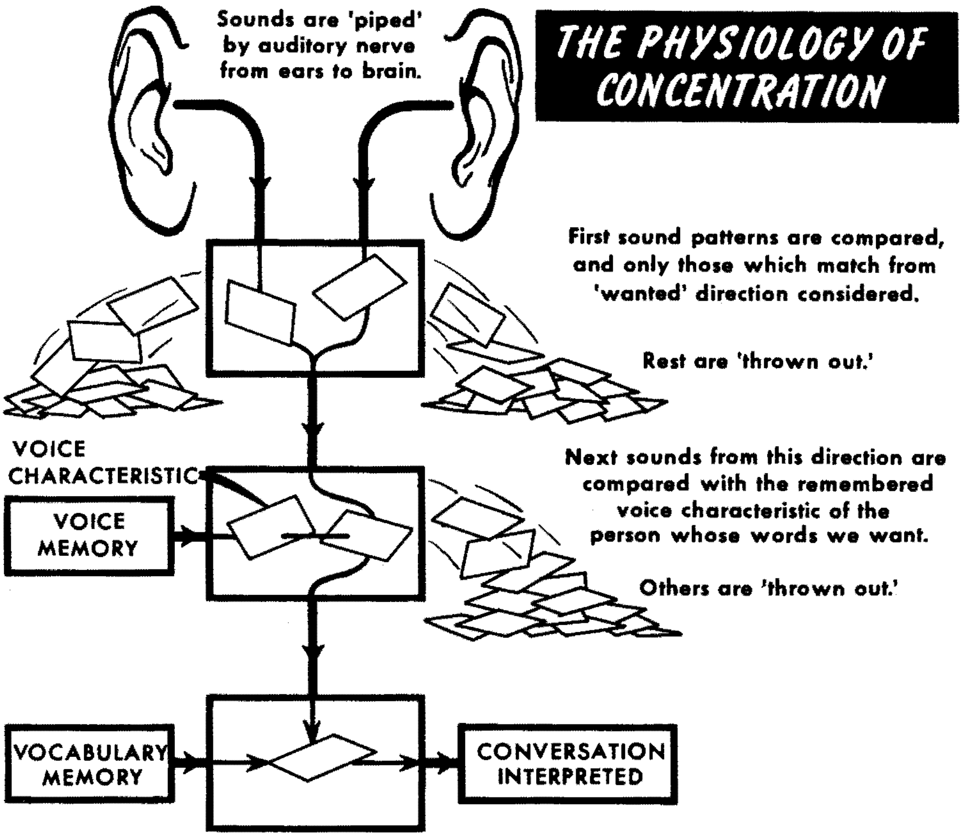
The first pulse to be transmitted to the brain is sent when that particular frequency is first detected by the ear. Thereafter, a sequence of pulses is sent along the same nerve fiber, dependent upon the intensity of that frequency at the instant. The louder the sound component of the particular frequency, the more frequently this nerve carries pulses to the brain.

The section of the brain devoted to the analysis of sounds heard "recognizes" a complicated pulse pattern by comparison with patterns already familiar. There are a number of characteristics by which the patterns can vary, enabling deliberate "differential listening." The most important part of any group of patterns is the first set of pulses along any particular grouping of nerve fibers. This is the transient effect of the sound heard. The following grouping indicates the way the sound varies in intensity or tone quality after it first starts.

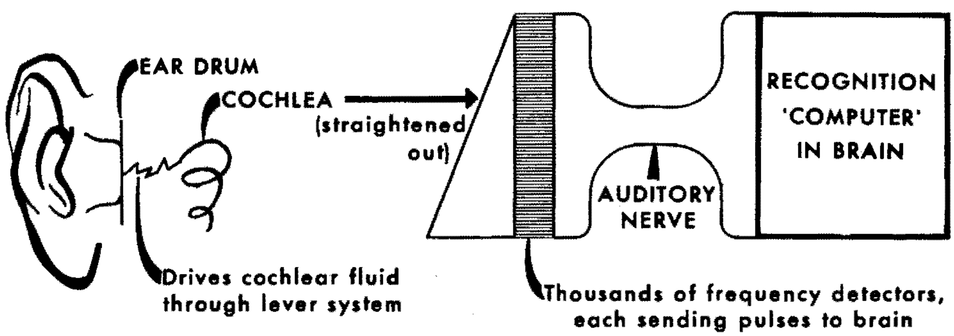
The auditory nerve conveys to the brain a pulse code pattern corresponding with all the sounds being heard all the time. The code can be decoded in a variety of ways because of its enormous complexity and the vast resources of the brain cells that interpret sound. Sounds can be grouped, for example, according to the particular pattern of fundamental and overtones characteristic of a violin. Thus, when listening to orchestral music, the sound of the violin can be distinguished from all the other musical sounds going on at the same time. Similarly, one person's voice can be singled out from the voices of many other people by noticing the pattern peculiar to that person's voice personality. The automobile mechanic's ear becomes tuned to the sound patterns that come from each component in an automobile engine. As a result, he can distinguish a sound due to a knock in an end bearing as different from one due to loose valve tappets, among all the other engine noises.

Although the hearing faculty is extremely critical and can detect any particular sound with high precision, it is also extremely tolerant: it can pay attention to what it wants to hear and ignore everything else.

The Physiology of Concentration (contd.)

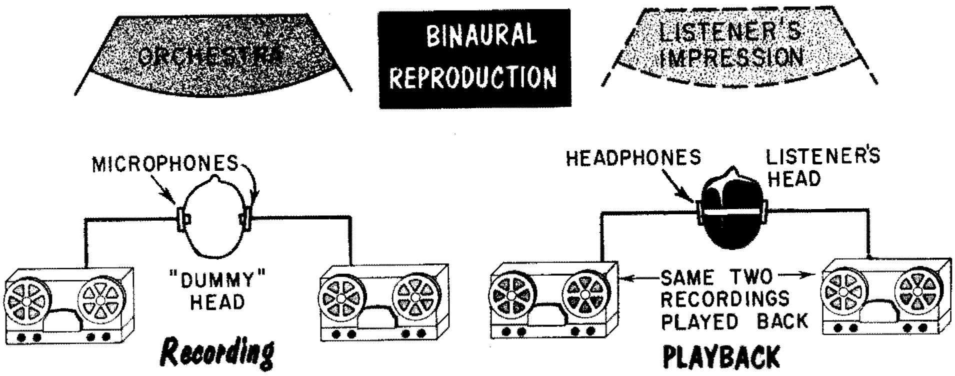


HOW WE CAN HEAR SELECTIVELY



EACH NERVE SEPARATES A "NEW" PULSE FROM FOLLOW-THROUGH ONES OF THE SAME FREQUENCY.

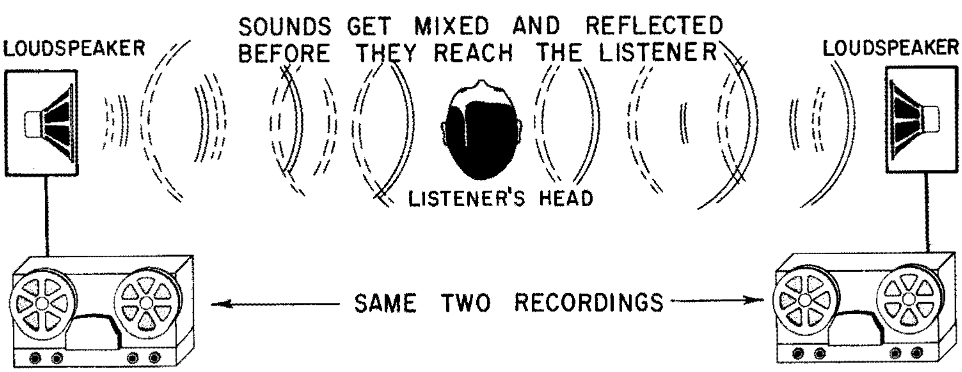
Binaural Reproduction



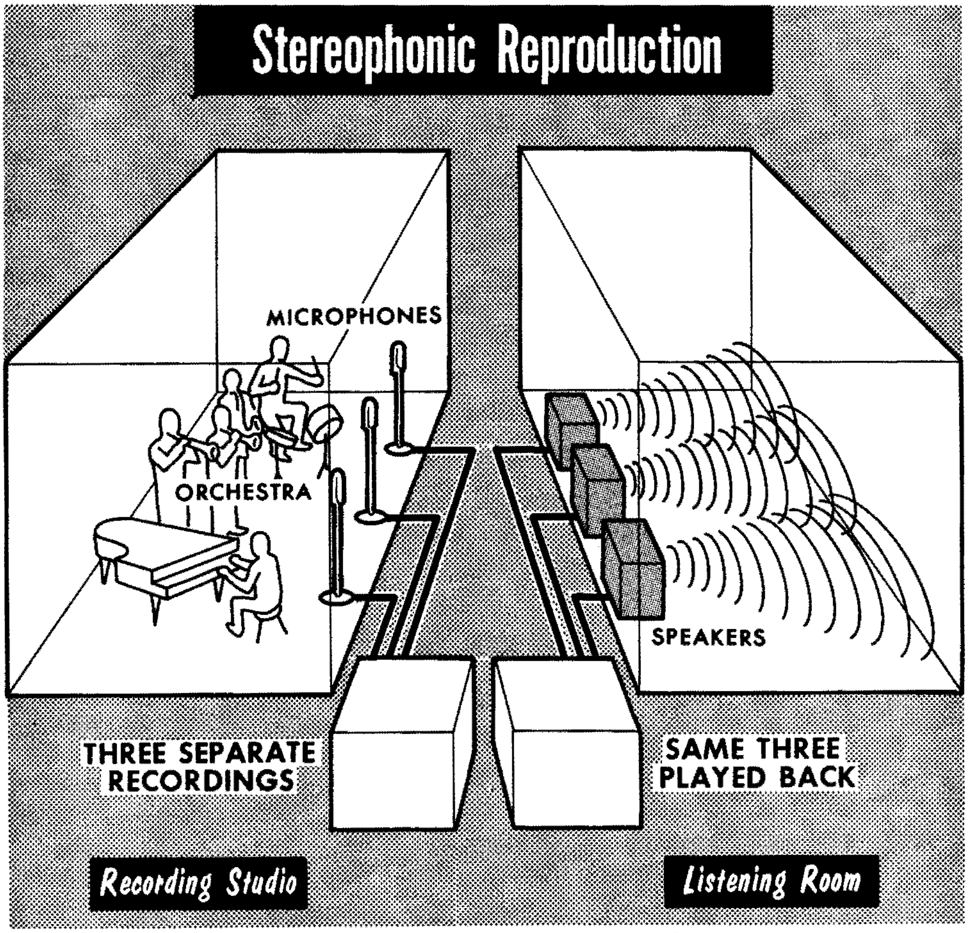
This brings us to some final aspects in achieving realism in high-fidelity reproduction: the artificial reproduction of sound that gives an impression of coming from different directions and enables the hearing faculties to draw similar conclusions to those that would have been received in the original program.

There are two main approaches to this objective. One uses a dummy human head with a microphone in the position of each ear and makes two-channel recordings. By means of a pair of headphones, one of these channels is delivered to each ear of the listener when the record is played. This system is called *binaural* reproduction. This same two-channel sound could be reproduced over two loudspeakers, but the effect would not be as good. When sound is released into the room by two loudspeakers, the waves from each loudspeaker reverberate around the room and recombine. The impression conveyed to each ear in this way is not the same as that conveyed by the earphones and the sound field created is not like the original field.

TWO-CHANNEL SOUND USING TWO LOUDSPEAKERS



Stereophonic Reproduction

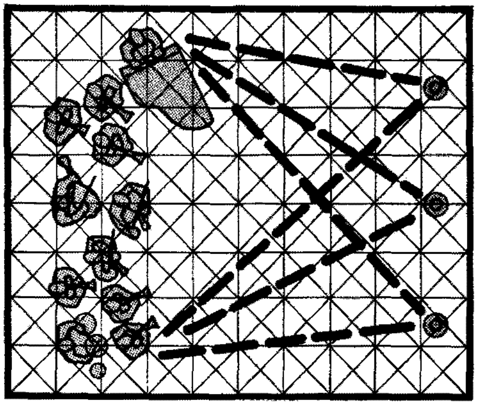


The stereophonic technique of reproduction takes the approach of endeavoring to recreate an original sound *field*, avoiding the necessity for wearing headphones. To do this, the recording microphones are spaced a given distance apart and the loudspeakers are spaced a similar distance apart. This, to some extent, recreates the original time and phase differences that set up the directional effect in the sound waves. It cannot completely achieve the desired result because the acoustic characteristics of the listening room as well as those of the studio are present. (The original sound field will never be fully recreated because of the additional reverberation added in the listening room.) If the reverberation in this listening room is reduced by putting very absorbent material on all walls, the floor, and the ceiling, the sound is still unnatural because then the reverberation would not appear to come from all directions in the same way as it did in the original performance.

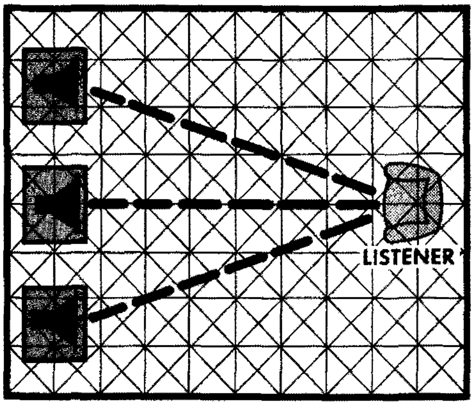
Stereophonic Reproduction (contd.)

In practice, the program is recorded in a studio that has less reverberation than the final reproduction should have, which permits the listening room characteristics to complete the sound pattern. It is better to allow the sound pattern to be slightly different from the original, but still a *credible* reproduction of it, than to try to reproduce the original sound exactly and fail in some more important aspect.

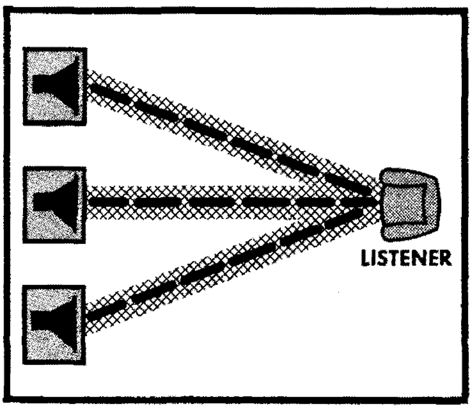
STEREO HAS ITS PROBLEMS



Original studio reverberation is recorded.



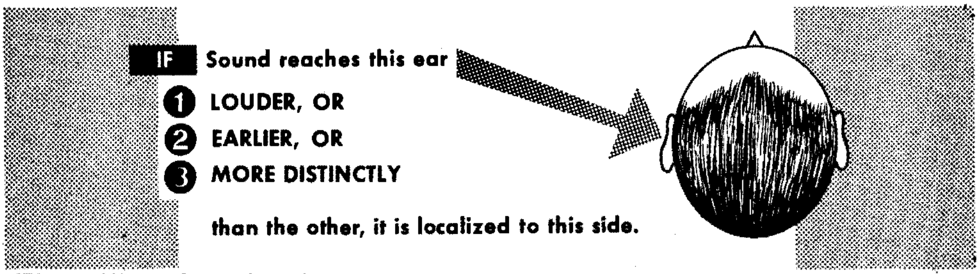
Playback room adds its own reverberation.



Using a 'dead' room eliminates playback room reverberation but the recorded reverberation comes with the original sound, which is still not natural.

For this last reason, it is particularly important in stereophonic reproduction to pay particular attention to how transients reach the listener and the method by which the reproducer systems throws them out into the room. If the transient pattern is a realistic recreation of the original, the reverberation effects will be relatively unimportant.

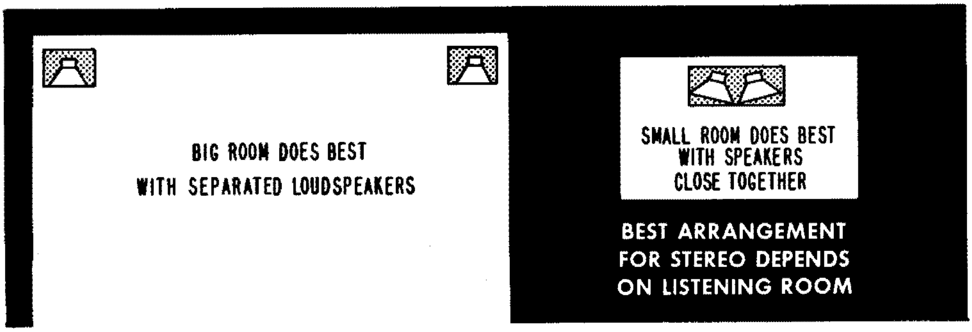
Stereophonic Reproduction (contd.)



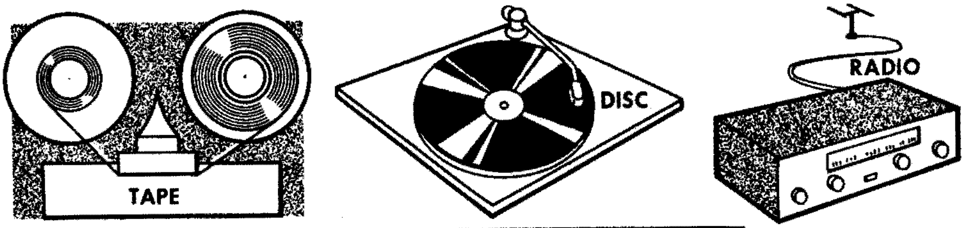
The ability of our hearing faculty to locate sound sources depends on three main differences in the sound at our two ears: *intensity, time,* and *quality.* If one sound is stronger at one ear than the other, the apparent direction is toward the side where it is received the strongest. If the sound is heard by one ear a fractional interval before the other, it will seem to come from the direction where it was heard earliest. Finally the hearing faculty notices the frequency content reported by both ears. It will also interpret a clearer sound, because more of the higher frequencies are present, as meaning that the sound originates from that side.

The hearing faculty uses all of these differences at once, but the relative importance it attaches to each will depend on room acoustics. In a big, reverberant room with a lot of echos, whether these are noticeable separately or just add confusion, time difference becomes almost meaningless. So our hearing concentrates on the other differences. In a smaller room, where reverberation is almost absent (and by the very size of the room, intensity must equalize somewhat), the hearing faculty gets extra critical about time differences.

This means that, in a large room, wide-spaced speakers fed with audio in which the intensity corresponding to different locations is emphasized will give the best stereo. In a small room, wide spacing can result in unnaturally large time differences *for the room acoustics.* Thus a system with loudspeakers close together, which relys on *the way* they radiate rather than on relative intensity, will do the best job.



Stereophonic Source Material



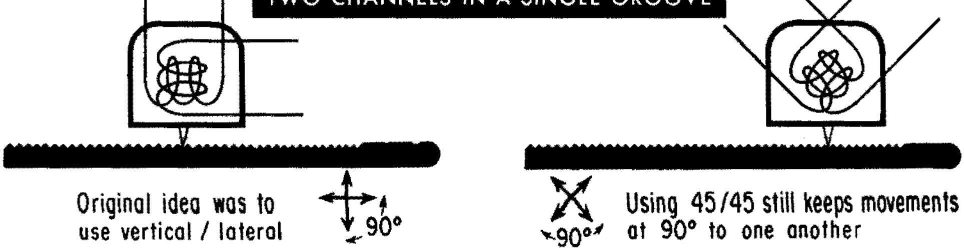
DIFFERENT SOURCES FOR STEREO PROGRAM

To get the most from a home form of entertainment such as stereo, we need a variety of sources. The first — because it was the easiest to adapt — was the two-channel tape recorder. An early attempt to apply it to discs used two separate grooves, spaced apart by a fixed distance. But this was very clumsy to use. What was needed was a means of putting two channels into one groove. This is not basically difficult, but it has practical difficulties that are not really audio problems.

The first idea was to use up-and-down motion of the stylus for one channel and the regular sideways motion for the other. As both are mutually at right angles, neither should affect the other, and they can carry essentially independent programs. With this method, the quality of the two channels was not equal. To overcome this, two motions mutually at right angles are used, but each is at 45° to either vertical or horizontal. This enables equal quality to be used in both channels. By phasing the channels so that unison between them results in the vertical components canceling and the horizontal components adding, the new 45/45 records are compatible with the unstereophonic LP's that only carry sideways-motion grooves.

For radio there are several possibilities: an AM broadcasting station can carry one with the FM station carrying the other; two FM stations in the same area can work together to send out half each of the same program. Various other combinations between TV and either AM or FM have been used. But the most promising for general use is one of several multiplexing methods (two channels on a single FM channel) that an adapter can "decode" at the receiver.

TWO CHANNELS IN A SINGLE GROOVE



QUESTIONS AND PROBLEMS

1. What kinds of storage medium can be used for audio recording? Name some uses for recording other than high fidelity.
2. What two basic ways are there of recording on discs? How does the problem of making a cutter and pickup for disc recording compare with that of making a loudspeaker and microphone?
3. What kinds of transducer element can be used for a phonograph pickup? Indicate which ones use velocity and which the amplitude principle.
4. Even though velocity recording is the basic form, why do modern recordings deviate so as to approximate nearer to constant amplitude? What turnover frequencies are used for this transition?
5. What is the best way to demagnetize tape or wire completely, preparatory to recording on it, and why?
6. Explain how ultrasonic bias works to improve the quality of magnetic recording.
7. Discuss the basis for equalizer requirements in magnetic recording.
8. How does optical recording work?
9. What is acoustic feedback? Besides the well-known case of a loudspeaker feeding back to a microphone in a public address system, where else can this phenomenon cause trouble?
10. Explain why a loudspeaker needs electrical damping from the amplifier that drives it. What limitations are there to this method of damping? Describe an experiment that demonstrates this.
11. Explain why a recording made with a high quality microphone placed in a typical audience seat would be very disappointing. What ideology of recording or high fidelity does this fact disprove?
12. Explain the physiology of hearing on which this difference between direct hearing and attempts to reproduce the original depends.
13. What is binaural reproduction? How should it be listened to, and what happens if such a recording is played over a more conventional system?
14. Research into the stereophonic nature of sound reveals what important facts? How can the best illusion of realism be achieved?

CUMULATIVE INDEX

NOTE: The first number of each entry identifies the volume in which the information is to be found; the second number identifies the page.

CUMULATIVE INDEX