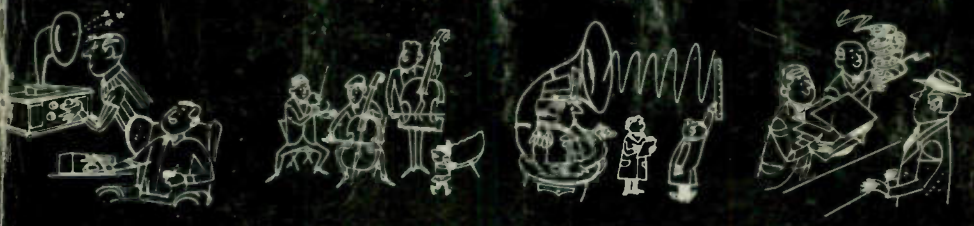
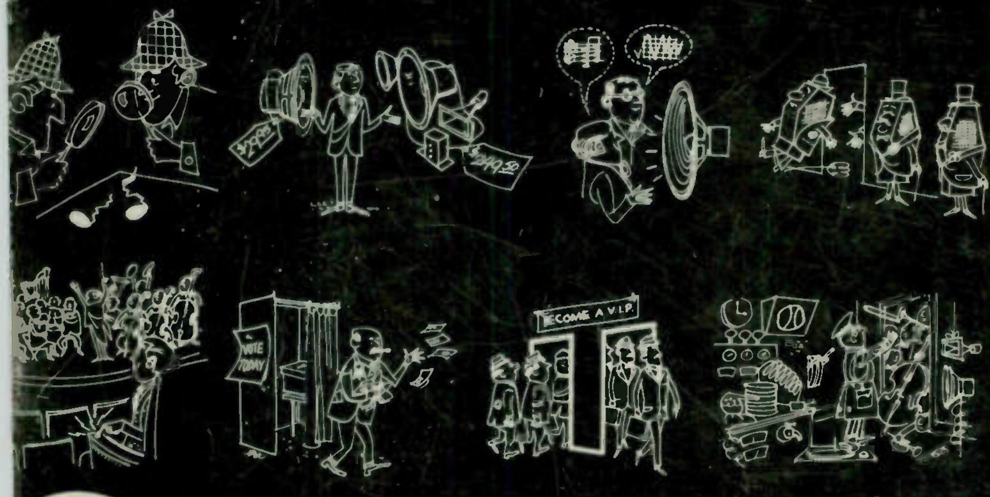


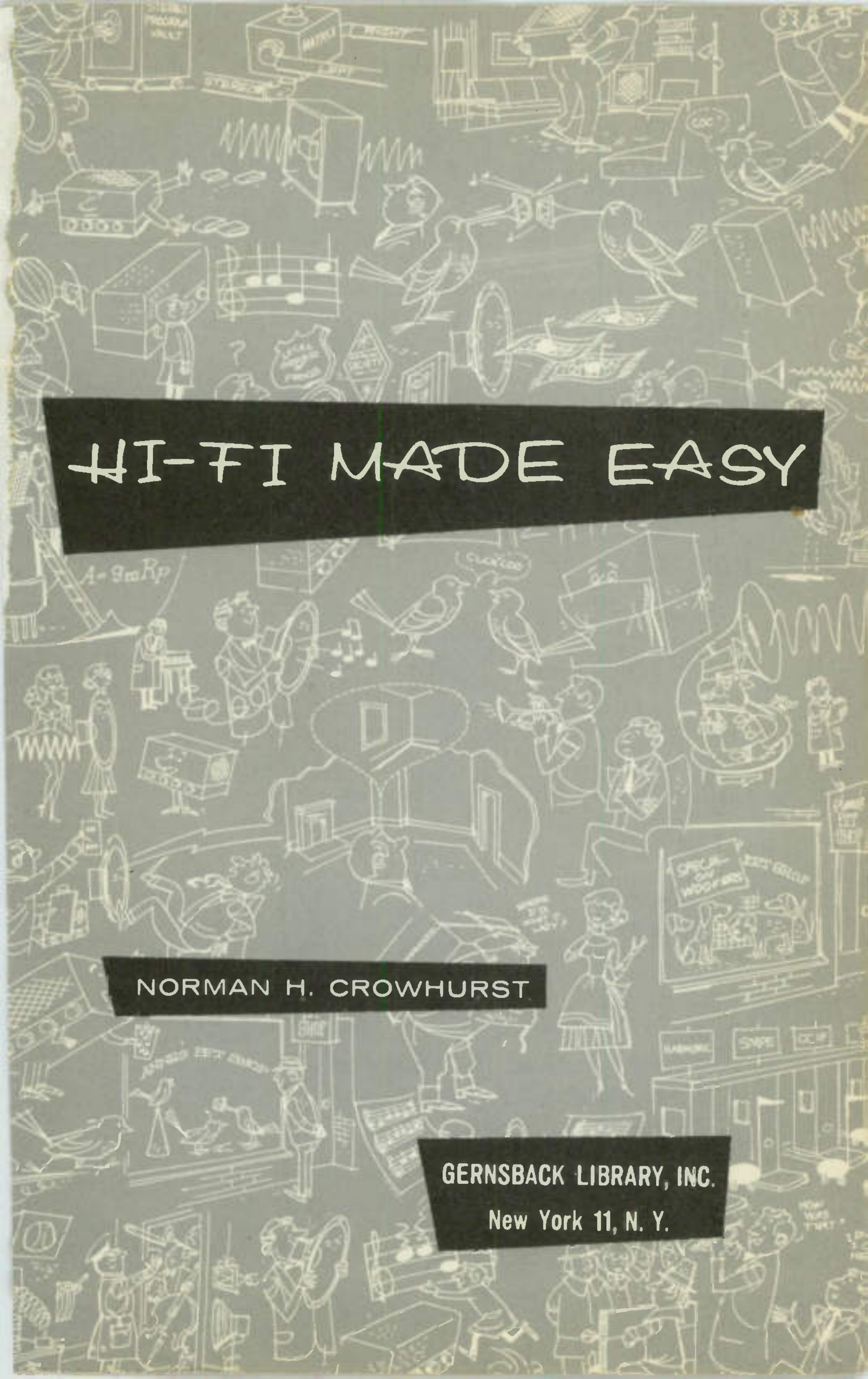
HI-FI MADE EASY

NORMAN H. CROWHURST



GERNSBACK LIBRARY BOOK No. 90

$2.90



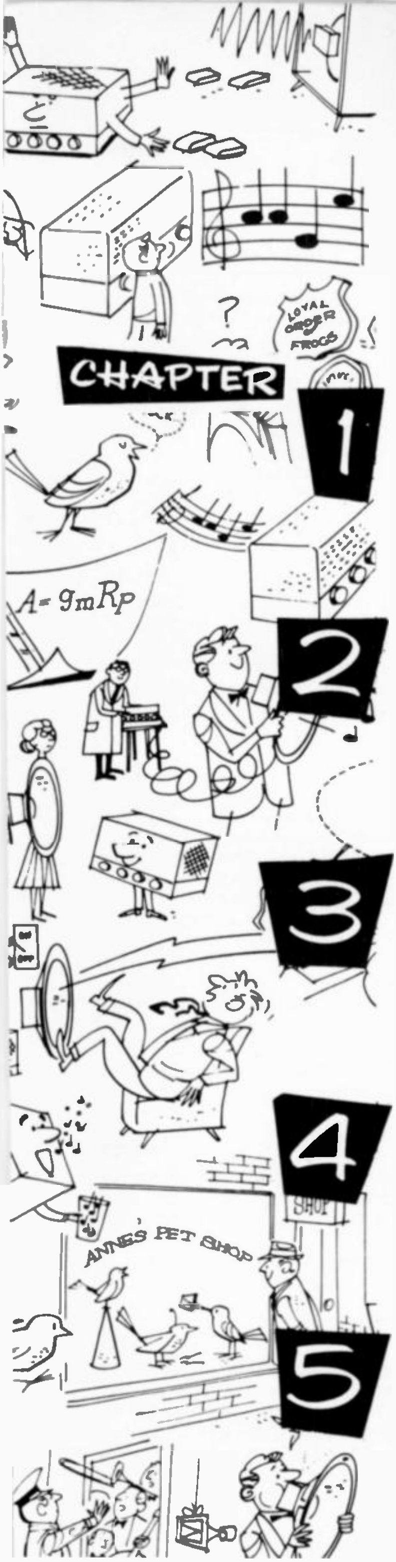
HI-FI MADE EASY

NORMAN H. CROWHURST

GERNSBACK LIBRARY, INC.
New York 11, N. Y.

First Printing—November, 1959
Second Printing—February, 1964

Library of Congress Catalog Card No. 59-15185



contents

Electronics 10.—


SPECIAL ON WOOFERS
PET SHOP
HARMONIC
SMPE
CCIF
STOP PAY TOLL

introduction

How did this high fidelity thing get started? A common question these days. It can be explained according to things or people. To some extent, of course, the two go together. And it's quite an interesting picture.

In any field of endeavor, the incentive toward progress is a desire to improve on something as it is, because it becomes evident it can do with it. Automobiles of today have reached their present standard because people a few decades ago found the initial models to be "bone shakers." So it was that early broadcast and recorded music produced "earshakers."

Yesterday's radios and phonographs left much to be desired in the way of quality. So while some people were still gaping at the amazing fact of recognizable reproduction, others said to themselves, "it's awful; it should be possible to do better than that." So, with the means available to them, that's just what they proceeded to do.

At first the handful of enthusiasts called what they were doing better reproduction, or quality reproduction, or any convenient description. They didn't need a name to merchandise it, because they didn't want to merchandise it. It was the hobby of a select few scattered around the face of the earth, a quest for quality.

But it was a good thing they were onto. Friends would drop in and be invited to listen. Gradually a wider circle began to realize that, if you were so disposed, the radio and phonograph could be transformed into a thing of beauty—at least to one's ears! It wasn't until later—much later—that anyone thought of making hi-fi equipment attractive to the *eye*.

There began to be a market—a very small market, but one worth someone's taking the trouble to serve—for the parts with which to do these fantastic things. This meant it had to have a name. You could not go into a store and say "Do you stock parts for making radio and phonograph reproduction sound better?" or something like that. So the hobby became known as high fidelity.

But so far the expanding members of hobbyists were strictly what today we would call a "hard-core" group. They weren't the long-hair music appreciationists. Their interest was purely and simply in sound quality. They didn't know much about music, but they liked to hear it *clean*.

These enthusiasts can be an argumentative lot. They even had an argument about the name for their hobby. Some favored quality reproduction, or some other variation. Not a few objected to *high* fidelity. "If you must use the word fidelity," they argued, "don't try to qualify it." As they pointed out, according to the dictionary, fidelity is a synonym for faithfulness or loyalty, about which no comparatives are usually allowed. A wife is faithful to her husband, or . . .

Applied to their hobby, they felt the objective should be perfection in reproduction. One does not have "high perfection." Perfection is the ultimate – perfection. Had they attained perfection? That's another matter.

Often they would feel they had "reached it." But after a little while they would become dissatisfied, make some further change which definitely made a further improvement. So gradually they realized they were on a quest. A path always leading, but never arriving. While the latest was always the best, the mentality of the enthusiast will never accept second best. When an improvement has been made, he has to have it.

It seems that certain magazine editors were influential in retaining the high before fidelity, in spite of the critics. These editors were practical men. They realized the name needed should describe the present state, not some never-to-be-attained ideal. So it's high fidelity today. Today's high in fidelity is higher than yesterday's, and tomorrow's will be higher yet.

Along the way, many divisions of opinion have occurred about the "right" way to proceed. For example, when indirectly heated tubes first appeared, making practical a radio or electrical phonograph without batteries, the high-fidelity enthusiasts were divided into two camps.

The new tubes caused hum, both by using a B-voltage rectified from the line and by induction from a heater inside the tube. Never would it be possible to make them as quiet as battery tubes. That was one view. The other group envisioned what eventually happened. The new tubes made amplification to much higher power levels practical, and it was only a matter of time before tube designers licked the various hum problems.

Many were the areas of contention and strongly held were the contending views. Some are held yet, although we doubt whether anyone still prefers the entirely battery operated high-fidelity set, on the score of quality! But there are still people who think triodes were never superseded by pentodes for fidelity, or that feedback was not the heavenly-sent-gift the amplifier manufacturers claimed.

On the average, there are just as many people inclined toward the progressive view as there are favoring the conservative. Some acclaimed indirectly heated tubes, pentodes and other developments, realizing the bugs would eventually be ironed out and the idea itself represented a step forward. This story has been repeated over and over, in each branch of the subject. It will be encountered time and again as you go though this book.

The latest development that has caused quite a similar division of opinion is the transistor. Some take facts as they are — or have been reported — and say that "the transistor is probably ideal for computers and missile guidance, but it does not replace the tube for high fidelity. Distortion and noise are not as good as in tubes." Others realize the transistor is still relatively new. It has already proved capable of achieving lower distortion and noise level than tube circuits. So it is really only time that is needed to develop more consistent, less expensive transistors and to learn more about using them, before they will be extensively employed for high fidelity.

The hard core of enthusiasts always were, and still are, individualists. They develop and hold preferences and contentions, sometimes with valid reason, sometimes not. Because of this, the newcomer may find the whole thing confusing. What should he get? Three enthusiasts will give him three completely different answers, along with equally good reasons why the others are wrong.

Of course, you could ask as many as you can find, and then work out the averages. In some instances, you would get the right answer. At least you would get the consensus. But the consensus is not always right.

In this book, we set forth the background of most of the things you'll encounter. We don't say this is right and that is wrong, unless there is scientifically irrefutable evidence (we hope). Rather, we tell you the arguments and how they arose, so you can get in the swim and enjoy the fun yourself.

On the "people" side of our opening question, there is an interesting reason for the wide popularity of high fidelity. At least we feel so. In grandfather's day, every family had its family circle. Someone in the family had a "musical bent" and there was always a family piano. In many families there were quite a few amateur musicians, and they had a real musical evening from time to time.

Those days were supplanted by radio, the movies, television. Entertainment became the business of professionals. What most of us did not realize was that this was a suppression of a very human need, the need for self-expression. This may have been temporarily warded off by the kind of movie that allows the viewer to identify himself with the hero or heroine. But this, being an inactive outlet, was little outlet at all.

The home music circle has become a lost art. Beside the performances available on records or over the radio, the family group appeared very amateurish. They became self-conscious about exhibiting their "art".

Reproduced music, which forms the basis for high fidelity, is an art in itself, over and above the art it reproduces—that of the musicians. There is an art in achieving balance, clarity and a satisfying illusion of realism. This does not end at the recording or broadcast studio. The high-fidelity enthusiast can participate in it.

This is a somewhat different enthusiast from the hard-core variety we introduced a minute ago. Theirs, to them at least, is a science. But the new high-fidelity enthusiast, who outnumbers his forerunner many times over, sees high fidelity as an art, a means of enjoying art, a means of participating in it. Most often he does not really want to be interested in tubes and tuners, distortion and dynamics. He merely wants what will give him the best in reproduction, and enable him to satisfy his artistic urge for self-expression.

But there he encounters a hangover — or should it be heritage — from the hard-core enthusiasts to whom he really owes the whole thing. After all, without the hard-core enthusiasts of yesterday, there would have been no high fidelity today. Their strongly held opinions persist in the different answers — and the strange new language — you will get as soon as you ask an "expert" what you need to make your new system the ultimate.

That's why we think you might as well use high-fidelity as an outlet for your self-expression. These days you don't need to become an expert with a soldering iron, or know the latest developments in electronics, to get a passing acquaintance with what goes on. This book aims to make this strange language meaningful to you, so, when an argument starts, you can enjoy the fun — and maybe feed it a little!

Many "science types" have an almost fanatic sense of loyalty to something they feel is an ideal — although really it's only an idea. We remember one argument where a man with certain scientific background started to contend the earth is flat. Believe it or not, for every piece of evidence the rest of us produced to prove it is round, he had an answer. That man left the argument still convinced the earth is flat. But the argument was a lot of fun.

Some people's main interest in politics is knowing all the arguments of both sides. When one meets an ardent Democrat, you'd swear, from the way he argues, he must be Republican. But listen again when the same person meets a Republican. He delivers all the Democrat lines. You can have fun doing the same kind of thing with high fidelity as your subject. But to do it, you must know the arguments. You'll find many of them — more than enough to get started with — in this book.

Then, of course, lots of people get confused by politics: what each party says sounds so reasonable until you hear what the other side says on the same topic; people who have only heard the "sides" are confused. It can very easily be that way in high fidelity. Often, as in politics, there is no absolute black and white about certain issues. We've set out to give you the background of such issues, so you won't be confused.

Why the pictures? We didn't want this book to be technical, in the usual sense of the word. We have a chapter on circuits, but this must not be made dependent on your being able to read an electronic schematic. Some of the pictures are there to help give the idea. Others are just humorous quips triggered by the words, mostly unrelated to the subject matter. It's been our experience that people look at pictures in a book first. Then they read the text to see if they can find what the pictures mean. That's just the idea. Using this approach, the dry-sounding terminology of high-fidelity jargon takes on meaningful significance.

Norman H. Crowhurst

About The Author

MENTION Norman Crowhurst's name to any technician or engineer in the audio—high-fidelity field, and you will get an instant response. His articles have appeared in more than a dozen technical magazines in the United States and Great Britain. In addition, he has written a number of highly informative books on the subject. Although this is his first venture into the non-technical, a great deal of his work has been devoted to "making easy" at various levels. He has contributed much of value in this direction, and to establishing an accurate, as well as simple, concept of things as they really are. Recently, he was awarded a Fellowship of the Audio Engineering Society in recognition of the importance of his writings in the formation of a comprehensive audio technology.

Born and educated (electrical engineering) in England, he had some twenty years' experience in electronics and high-fidelity audio before coming to the United States in 1953. He is now a consultant to some of the most respected firms in the industry, as well as being a widely published author.

Mr. Crowhurst is an Associate Member of the Institution of Electrical Engineers (Great Britain), Senior Member of the Institute of Radio Engineers, and Member of the Acoustical Society of America, the Society of Motion Picture and Television Engineers, and the British Sound Recording Association.

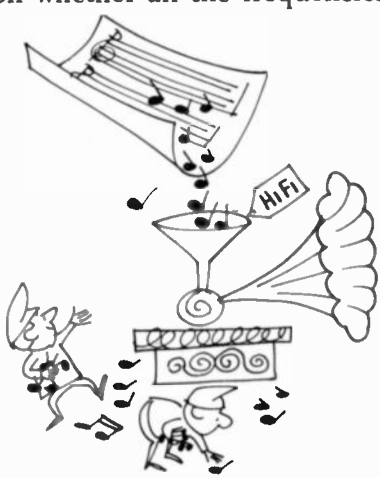
THE PUBLISHERS

chapter

1

frequency response

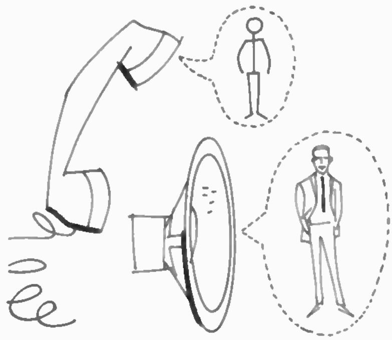
THE long trail toward high fidelity began when it was first discovered that the degree of realism in reproduced music and other kinds of program depended on whether all the frequencies were present. Looking back from the lofty heights of fidelity we have now attained, those early efforts seem poor indeed. But at the time, listening to the local radio station, with a pair of headphones was described as "life-like." "It is as if you were there!"



We were startled by the fact that we could identify sounds, whether of individual instruments or people's voices. Sometimes people are still amazed when we recognize their voices on the telephone. But knowing a person's voice on the telephone, when you have previously heard it directly, is much easier than the other way round. Have you ever been first introduced to a voice over the telephone and later met the owner, without recognizing him? It's happened to me. And there's a reason for it.

Cartoon or portrait?

Have you ever studied the cartoonist's art? It always amazes me how an artist, with so very few lines on paper, can convey such an apparent wealth of detail to our minds. When you examine the drawing carefully, the detail is not there. But little kinks in the right places suggest the detail to our minds and we fill in mentally what really is not there. Making the right optical suggestion is the cartoonist's skill.



With just a few lines, he can create quite a character, who smiles, frowns and expresses the vast array of human emotions we know. But we recognize these emotions in the cartoon character only through familiarity with them in real people. In fact, that is the real appeal of the "comics".

If we were not familiar with the things that happen and the emotions felt and expressed by real people, cartoons would mean nothing to us. The skill of the artist gives the merest suggestion that enables our mind to fill in the detail from the familiar storehouse of our memory. We do not even realize we do it.

This explains why you may recognize a voice over the telephone of someone you have previously met, but not vice versa.

The telephone does not really carry *all* of a person's voice, only certain parts of it – a caricature. Enough to enable the listener to understand what is said, and usually to identify the person speaking. Through familiarity, from having met the owner of the voice or being familiar with the language spoken, you subconsciously "fill in" the missing parts. You can even "see" the smile on his face sometimes, by the intonation of his voice.

When your only contact with a person has been by telephone, you do not really know so much about him – although you may *think* you know his voice. This is because you have never really heard *all* of his voice. So when you meet him in person, you are surprised. The additional details are not what you imagined at all. You may even fail to recognize him, or think he must be someone else you have never met.

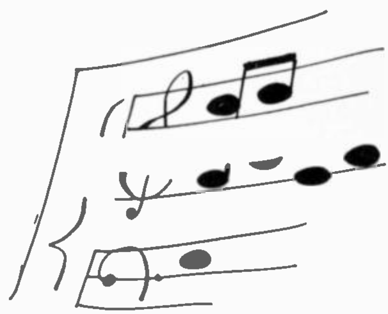
The difference between telephone communication and high-fidelity reproduction is very much like that between a cartoon and a portrait. A good portrait is a lifelike image of the original. From it you could recognize an original you had only seen in portrait. A caricature or cartoon of someone you know you may recognize. But you could not get a lifelike picture of someone you have never seen (either in life or portrait) from a caricature.

The missing details

What are the missing qualities about voices and music — the ones we hear in personal contact but that get stopped by a telephone, or those early radio headphones? Primarily, certain *frequencies* are missing.

This can be illustrated by an ambiguity that often happens when you try to give spelling over the phone. "F for Freddie?" your listener may ask. "No, S for Susie," will be your reply. Why should it be so difficult to distinguish these two letters, *s* and *f*, over the telephone, when they sound quite different in direct conversations?

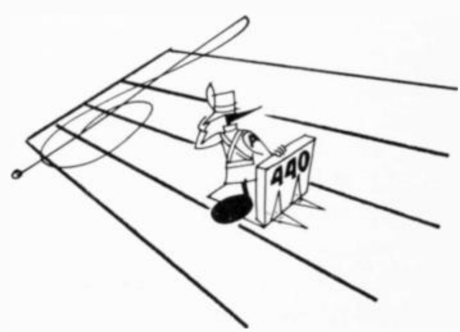
Because the telephone, among other things, can drastically cut down the higher frequencies in the sibilant *s* sound. When you name the letters *s* or *f*, all that really comes over the phone is the e part of *e*ff or *e*ss. There is not enough of the sibilant sound that is really *s* to be sure which is which.



This does not usually bother you, because you recognize words by the part of them that you can hear, and don't even realize part is missing. Fortunately, there are no such names as "Sreddie" or "Fusie" so using the appropriate letter that way settles the matter in your mind.

What are frequencies?

Whether a sound is musical or not, it is made up of a whole range of frequencies. Have you ever stood near a piano and sung a steady note? The corresponding piano string will go on "singing" after you leave off. The string that sings has the same *frequency* or *pitch* as the note you sing. Pitch in music is the same as frequency in high-fidelity lingo. A low bass note on the



piano has a low frequency, while a high treble one has a high frequency.

Frequency refers to the number of vibrations every second that produce the musical effect we identify as that particular note. The tuning note A in concert pitch corresponds to a vibration rate of 440 times per second, called 440 *cycles* in high-fidelity jargon.

What makes music?

An interesting number pattern exists between frequencies in musical notes. For example, the note A one octave below the tuning note has a frequency of 220 cycles, exactly half of 440. The note A two octaves down has a frequency exactly half again, or 110 cycles. Going up, the next A above the tuning note is twice the frequency, or 880 cycles. Two octaves up brings us to twice again, or 1,760 cycles.

Did you ever wonder why notes an octave apart sound almost like the same note over again? It's because the strings do not vibrate at just one frequency. Each one vibrates at a whole range of frequencies, related by this interesting number pattern, called a *harmonic relationship.*

For example, when the tuning note A is struck, not only 440 cycles is produced. At the same time, in lesser quantities, the string vibrates at twice, three times, four times the rate, and so on, giving frequencies of 880, 1,320, 1,760, etc. The strongest, next to the *fundamental* of 440, is the *second* harmonic (so-called because it is the fundamental multiplied by 2), or 880 cycles. This explains the marked similarity when two notes an octave apart are played either together or in succession, as compared with any other note combination.

The next note, G, has a frequency of 392 cycles. Because this does not fit the regular pattern of note A's frequencies, it sounds like a different note, a little lower, because the frequency is lower. But the note G will likewise contain a whole family of frequencies, twice, three times, four times, and so on, giving 784, 1,176, 1,568, etc. Although these numbers follow the same "pattern," they are not related to those for note A.

And other sounds?

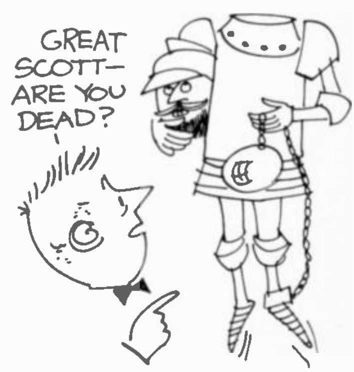
It is this harmonic relationship between the frequencies produced that makes the sound musical. Nonmusical sounds do not have this definite number pattern. But they are made up of frequencies.

For example, the sibilant *s* sound is made by blowing air between the teeth. This causes very rapid vibration of the air molecules at this point, making the sound. The drummer's wire brush on a "washboard" makes a quite similar sound. Both sounds are characterized by a vibration rate between 3,000 and 5,000 times a second, but not at a regular rate that would give them a musical tone identity.

However, these sounds can have a sort of tone coloration, as you can verify for yourself by changing the position of your lips as you make the *s* sound with your teeth. The rushing sound you can hear by holding a seashell close to your ear has no magical connection with the sea. The cavity inside the shell merely emphasizes certain frequencies in the general noise going on around us all the time, to give this effect. This is very similar to the way your lips change the tone of the sound produced by your teeth when you hiss.

A hiss is just one kind of nonmusical sound. There are many others: motor noises, rattles and, not the least, speech. All this variety of sound is characterized by containing groups of frequencies, each unique to its own sound, but without the special number pattern that makes a sound musical.

Stating the obvious

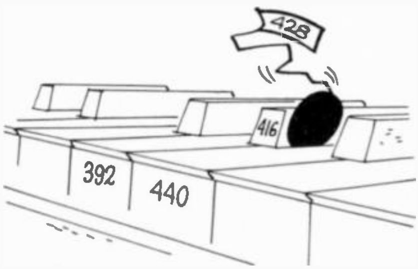


In view of the fact that all kinds of sound consist entirely of varying quantities of the different frequencies within the audible range, it seemed axiomatic (or in everyday language, it didn't need proving) that the basic requirement for fidelity is quite simple: All the component frequencies of any sound to be reproduced must be kept, and in the same proportions as the original. Viewed this way, fidelity seems to be entirely a matter of *frequency response.*

This is not just a matter of making sure the high-fidelity system covers the required frequency *range.* It must do more than this. It must respond *uniformly* to all frequencies within the range. To illustrate:

The frequencies between

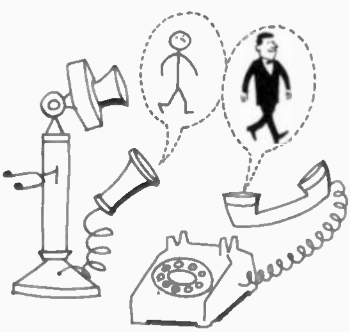
A piano responds to a singing voice in its vicinity by having the



string for the same note sing in unison. But each string sings at its own specific frequency, and possibly some harmonics or multiples of that frequency. The string for A will sing at 440 cycles. The G string sings at 392 cycles, while the string for G♯ (or A♭ is halfway between and sings at 416 cycles. Suppose the note you sing happens to be 428 cycles, what then?

Because you have split the difference between A and A♭, both of them will sing a little because you are so close to both of them. But the piano itself cannot sing at 428 cycles, because it does not have a string tuned to this frequency.

So, while a piano will respond fairly uniformly to any note for which a particular string is tuned, it will respond only at these notes. It cannot respond to *any* frequency within the audible range. However, this is what a high-fidelity system is expected to do.



Sonic caricature

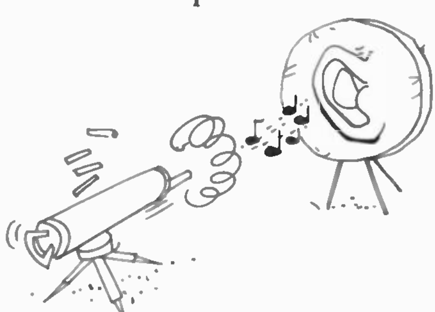
The telephone instruments, microphone and earpiece, use simple disc diaphragms, with various holes and cavities arranged so they respond to selected frequencies throughout the speech range. This is why everyone's voice tends to sound similar over the telephone, and the individuality does not come through as well as it does over a high-fidelity system.

While the piano and the telephone, in different ways, utilize the fact that their parts are "tuned" to specific frequencies, each to achieve its own purpose, a high-fidelity system must do the opposite. Tuning parts, whether strings, telephone diaphragms, organ pipes or whatever, to respond to specific frequencies is called *resonance*. The job of a high-fidelity system is not to introduce any frequencies of its own, but to render a faithful copy of whatever frequencies are in the sound presented to it to reproduce.

Early telephones were not as good at getting the message through intelligibly as modern ones. Bell Telephone Laboratories did a large amount of work to find what was needed to get messages through the human hearing faculty to the brain. They wanted to know what is the essence of good, recognizable, sonic caricature. What they found, and what later workers repeating the experiments in greater detail have found, has been helpful for high fidelity, as well as in making better telephones.

Frequency range

One of the first questions was: What range of frequencies can we humans hear? These experiments turned up some facts that



surprise us when we learn about them for the first time. The first surprise may be the fact that there are sounds we cannot hear. The natural reaction, if we cannot hear anything, is to think there isn't any sound to hear. But there can be sound whose frequency is too high or too low to hear, or it may just not be loud enough to be audible.

However loud the sound is made, though, there are frequency limits to everyone's hearing. These limits differ from person to person, so we cannot quote specific frequencies in answer to the question.

At the top end, the frequency limit varies from a little below 10,000 cycles to almost 20,000 cycles. Most people can hear 10,000 cycles, although some have difficulty, and very few can hear 20,000 cycles.

At the low-frequency end, the test gets more complicated. Below about 50 cycles it is difficult to get pure enough notes to tell. If a 20-cycle note has second harmonic, 40 cycles, along with it, you may be hearing a 40-cycle tone and thinking you can hear 20 cycles. Few people can hear 20 cycles by itself. Many can hear 30 cycles, if it is loud enough. Practically everyone can hear pure 40 cycles, but it has to be at fairly high volume to be audible.

Disappearing bass

The tests at the low-frequency end revealed another interesting fact that you have to experience to believe. Loudness makes much more difference to the audible range at the low frequencies than at the other end. The cause of this is connected with something

that had already been noticed, but had been blamed on the equipment.

As you turn down the volume control on a radio receiver or high-fidelity set, the bass seems to disappear. It is not due to a deficiency in the radio or hi-fi, but to the natural characteristic of human hearing. This fact also explains why many high-fidelity enthusiasts feel they must operate at high volume, so they can hear the full frequency range. Unfortunately, the neighbors don't always have the same taste in program material!

Can you trust the obvious?

Here we encounter the first challenge to the correctness of what we at first took to be axiomatic. When a high-fidelity program is

played at reduced volume to avoid annoying the neighbors or so we can enjoy the music ourselves *and* think at the same time (although some enthusiasts will mutter 'sacrilege' at such a suggestion), the bass seems to disappear. This happens when we reduce the volume of all the frequencies *in proportion,* which is what we assumed *ought* to be done.

But *music can be played quietly,* with all the frequencies still audible. Small ensembles in restaurants do it all the time. So why cannot high fidelity be played quietly, too, without losing some of it?

Music can be played quietly because musicians have ears like other people (a "musical ear" is not basically different from any other kind). When a bass player reduces his volume, he judges it the same way as anyone else would, not by turning down an imaginary hi-fi volume knob. Without realizing it, this means the bass player will reduce his actual volume of sound to a lesser extent than other instrument players, although the *apparent* loudness reduction is consistent with them.

Smaller or farther away?

This is the argument for using loudness compensation in a high-fidelity system. If proper loudness equalization or compensation is used, turning down such a control will give the effect of the music's being played more softly, rather than its being removed to another room or taken farther away.

A few die-hards, who still believe proportionate representation of frequencies is axiomatic, argue against such loudness compensation. If they're so democratic about their high-fidelity systems, they should investigate the principle of proportionate representation in government sometime!

If all groups are to be represented, then some minor groups will need more than strictly proportionate representation, otherwise they would not even rate one representative. Similarly, some adjustment to the theory is needed so the deficient bass frequencies do remain audible.

Hi-fi for bats?

Another question that frequency response introduces is whether frequencies beyond audibility help the realism in any way. In the first place, the reason systems providing responses 'way beyond audibility got started was a matter of advertising claims — the endeavor to outdo competition.

Early reproducers did not even cover the full range. When the bass was added by dynamic (moving-coil) speakers mounted in various baffles, and when tweeters added the extreme high frequencies, a definite improvement in realism was noted. Claims for frequency response got progressively wider, at first approaching the full audible range, then slightly exceeding it and finally going a long way beyond it.

The idea seems to be that a system that goes from 10 to 50,000 cycles must sound better than one that goes only from 20 to 25,000 cycles because the numbers say so. This, even though we can only hear, for example, 30 to 15,000 cycles as an average, and 20 to 20,000 cycles as an absolute maximum. Those extra frequencies are strictly for dogs and bats.

Some have "explained" that we are conscious of these extreme frequencies, even though we cannot consciously *hear* them. To "prove" their point, some have demonstrated audible differences between a system that responds only to the audible range and one that responds far beyond it.

The magic of sound

But there is one thing we should learn as soon as possible in this high-fidelity game. And that is that demonstrations do not always prove what their sponsors claim. In this particular case,

we will not deny that a *difference* may have been heard and that it favored the system with frequency range far beyond audible limits. But was the improvement due to presence of these inaudible frequencies, or to something else more or less coincidental?

One reason put forward for the difference observed is the possibility that a system that has a frequency response extending far beyond the necessary limits will have a more uniform response *within* them.

This could be a reason, but it would not explain all of the tests. Some were made without changing any of the components, just by inserting electrical filters that removed all frequencies beyond a certain point. If the difference is not due to removal of these frequencies, then what could it be due to?

The important transients

The answer to this is a change in transient response. When the piano responds to your singing a certain note, the loudness of its response depends, not only on how loudly, or how close, you sing, but also for how long.

If you give only a very short *staccato* burst of the note, the piano will hardly get started responding. If you sing for about a second on the note, it will respond about as much as it ever will. For intermediate durations of your singing the note will produce intermediate degrees of loudness in the piano's response.

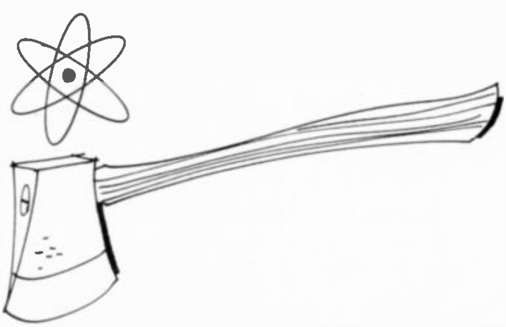
Also you will notice (if you hold the loud pedal of the piano down as you do this so as to keep the *check* pads off the strings) that the note sung by the piano string is sustained for quite a little while.

Not only does a piano not respond to all frequencies, only to those at which strings are tuned, but it takes time to respond. While you sing it builds up its response and, after you leave off, it takes time to die away. In a high-fidelity system, this would be poor transient response.

Reproduced sound should start the notes at the full volume of the original, not take time to build up to it. They should die away like the original, not take longer. The importance of responding, not only accurately to the different frequencies, but with uniform promptness, will be seen more and more as we delve deeper into high fidelity.

Don't use an electronic ax

Electrical filters, used to "remove" all frequencies above a certain point, unfortunately do introduce transient distortion. Fre-



quencies nearer the limit of those allowed to pass get progressively more delayed in their build-up and die-away time, compared with the original. Because this is a noticeable difference, it is observed when the test is made. It is not due to the fact that ultrasonic frequencies are no longer there, but to what has happened to the upper *sonic* frequencies.

But can the frequencies beyond audibility be left out without causing this kind of trouble? Actually, the answer to this hinges on what you mean by "left out." The trouble arose because the filters aimed at removing the ultrasonic frequencies almost completely.

If the filter was designed for 15,000 cycles, then 14,000 cycles is undiminished while 16,000 cycles just isn't there at all. Frequencies closer to 15,000 cycles really have a bad time—they don't know whether they're supposed to be "in" or not, and go crazy trying to find out.

There's no need to be so drastic. The important thing is that response to audible frequencies should be uniform. Beyond that the response may die as it chooses – it doesn't have to "drop dead"! But to make sure it doesn't, the response need not extend to 50,000 cycles or higher, either.

Watch those transients

Accurate handling of transients is important not only at very high frequencies, it is equally necessary throughout the audible range. All of them have different ways of starting and stopping

that are characteristic of the sound to which they happen to belong at the moment. This should be accurately copied by the high-fidelity system or the resulting reproduction will lack realism.

There are many reasons why transient performance may suffer at various frequencies, which we shall discuss in more detail as we come to the offending parts.

A frequency who's who

The basic identity of frequencies can be deduced from the values of the musical scale. As we have already stated, the A used for concert-pitch tuning is 440 cycles. Middle C, the note that divides bass from treble on the musical scale, is 261.6 cycles. Top

C, the highest note on many piano keyboards, is 4,185.6 cycles, while the lowest note, usually A, is 27.5 cycles. The E string of a double bass, played open, gives 41.2 cycles while that of a guitar is 82.4 cycles.

The deep bass isn't always

Many instruments with bass notes go down to the same fundamental frequency of 41.2 cycles. An accordian is a good example. But its tones, like those of many other instruments, have more

harmonics than fundamental, so the ear hears much more of 82.4 cycles, 123.6 cycles, and so on, than it does of 41.2. Here we find another of the "hearing is believing" surprises.

All of us have a subconscious musical "training," whether we have had lessons or whether we have only listened to any music that came our way since we were born. This subconscious training, or "conditioning" as it is called, causes us to identify that note as the one corresponding to 41.2 cycles, even if we cannot hear that frequency at all. Our hearing bases that deduction on the number pattern of the frequencies we do hear.

What is treble?

Practically no musical tone ever goes above that 4,185.6-cycle note on the piano — in fact, that one is very seldom played! Musically, treble starts at 261.6 cycles. Fundamental musical tones

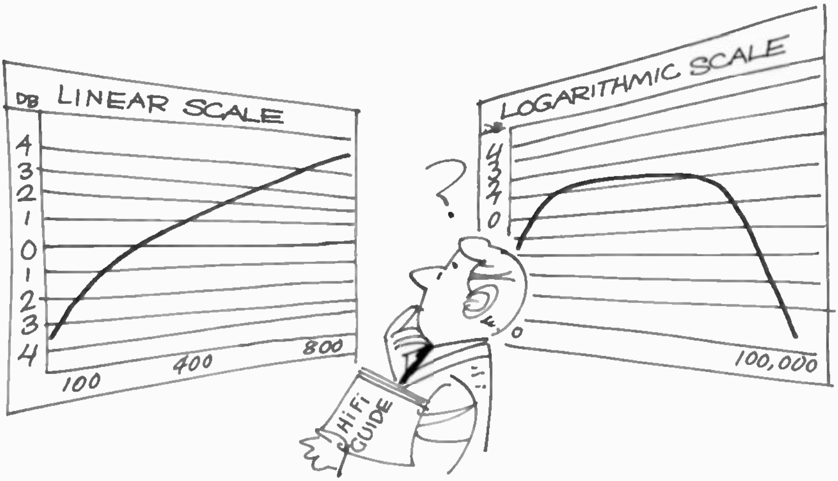
do not often go higher than somewhere between 1,000 and 2,000 cycles. The percussion effects — cymbals, triangle, the washboard we already mentioned — dominate in frequencies from 3,000 to 5,000 cycles. The intelligibility and personality parts of speech also lie in this range.

Thus treble consists of frequencies from 261.6 cycles up to somewhere between 1,000 to 5,000 cycles, according to whether you call percussion music. So why does high fidelity need frequencies up to 15,000 or 20,000 cycles?

Because harmonics and other "character-forming" parts of the sound come in the range from 1,000 to 15,000 cycles. Extending the range above about 5,000 cycles will not make any new sounds audible, but it makes them more accurate. Without them, high-frequency sounds, due to percussion for example, become indefinite and confused in just the same way that removal of sibilants from telephone transmitted speech causes confusion between *s* and *f*.

Why the fancy scales?

At one time, much capital would be made of a system that extended the range "by another 2,000 cycles" — from 14,000 to 16,000. By the numbers, this would appear to be a 15% improvement.



ment. But the modern practice of putting frequency on a logarithmic scale, and thus making the distance between 50 and 500 cycles the same as that between 500 and 5,000 cycles, has changed this. Judged by this spacing, the same improvement looks like less than 2.5%, which it really is.

The logarithmic scale fits in with the musical values. Each octave seems like an equal tone interval, but the frequency is always in 2-to-1 ratio. The logarithmic scale shows it this way.

By a *linear* scale of frequency, a response from 1,000 to 16,000 cycles would look good. The piece from 0 to 1,000 is only 1/16 of the range. But in reality this "1/16" includes practically all of the useful musical scale!

The popular frequency range, usually regarded as the hi-fi range, from 20 to 20,000 cycles, contains approximately 10 octaves. Starting with 20 cycles, successive doublings give frequencies of 40, 80, 160, 320, 640, 1280, 2560, 5120, 10,240 and 20,480. Although only the range from 40 to 5120 — at the outside — can be taken to have musical significance, this whole succession of 10 steps represent apparently uniform intervals of pitch.

Each octave is divided into 12 semitone intervals, according to Western musical standards, and each of these intervals is an approximately equal frequency ratio. The ratio is about 1.059. Multiply 1.059 by itself 12 times, and the result is almost exactly 2 – the octave.

In older books on sound and music, the relationship between frequencies is given an importance exaggerated almost to the point of magical significance. To aid in this, an old-fashioned scale, called the diatonic, is used. In this scale, all the notes have a simple relationship one to another, instead of the decimal value 1.059, which is the twelfth root of 2. In the tabulation given here, the frequencies based on this diatonic scale are compared with those according to the modern tempered scale.

Some instruments use the true tempered scale. Examples are the electronic organ and some pianos.

Table of frequencies for one octave

note	diatonic reference ratio	frequency in diatonic scale	frequency in tempered scale
C	1	261.6	261.6
C♯	17/16	278	277.2
D	9/8	294.3	293.6
D♯	19/16	310.6	311.2
E	5/4	327	329.6
F	4/3	348.8	348.8
F♯	17/12	370.6	369.6
G	3/2	392.4	392
G♯	19/12	414.2	416
A	5/3	436	440
A♯	7/4	457.8	465.6
B	15/8	490.5	494.4
C	2	523.2	523.2

A piano tuner uses what is officially the tempered scale. But a truly tempered scale makes a tune sound the same whatever key it is played in. Most pianos give an impression of flats and sharps, according to the key used. A tune played in flats has a mellow sound, while the same tune played in sharps give a brilliant sound. This is because the tuning is not strictly according to the figures in the tempered scale.

In the old diatonic scale, chords in open key, C or F or G have perfect harmonic relationship: in each the second note is 5/4 times the first and the third is 3/2 times the first. If you want to check, these chords are: for C: C, E, G, C; for F: F, A, C, F; for G: G, B, D, G. This was claimed to account for the smooth sound of chords in open key, while the deviation from exact ratio accounted for the sharp or flat sound of other chords.

But, for a long while, instruments have been tuned much nearer to the tempered scale, so this no longer provides the correct explanation. The fact seems to be that the human ear does not make a too critical comparison of ratios, with the exception of the octave, which is such a simple one. What it does notice is the interval, *relative to the intervals to which it has been educated.* The human hearing is very sensitive to pitch. It is even more sensitive to *relative* pitch, which is what any change of frequency is; either when a note changes or when different notes are played together.

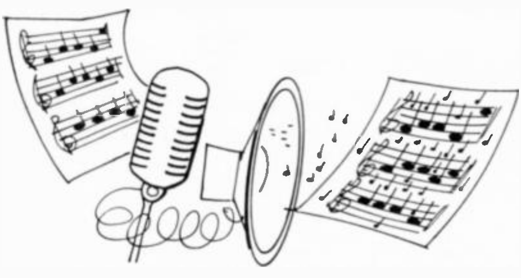
Any published data about frequency response uses (if it means anything, it does!) another strange expression, the decibel, contracted to db and usually said that way too. In this chapter we have contented ourselves with showing what frequency is, why it needs to be uniform and a few other things like that, without getting involved in the details of quantity—*how* uniform. As this really pertains to another property of our hearing faculty, as well as another feature of a high-fidelity system, an explanation of decibels and all they entail is held over to Chapter 3.

chapter 2

distortion

SOMETIMES the electrical or electronic transmission of sound is explained so as to make it appear deceptively simple. One such explanation tells us that the microphone converts the sound-pressure undulations into corresponding electrical-current fluctuations, which are then faithfully amplified, to be converted back to sound-pressure undulations again by the speaker. That's a lot of things happening to explain so simply.

From that simple explanation, one would visualize a conversion of a sound waveform into facsimile electrical currents, which are



then amplified and converted with precision fidelity back to sound waves. But it is not at all that simple. The previous chapter showed that the reproduction could be spoiled if some of the frequencies in the original sound got lost somewhere along the line. In this chapter, we find the reverse action to be a possibility — that some frequencies may appear in the reproduction that were not in the original. That is called distortion.

What is distortion?

Strictly, anything that changes the waveshape at all might be called distortion. But waveshapes can be changed by adding or subtracting any of the frequencies present. Just altering the timing between them will change the shape. Small amounts of such time differences, called phase shift in the jargon, and even some change in the relative size of different frequency components, can be accepted without being *audibly* detected. For this reason they are not usually called distortion in the hi-fi sense.

The word distortion is reserved for anything that gets into the waveform that should not be there. If you hear a nice clear musical program, and then hear the same thing reproduced with a lot of harsh grating sounds, you know there is distortion all right. But even a little of this kind of distortion can spoil high-fidelity reproduction.

Getting after distortion by just listening for it, as something that should not be there, can become very involved. Did the trumpeter blow a bad note, or was that distortion? Most of the work toward eliminating high-fidelity distortion has so far been achieved by measuring it, rather than just listening to it.

Distortion classified by order

The simplest way of measuring distortion is by using a single-frequency test tone, produced by an audio oscillator, and then

checking the output for any other frequencies. What other frequencies may appear are determined by what is called the *order* of the distortion. The full mathematical explanation of how this works is quite complicated, but with some simple numbers it's not difficult to grasp.

Suppose the single tone is the tuning note A, of 440 cycles. For first-order distortion, you just write the same figure again, underneath it, so it could be added or subtracted,

$$\begin{array}{r}440\\ \underline{440}\end{array}$$

If you add these numbers, the result is 880. If you subtract, the result is zero. Zero frequency is direct current in electrical circuits, and just air in continuous motion from the acoustical viewpoint, neither of which contributes anything audible. But 880 cycles is the second harmonic of 440.

For second-order distortion, repeat the same figure twice more under the first,

$$\begin{array}{r}440\\ 440\\ \underline{440}\end{array}$$

In this case, the three numbers may all be added, or two of them can be added and one subtracted. The first gives 1,320 and the second 440. A frequency of 1,320 cycles is the third harmonic of 440. The 440 component of distortion is indistinguishable from the original 440, being the same frequency.

The number of tones involved

But suppose you used two tones and did the same thing, adding or subtracting any combination of numbers and taking one from each line. Suppose we use G and A. For first-order distortion, writing them the same way,

$$\begin{array}{rr}392 & 440\\ \underline{392} & \underline{440}\end{array}$$

Writing out the different possibilities, we have:

392	392	440	392	440	440
+392	+440	+440	−392	−392	−440
784	832	880	0	48	0

The zeros, as before, do not represent audible distortion components, but we now have 784 and 880, the second harmonics of 392 and 440, respectively. We also have 832 and 48, which are not directly related to either of the original tones. These are the first-order intermodulation-distortion products.

For second-order intermodulation possibilities, we write the same numbers in two extra lines,

392	440
392	440
392	440

Adding, different ways, taking one number from each line, we can get 1,176, 1,224, 1,272 and 1,320. Subtracting, the possibilities are 344, 392, 440 and 488. The 392 and 440 components are indistinguishable from the original frequencies. The 1,176 and 1,320 components are third harmonics of 392 and 440, respectively. Then 344, 488, 1,224 and 1,272 are second-order intermodulation products, not directly related to either frequency by itself.

Harmonic and intermodulation

These are the possibilities, just taking two frequencies and the first- and second-order causes of distortion only: 10 distortion components have appeared: 4 harmonic—784, 880, 1,176 and 1,320

cycles, and 6 intermodulation—48, 344, 488, 832, 1,224 and 1,272 cycles.

This may seem complicated enough, but related to actual possibilities in performance it is simple! Music contains many more than two notes at once, as a rule, and there is no guarantee that distortion stops at the second order. Addition of extra components, in either the number of program frequencies present or the number of distortion orders, very quickly multiplies the number of distortion frequencies possible. So we can only come to one conclusion: distortion is a very important thing to do without!

On the warpath

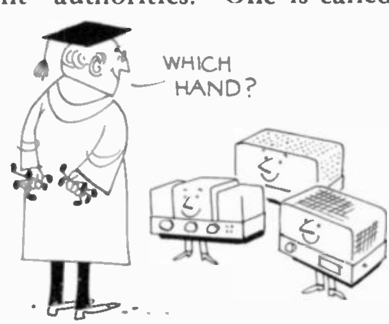
Unfortunately, though, no system, or any part of one – microphone, amplifier, speaker, pickup or what-have-you – is *completely* without distortion. If we want to know how good a system is with regard to distortion, we need some way of measuring it rather than relying on an estimate that might be prejudiced, "I thing this one has less distortion than that one."

If you use only one test frequency and then measure the output for any *other* frequencies that might appear, you will find harmonic distortion. Using 440 cycles and assuming only the first two orders of distortion, you may find two frequencies, 880 and 1,320.

If you use two test frequencies and then measure the output for any other frequencies, you will find, in theory at least, both harmonic and intermodulation components, all 10 of them for just the first 2 orders. From this, it would seem obvious that the intermodulation test is much better than the simpler harmonic test. But there are practical limitations.

Pick your frequencies

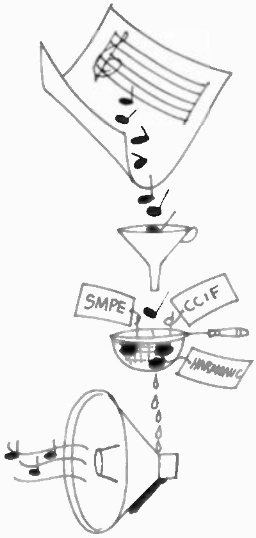
You have to decide *what* two frequencies to use. Two basic possibilities are favored by different "authorities." One is called the SMPE method, because it was first standardized by the then Society of Motion Picture Engineers. This uses one low frequency and one high one; for example, 50 and 2,000 cycles. Let's see the possibilities with this method first.



First-order components would be 100, 1,950, 2,050 and 4,000 cycles (apart from the zeros). Second-order components are 150, 1,900, 2,100, 3,950, 4,050 and 6,000 cycles. While all 10 of these frequencies are produced (and more if higher-order distortion is present), the meter normally used does not measure all of them.

One might think the simplest thing to do would be what the harmonic meter does — just remove both the original frequencies, 50 and 2,000 cycles, and measure what's left. But very few, if any, meters work that way. With the idea of making operation easier, they remove whole blocks of frequencies, with filters.

As a result, the frequencies usually found by the SMPE intermodulation test will be only those near to 2,000 cycles: 1,900, 1,950, 2,050 and 2,100 for the first and second order. Only 4 out of a possible 10!

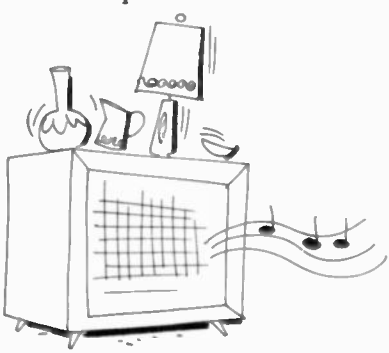


The other intermodulation test, called the CCIF method, uses two relatively high frequencies; for example, 4,000 and 4,500 cycles. In this case, first-order components will be 500, 8,000, 8,500 and 9,000 cycles. Second-order components are 3,500, 5,000, 12,000, 12,500, 13,000 and 13,500 cycles. But the practical difficulties of removing just the original frequencies are greater with this method than with the SMPE.

So the usual method adopted is to "look" for specific distortion components, with a simple filter, usually designed to pick off only the low frequencies well below the original ones. This means only 500 cycles, in this example, is found.

What have we caught?

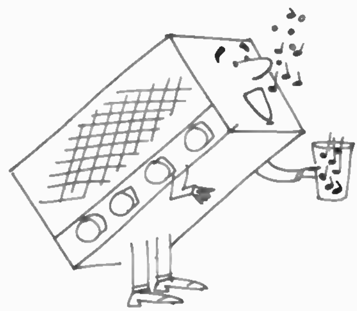
The CCIF method of measurement finds only one of the first-order components — the other nine components, including all the second-order ones, are ignored. So why would anyone want to use this method? Actually, both tests show up different kinds of distortion. So each one has its useful place.



Some buzzes

In the CCIF method, two high frequencies are used, which produce just one low-frequency intermodulation component. This may be only 1 of the 10 possible products of this combination (taking only the first 2 orders) but it is an important one, because it can be *very audible*.

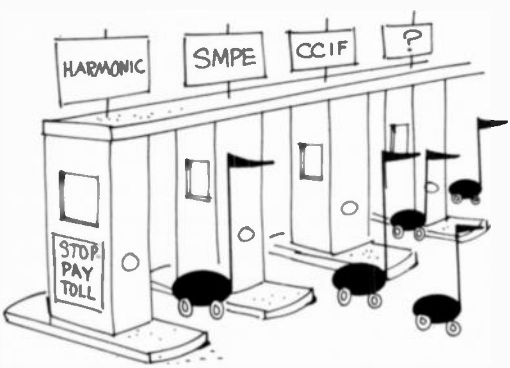
If notes G and A are played together, for example, this component will be 48 cycles. It appears as a low-frequency buzz, whose only connection with the music is that it appears every time those particular two notes get played at once.



And some gargles

The SMPE method uses one low frequency, which may be 50 cycles or some other one, and one high frequency. The components normally measured are all intermodulation components involving both frequencies. But while the low frequency may have one or two parts in the figuring, the high frequency always has one part only. This means the measurement is sensitive to the kinds of distortion produced by the amplifier, particularly the lower test-frequency waveform.

This method measures an effect that can appear in music, caused by distortion that occurs when low notes are played and evident as a fluttery effect in the reproduction of other notes high up the scale.



So we have two completely different intermodulation tests that are related to correspondingly different distortions that can happen to actual program material.

But why so many tests? Presumably an amplifier or other component that would produce one form of distortion would produce all of them. Would not one of the tests be enough to show relative freedom from all forms of distortion? Unfortunately, no!

Why not one overall test?

Amplifiers may produce different proportions of the different orders of distortion, and these may get severe at different frequencies. So it happens that there is no consistent relationship between readings obtained with one method and another. In fact, there is no predictable connection between readings taken by the *same* method — say the SMPE — if the frequencies are changed.

Not only this, but the kind of distortion most prominent in an amplifier, or other part of a hi-fi system, will also vary according to the volume level. Distortion that accompanies low-level, quiet passages is usually — but not always — of low order, mostly first and very little above second. But distortion at high-volume-level, loud passages is higher order, second and upward. So comparison between systems, made at a certain volume level, is no indication what comparison between the same two systems would show at another volume level.

It's always better to have less

One thing, of course, is important: the smaller any distortion is, the less likely is it to be audible. This simple principle has led to a struggle to get distortion figures as near to zero as possible.

When distortion first began to be measured, 5% at maximum output was considered a good figure. Today, whichever method of measurement is used, that would not be considered a very good figure. About 1% is considered satisfactory, while many amplifiers have distortion figures down to 0.1% and even lower.

Does reduction in the figures always mean improvement?

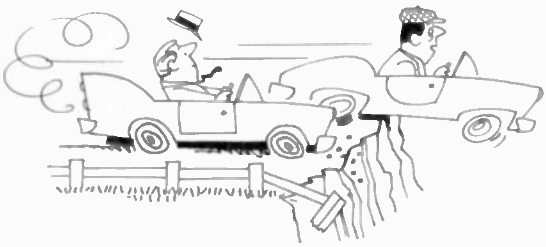
A few facts often seem to be overlooked. Very few speakers achieve distortion anywhere near as low as common figures for amplifiers. Phonograph pickups are not much better. A "low-distortion" speaker may produce in the region of 4% or 5%, particularly at the lower frequencies, while "average" figures exceed this considerably. A really good pickup may be lower than 2%, but very few are.

Under these circumstances, one would expect any amplifier with less than 1% distortion to be inaudibly different from any other that good, because the distortion of either would be swamped by speaker-caused distortions. Yet this is not so. Quite audible differences between such amplifiers can be found.

In fact, often the amplifier with the lowest *measured* distortion produces more *audible* distortion, at least when both are compared in one particular system.

Why the vice versa?

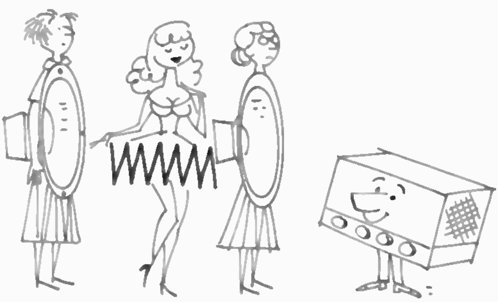
The truth seems to be that designers have been so preoccupied with measured distortion that they have overlooked some very



important causes and forms of distortion. Some of these have more bearing on, and are more basic to, real high fidelity than the acquisition of virtually meaningless distortion figures obtained by "standard" methods. So let us take a look at some of them.

That dumb load!

One of the most obvious things (at least when you know about



it) is the fact that all tests are made with the amplifier feeding its power into a "dummy resistance," which gets hot as a result but does not emit a note. Unfortunately, no one has found a way of making a dummy resistance produce program sound, so we have to substitute a speaker when we want to *listen* to the output power. And no speaker is electrically identical with the dummy resistance.

Just because they both bear a label "16 ohms" (or some other appropriate number) does not mean the impedance to which the amplifier feeds its power is identical. The resistance measures its label value (or close to it) at any audible frequency you want to name. The speaker will measure a different value at almost every frequency. This measured value may vary over a range of about five times.

Crazy speaker impedances

If the speaker's impedance label says 16 ohms for example, (other cases are likely to be in proportion) one speaker may measure this at about 400 cycles. At its low-frequency resonance, usually about 100 cycles, it probably measures about 80 ohms, which it will also do again at about 8,000 cycles or higher. The measured value is continuously changing as you try different frequencies.

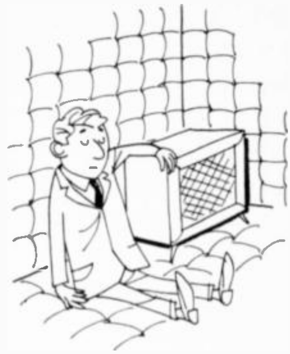
Another 16-ohm speaker may measure what the label says at, say, 60, 135, and 2,350 cycles. But at 400 cycles it measures only 6 ohms; at 90 cycles it measures 35 ohms, which it also measures at 5,200 cycles. This too changes rapidly as you vary frequency.

Whichever of these two speakers you connect to the amplifier, the power it delivers will not be the same as when the dummy resistance was connected for test, neither will its distortion. But nobody ever measures the distortion an amplifier gives when it feeds a speaker.

Why the fictitious figures?

Why not? Would it not be more logical to measure distortion as you hear it, than when the power is used only to warm up the dummy resistance? The reason is not far to seek. Manufacturers want figures that mean something, if possible.

It's difficult enough to make any meaningful comparisons between distortion measurement methods. But if you measure the distortion with a speaker, you will likely get a totally different figure with every speaker, and the result will change much more critically as you vary the test frequency.



At least, by using a dummy resistance, the result can be repeated or verified anywhere, anytime. If the amplifier does not give what the specification says in the way of permissible distortion at maximum power, something is wrong. To measure the same thing with a speaker introduces so many variables you could never be sure of anything.

There have been some attempts at making other checks to find how much the amplifier will change its performance with different speakers connected, but so far none of them is too meaningful.

As well as differences between the way the amplifier is connected to measure its distortion and when you listen to it play music, there are differences because it is not tested with music.

Why not a music test?

The test method uses one, or at most two, frequencies at one time. Music contains a whole range of frequencies all the time, and these are continually changing. Here are two differences. As

we saw earlier, the presence of many frequencies will rapidly multiply the number of distortion components. But if the distortion is small, then *all* the components will be small.

The other difference is responsible for many of the audible distortions that do not show in measurements: musical waveforms are constantly changing. Different frequencies are being started all the time, as the various musical notes are played. Distortion of the kinds we discussed is always measured while the tone or frequency (or two frequencies) is steady — a condition that does not represent much of a musical program.

Transients

Certain distortions happen when these changes occur, particularly when each tone starts, whether blown, plucked, struck or bowed by the original instrument. In high fidelity, these parts are called transients, from the fact that a change occurs as the sound wave for the tone gets started. They can be distorted by the amplifier, causing extraneous "noises," as well as by erratic frequency response during them.

When this fact was first realized, tests were sought for measuring the effect. One of the first was the square-wave test. Musical tones, after the initial startup time, have a smooth back-and-forth waveform or movement, the frequency of which fixes the tone. The pure test tones used for harmonic or intermodulation measurement have no sharp corners on their waveforms.

But a square wave can be regarded as a kind of repetitive transient. The electrical wave consists of a steady voltage for half the period of a complete cycle, a sudden change to another voltage, at which the waveform stays for the other half before suddenly changing back to its original value. So it is like a succession of sudden changes, occurring at twice the frequency of the square wave.

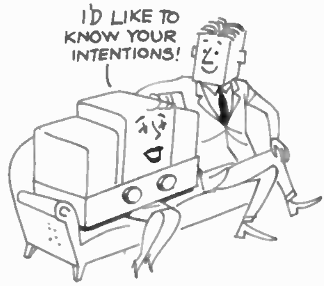
After the square wave has been passed through the amplifier, or other parts of a hi-fi system to be tested, it is displayed on an oscilloscope screen. Much equipment fails to give a true square wave again at the output for various reasons. Some do give a fairly close copy of the original.

Do square waves mean anything?

But does ability to "handle" a square wave indicate good transient response for the kind of waveforms encountered in music? Not necessarily.

In a sense, the square wave is still a steady or continuous waveform, although its shape may make us think otherwise. It repeats continuously at its set frequency. A square wave reproduced by a speaker may sound somewhat like some musical tones, but it is a continuous tone. And actually you cannot *hear* a square wave, as such.

To do so, the speaker cone would have to hold quite still for half a cycle, then move infinitely quickly to the other position where it must hold quite still for the other half-cycle. No speaker cone can move so quickly. Even if it could, the air it drives could not and neither could your eardrum at the receiving end. So a square wave is not a realistic representation of a musical effect. As someone truly said, "We don't listen to square waves, we listen to music."



An amplifier and the other parts of a high-fidelity system may settle down to "handle" a square wave quite "happily" but still get thrown off balance by the sudden or unexpected type transients encountered in music. This can happen in lots of ways, both in amplifiers and other parts of the system.

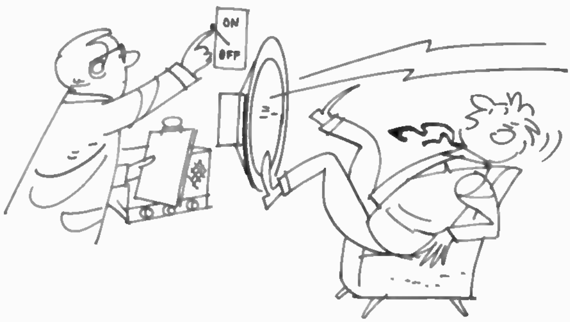
A question frequently asked at this point is: "But surely, other things being equal, a system with good square-wave response will sound better than one without?" Unfortunately, this question contains a fallacy when it says "other things being equal." Even though both systems are as nearly alike as possible, the fact that one has a good square-wave response, while the other has not, means there is a difference. And it is impossible for this difference not to affect something else besides how it responds to square waves.

Just how this difference, whatever it is, will affect performance on musical transients is something that cannot be answered with a simple yes or no. In one amplifier, the response to musical transients may have been spoiled by what was done to get good square-wave response, while in another the reverse may be true. The fact is, there is no direct connection between the two forms.

Tone-burst testing

Another attempt to simulate the effects produced in handling musical transients uses what is called the "tone-burst" test. This uses a regular single-frequency tone but switches it on and off to simulate starting and stopping of a musical note. The test then investigates whether the tone is distorted in any way, either just after it is switched on or as it is switched off.

This test is certainly more effective in "finding" some of the distortion, at least, that happens on musical transients. But it has limitations. To be able to view the effect on a screen or to take



continuous readings of it, the tone must be switched at a fairly rapid rate. Also, the oscillators available for controlling the "switch" do not operate at much slower than 40 times a second (20 on's and 20 off's).

A tone repeated 20 times a second is a little more "staccato" than anything encountered in music! This test is helpful, but it does not give sufficient time for some of the distortion that occurs in handling musical transients to happen.

Wanted — better tests

Other tests are at present being devised with a view to finding a way to get a more accurate indication of the things that really happen during the playing of music through a high-fidelity system.

A lot of tests are available. Each of them finds some forms of distortion. It is very evident, the more we look into the matter, that no one test will ever produce a single answer as to how much distortion an amplifier or system produces. The question is too complicated. But we do want to know the importance of the various kinds being measured.

To find this out, a variety of tests have been made to see how objectionable different kinds of distortion are.

Harmonic distortion, the kind produced when only one tone is playing, and found by a single-tone test, is almost undetectable by itself. The 5% figure once considered good for this purpose still is. But since we do not play music with only single tones all the time, and as equipment that produces harmonic distortion will usually also cause other kinds, this is small satisfaction. The harmonic-distortion figure tells us little about what the same equipment will give with other kinds.

Intermodulation distortion, measured by each method, is directly related to a kind of distortion that is very audible in repro-

duced music. But the degree of audibility in music is not necessarily indicated at all reliably by the measured value. This is because neither of the measurements gives complete information about the *order* of the distortion components — how *many* unwanted frequencies they are liable to produce — only how *much* of them.

Although transient distortion has been discussed for a long while, proving that high-fidelity people were at least aware of its existence, it is only recently being satisfactorily analyzed to find the different varieties that can happen. Some tests have been made to see how audible some forms are, but the work is far from complete. This is an area where there is room for a lot more work toward the attainment of high fidelity.

Looking for trouble

The proportions of the different kinds of distortions to be expected will vary with the component causing it. To some extent this is a means of locating the origin of distortion. You will find that careful listening can often help you track down the cause of unpleasant sounds that should not be there.

Identifying distortion

A sound like the grille cloth of the speaker buzzing may well be just that. Check very carefully to see it isn't, before looking elsewhere. Some forms of intermodulation distortion caused by amplifiers can produce a similar sound. A poorly aligned speaker is another possible cause.

If the voice coil of the speaker is knocking against the pole pieces or some form of end stop, that may be what it sounds like too. It can also be due to amplifier overloading producing clipping. This form of overload is far more likely to be in the output section (basic amplifier) than the front end (preamplifier).

When a lightweight pickup collects dust under the stylus, the sound gets "fluffy." This kind of sound you will recognize after you have once heard it. But it can also be due to some form of amplifier transient distortion.

In trying to identify distortion by listening, don't forget that it's possible for distortion to be in the record or transmission. After all, they use amplifiers, too. For the most part they use every possible care to avoid all possible forms of distortion. So, in most instances, distortion is far more likely to be in home playback equipment than in the recording or transmission. But if you have a really good distortion-free system, you will soon be surprised how much program material is *not* as free from distortion as you might expect. The more you listen, the more critical you will get. You'll find yourself hearing distortion that others cannot detect, but it's there.

This has been the story of high fidelity. As systems generally improved, the quality of records became more noticeable. Record manufacturers had to improve their recording and pressing techniques. Then pickup makers had to improve pickup design so all the fine quality recorded could be appreciated. Every component along the line could now benefit by being improved. Presently the situation would be repeated (at a much higher quality level). Systems were again good enough to be critical of the quality of the record.

But what do you listen for in trying to get quality? As there are many forms of distortion that can occur, you need to listen to a variety of program types, for a variety of "signs" to determine the overall quality. And if you hear one of the signs, check with a variety of similar program material to be sure whether the fault lies in the recording or in your own system. If it's always there, it's most likely at your end; if it's evident only in some cases of a certain type of recording, then obviously those particular recordings are at fault. Similarly, if you notice it on transmissions from certain radio stations and not others, you know where the fault is.

Low-frequency distortion

Low-frequency distortion can take two forms, irregularity of response or distortion of the wave. Irregular response makes some bass notes unnaturally louder than others. This is more likely to be in the speaker than anywhere else. The speaker or any other component can also cause either deficiency or excess of low frequencies in general. This was discussed in the previous chapter. One more point: if the bass is naturally deficient, due to poor design, and some attempt is made to correct this by boosting it, most probably the low frequencies will cause distortion.

Harmonic distortion of low frequencies is seldom if ever noted. What does cause trouble is the intermodulation distortion. The presence of low frequencies, such as the deep pedal tones of an organ or a heavy string bass, causes a fluttering effect in the reproduction of other tones. If a pickup is not quite heavy enough to stay in the groove where these heavy low frequencies are, it will bounce off the groove walls, causing this kind of distortion. Amplifiers and speakers can cause quite a similar effect.

Distortion at higher frequencies

Distortion at higher frequencies can also take the form of either irregular response or actual waveform distortion that may cause spurious sounds altogether. The pickup or speaker is most likely to cause irregular response. At the higher frequencies, this is noticed by an irregular quality about treble tones. Their coloration is not consistent. Some will sound "sharp" and others "flat." This effect will occur on the same tones, regardless of what instrument is playing them.

Buzzing

Waveform distortion in the higher frequency range produces buzzes and tinkles that seem unrelated to the music. Actually, their only relation is that they occur at a particular musical tone combination or chord.

Shrillness

Excessive response to higher frequencies is marked by a general shrillness. Sounds that are normally brilliant, such as a xylophone, take on a "cracked" quality. Deficiency in this range is less noticeable, provided some attempt has not been made to correct it. Then, very often, the effect is a "fluffy" reproduction, as the pickup is playing with a considerable amount of dust collected under its stylus.

Flutter

Another form of transient distortion occurs particularly in amplifiers — usually the ones with "plenty of feedback." This produces a fluttery effect, like the low-frequency intermodulation — except that it occurs when there aren't any low frequencies present to cause it! Music from wind instruments or strings played by plucking (and sometimes by bowing) is most likely to cause this.

We could go on describing a lot more forms that have been noticed. Those are some of the main ones. You'll find the lesser ones as you go. Any distortion that robs an instrument of its realism is due to some transient effect.

If the piano hammer strikes sound tinny, the bass drum sounds like dead wood instead of pigskin, or you hear a wavering vibrato-like effect (and have checked it isn't wow or flutter), it's almost certainly a form of transient distortion that the tests didn't show up.

In listening for these things with new equipment — and when you are buying it — pick program material with which you are familiar. Then you will know that the program material itself is not responsible. Familiarity also means you will know how it *should* sound.

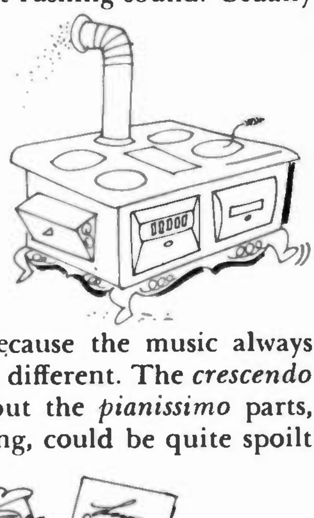
One more thing about buying: For example, when shopping in a hi-fi showroom you may spot some form of distortion — one of the types we have discussed. The salesman may try to convince you that you imagined it or you may be told that "the amplifier's new; that is a little roughness will clear up after a few hours' use."

An amplifier is not like an automobile in this respect. A car may run smoother after the first thousand miles or so, but an amplifier never does. Little distortions are not likely to disappear; they're far more likely to build up into bigger ones.

chapter
3

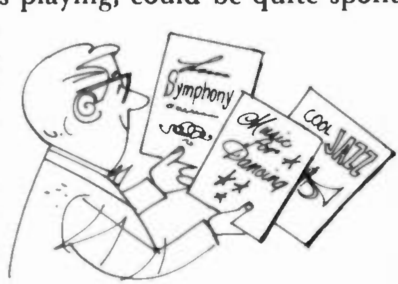
dynamic range

IN the early days of high fidelity – when the name was just beginning to be used – it seemed as if it differed from low fidelity by a characteristically ever-present rushing sound. Usually this was tube hiss from one of the earlier stages in the amplifier. Sometimes it was resistor noise. But the difference usually occurred because a high-fidelity pickup of those days did not produce as much electrical output as the low-fidelity variety.



What's your choice?

With dance music or jazz, this hiss or rushing sound was not objectionable because the music always drowned it out. But symphonic music was different. The *crescendo* passages might drown it out all right, but the *pianissimo* parts, where maybe just a single violin is playing, could be quite spoilt by it. The beauty of the instrument was lost in the rushing sound. So it was the disciples of high fidelity became aware of dynamic range.



The work at Bell Labs we mentioned in the first chapter uncovered an interesting fact in this connection. The range of human hearing, between sound that just becomes audible and sound that is so loud as to be almost unbearable, is enormous. It represents a change in sound intensity, over most of the audible frequency range, of over a 1,000,000,000,000,000 to 1. So when you hear someone say one sound is a million times as loud as another, he may not be exaggerating!

Exaggeration in reverse!

The fact is that, what seems to the ear like a change in level of, say 10 times the volume, may really represent a change in actual sound intensity of a thousand times or more. This does not follow any direct or simple relationship. Making a sound seem twice as loud does not require any consistent increase in sound intensity.

At the threshold where sounds become audible, a relatively small change in sound intensity makes the difference between hearing it and not hearing it. A further quite small change can also make it quite perceptibly "more audible" – it could hardly be called "louder" at this level.

Now suppose the intensity is 10 times that where the sound just crosses the threshold of audibility: it is still a very quiet sound. At this point, increasing the intensity by another 10 times will make the sound seem about twice as loud.

Increasing it another (third) 10 times, making it now 1000 times the threshold intensity, will make the sound seem about 3 times as loud as the first sound, which was only 10 times threshold intensity.

By the time you get to intensities in the range from 100,000 times threshold to 1,000,000 times, the 10 to 1 change in actual intensity will only *seem* like about 20% increase in loudness.

Now we begin to see why dynamic range is important. The dance or jazz music may represent a sound-intensity variation of maybe 100 or 1000 times. Played at a level of about 1,000,000 times threshold, it easily drowns out tube hiss, even if this should be from 10 to 100 times threshold. But the symphony is another proposition. This itself may represent a sound intensity range of 1,000,000 times or more.

Getting it all through

Suppose the loudest part of the music corresponds to a power of 10 watts delivered to the voice coil of your speaker, and this gives a good loud sound in the room. To handle this big dynamic range, the soft passage must produce sound from 10/1,000,000 watt (called 10 microwatts). The speaker has to be quite sensitive for the coil to move the cone at all with such very small power.

More than this. The power the speaker gets is amplified very many times from that produced by a high-fidelity phonograph

pickup, possibly a million times or more. This means the power at the input to the amplifier, corresponding to such a soft sound, must be very small indeed. It is almost inevitable that we will get trouble from resistance noise and tube hiss.

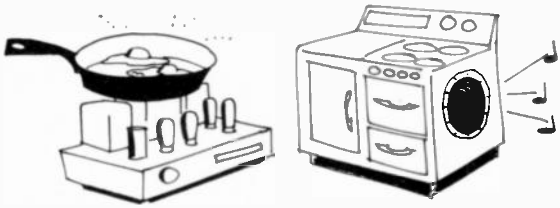
Watch the headroom

The expression *dynamic range* refers to the range of sound levels the system, or some part of it, can handle. It has two limits, a top and a bottom. The top limit is set by the maximum power capability. At the output end, this may be governed either by the amplifier or the speaker, depending on which runs into trouble first.

Power ratings

If a speaker is rated at, say 10 watts, this means it should be capable of converting 10 watts of electrical power from an amplifier into sound. It does not have to get 10 watts of power before it will work, nor does it mean it will automatically take 10 watts when it is connected to any amplifier. It means it will convert *up to* 10 watts of electrical power and not any more into sound power. If more than 10 watts are delivered to it, it is liable to buzz, rattle, fall apart or even burn out.

An amplifier rated at 10 watts is capable of delivering 10 watts



to a speaker, provided the impedance ratings match — the speaker voice coil with the amplifier output terminals. This is a figure quoted in *ohms*. So if the speaker is rated "Impedance: 16 ohms," it needs to be connected to the 16-ohm terminals of the amplifier.

Again, the amplifier can deliver less than its rated power, but not more. This depends on how much you feed into it. If you try to drive it harder — to give more than its rated power — it will produce severe distortion.

That's all nice and easy, if you connect a 10-watt speaker to a 10-watt amplifier. But can you connect different combinations, either way? You certainly can, and here's what happens.

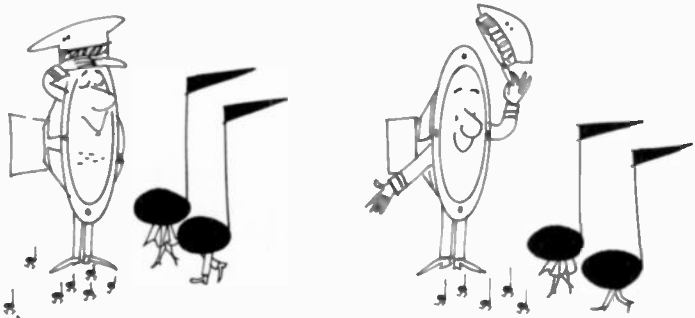
Which one stops it?

If you connect, say, a 10-watt amplifier to a 30-watt speaker, you will just never drive that speaker all the way. A 10-watt amplifier will not give any more power to a 30-watt speaker than it will to a 10-watt one. It will go its 10 watts and no further.

Taking it the other way around, if you connect a 30-watt amplifier to a 10-watt speaker, it is the speaker that sets the limit. It will still handle up to 10 watts from the amplifier although the latter is capable of 30 watts. But if you try to push the full 30 watts into the speaker, it will distort badly and probably damage itself seriously.

Speaker efficiency

While electrical power is an indication of *relative* loudness, using a particular speaker, loudness also depends on the speaker's *efficiency*. A high-efficiency speaker may convert between 10% and 25% of the electrical power into sound. A low-efficiency one



would convert the same electrical power into only one-tenth the sound power, ranging between 1% and 2.5%.

So a high-efficiency speaker fed with about 3 watts would be as loud as a low-efficiency one driven with 30 watts, regardless of power rating. There is no connection whatever between efficiency and power rating.

Efficiency determines how much of the electrical power gets converted into sound waves.

Power rating tells us how much *electrical* power the thing will take.

Possible sound volume depends on *both* of these properties.

Watch the lower level, too

So much for the high-volume end, which fixes the top limit of dynamic range. The bottom limit is fixed by various things that produce noise, like tube hiss.

You will never hear the tube hiss of the output tubes. But each stage of amplification amplifies, along with everything else, the noise that comes from the stage before. So you are most likely to get audible tube hiss from the input tube. And the more amplification you need, the more likely is tube hiss to become audible.

Other causes of noise that limit the bottom end of dynamic

range are atmospheric noises in radio reception, surface noise in phonograph reproduction and also tape.

In the sky

Atmospheric noise on radio includes man-made interference. Various methods have been adopted to improve dynamic range in radio transmission. Anything that will get the incoming radio wave farther "above" the noise will improve dynamic range.

Being closer to the transmitter is one way to achieve this. A good receiver operated with a good antenna near the transmitter will never have a noise problem. It may run into distortion,

because the front end cannot handle such a big incoming wave, but that's another matter.

Where the receiver is farther away, two things can help get the incoming wave above the noise: a more powerful transmission or a better antenna. Making the transmission more powerful has its limits, though. The program you want to hear is part of someone else's noise problem.

There are so many stations on the air that it is difficult to separate them. As we will see more fully in the chapter on radio, there are two ways to go about getting satisfactory separation. Raising the transmission power can do only so much toward reducing the effective noise at the receiver.

Making the transmission directional can be the equivalent of putting out more power in the favored direction. This can sometimes help extend the service area of a radio station without correspondingly imposing on others.

Making the receiving antenna directional is another way of improving the received wave and at the same time cutting down interference received from other directions. Use of FM in the vhf band has also helped the problem considerably.

The radio receiver itself can have its noise problems, the same as the first stage in a phonograph amplifier. Here tube manufacturing techniques and improved circuit design have helped over the years to make better radio reception possible.

On the platter

The earlier phonograph discs in use when high-fidelity first got started were the electrical recordings that first superseded the old acoustic recordings. They were made of shellac composition that included an abrasive in the mixture. Why the abrasive was put in these 78-rpm shellacs remains somewhat of a mystery.

In those days discs were played with steel needles of various grades of hardness. Possibly the disc manufacturers were afraid that the sharp steel needle would wear its way through the disc unduly quickly and incorporated the abrasive to wear down the needle first. Some of them claimed it was to make sure the needle kept the same shape as the groove *while* it was wearing down, and thus avoided wearing the groove out of shape.

In any event, the abrasive was somewhat coarse and resulted in excessive surface noise. Because of the molecular size of the abrasive particles, the noise had its own hiss coloration. This has been largely eliminated by the advent of LP's, which are pressed in a vinyl plastic without any abrasive and are unbelievably quiet after the others.

Phonograph pickups have always had a two-way problem. On the one side, they need to get enough to be well above circuit noise — tube hiss or hum picked up in the wiring. On the other side, they want to achieve a good frequency response with low distortion. For many years it seemed as if these needs conflicted. But various design improvements that will be discussed in Chapter 7, have certainly brought phonograph reproduction a long way since those days.

On the oxide

Tape is a relatively new member of the high-fidelity family. It too has had its noise problems. But improvements in both tape and head design have vastly changed this situation too, as Chapter 6 will explain in more detail.

Whatever "source" you use for program material — radio, disc or tape — the medium itself has limitations as to dynamic range.

Research has extended this over the years. But the medium itself has always been the basic limiting factor for the whole system. There is only so much range between the biggest and smallest signal it can take, due to the various limitations pertinent to whichever medium we consider, radio, disc or tape.

It can't go as is

This led to the question as to whether something could not be done to "squeeze" the range somewhat. Obviously, it is better to hear that *pianissimo* violin above the noise, so the hiss does not spoil its delicate quality, even though it may be a little

louder than strict proportional representation would dictate. That obvious axiom of Chapter 1 is coming in for another serious challenge of its validity.

As a point of fact, practically all program monitors do compress the program a little to achieve just this purpose. During quiet passages they turn the volume up a little, and during loud *crescendos* they back off a little, so the range between the *pianissimo* and *fortissimo* is reduced somewhat compared with life. This is done whether the program is transmitted over radio or recorded.

To avoid noticeable distortion, which would occur if the program monitor did not turn down the volume quickly enough on a sudden fortissimo, he works from a score, and takes care to have the volume turned down a little ahead of time, so the loud passage has enough headroom when it arrives.

Here we go again

This brings us into another area where the audio experts have argued for many years. One group is quite happy to accept this limitation, tastefully introduced by the program monitor to improve the overall effect. The other groups says, "No." This is a departure from their axiomatic ideal. They want the full dynamic range of the original restored — nothing less. They also object to the monitor

intruding what he considers to be an optimum "riding of the control."

So, many years ago, automatic compression and expansion circuits were evolved to achieve this electronically from the program itself. When a crescendo came along, the amplification would suddenly be reduced, to avoid distorting it. During following quiet passages, the amplification is gradually brought up to avoid their being "lost in the noise."

Played back just as it is, the compressed program will not be very different from the kind produced by a human program monitor. With a corresponding expander unit, the process can be reversed, bringing back the full dynamic range of the original program.

This was a wonderful idea. And, as far as improving dynamic range is concerned, it worked like a charm. The reduction in audible surface noise from the old 78 shellac discs, played this way, just had to be heard to be believed. So why didn't it get more widely used? While it certainly did improve dynamic range, it had various problems in other directions.

Compressed fidelity?

The most obvious one is the question of how much compression and expansion to use, and making sure both ends use the same amount, or near enough to be realistic. Suppose the available dynamic range on the disc or radio, allowing a suitable margin, is reckoned to be an intensity ratio of 10,000 to 1.

Dance or jazz music has a range of only between 100 and 1,000 to 1. So this could be expanded instead of compressed, if we wanted! In any event there is no need for compression.

But the orchestral symphony may use 1,000,000-to-1 range. So the compression will have to take this and "squeeze" it into a range of 10,000 to 1. Then the expander at the receiving or playback end "stretches" this back to 1,000,000 to 1.

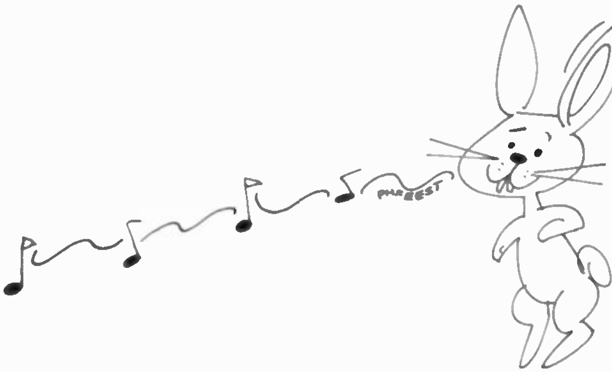
To be consistent, all program *could* be compressed in proportion. In this way, dance or jazz music with a range of 1,000 to 1 would be compressed to about 100 to 1, and some with only 100 to 1 would get compressed into narrower limits — about 20 to 1!

This would be all right. But what of jazz or dance music that has not been compressed? Play this back through the expander and it has a range of between 10,000 and 1,000,000 to 1. This produces a most weird and unnatural sound.

Those sibilants again

Another problem with this is what happens when speech is expanded. True, the background noise is reduced — fantastically. But a very peculiar effect is observed on words beginning with the s sound.

Although this sound normally may appear as loud as the word



that follows, the actual sound intensity is a lot less. This happens without our being aware of it, because our hearing is especially susceptible to the frequencies in the "s" sound. When a word beginning with this sound passes through an expander, the s does not work the thing, but the word that follows does. This way, it seems as if the speaker suddenly realized he was whispering and raised his voice after the first letter.

Passing the same word through the compressor produces the reverse effect. The sounds at the beginning of words get grossly overemphasized because they get through before the compressor turns the amplification down.

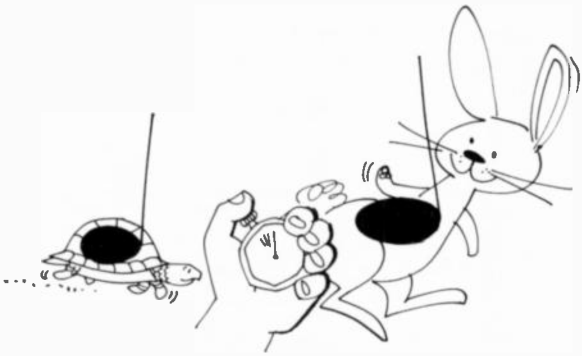
One way to get over this problem is to make the control circuit that works the change in amplification extra sensitive to the range of frequencies (3,000 to 5,000 cycles) in the s sounds. In a compressor, such an arrangement is called a "de-esser" because it prevents that overemphasis of the s sounds.

But this only complicates the problem more, because some program wants "de-essing" treatment while some doesn't. And, to be effective, the controls on both compressor and expander must always do the same thing, in reverse, of course.

A question of timing

Finally, even if these little difficulties are overcome, maybe with some kind of compromise so nothing suffers too much by not getting its ideal treatment, the system has more problems.

It will work perfectly only if the amplification changes *instantly* when a sudden *fortissimo* passage comes along. While it can be



made very quick in action, it is virtually impossible to make it instantaneous. So either some delay occurs in the impact of any *crescendo* or the crescendo gets distorted because of failure at the compressor end of the setup to get the amplification turned down quickly enough.

Actually, the delayed action is more difficult to avoid in the expander than in the compressor, because it is basically easier to compress than to expand dynamic range. But if the original program quality is to be preserved, the two rates have to be identical. The compressor should turn down the amplification at exactly the same rate that the expander can turn it up again. Then the *overall* amplification will be constant at all times. But this leaves the compressor working too slowly to handle the crescendo peaks without distorting them.

So, at present, the general consensus accepts the human monitor's compression as being a more tasteful solution, not requiring any restoring expansion at the playback end.

Often a modified form of compressor, called a limiter, is used as an electronic safeguard. This does nothing unless an extra loud passage overshoots the top. Then it very quickly turns down the amplification enough to accommodate the overshoot, slowly bringing it back again afterward.

A few object to both these features as taking unwarranted license with the program. They would prefer to have the full original dynamic range preserved, even if this make the hash much more audible or occasionally runs into a little more distortion. Such is their concept of high fidelity. Since high fidelity is, as we are discovering, just a relative term, they are certainly entitled to their opinion.

Finer details, too

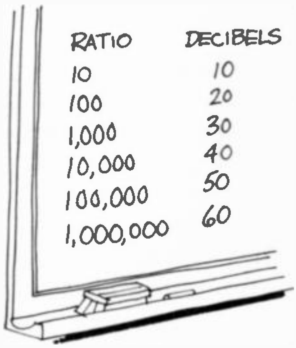
But the extremes are not all there is to dynamic range. Just as frequency response is concerned with more than frequency *range*,

so differences in intensity need attention in the fine as well as the coarser details. In fact, both are really interrelated.

When, in Chapter 1, we spoke of a "flat" or uniform frequency response, we did not go into detail about *how* flat. It is impossible to have response perfectly uniform at all frequencies. So how much deviation from perfection are we to permit?

Going back to the experimental work, measurements were made of the effect different changes in sound intensity had on the illusion of loudness. Some of this we have already related. Here are some more findings.

Perceptible difference depends to some extent on the starting-point loudness. But, in general, an increase in intensity of less than about 25% is almost impossible to detect. With very careful listening, about 60% increase can just be detected. A 100% increase is just comfortably heard, while a 300% increase (quadrupling the intensity) makes a readily noticeable difference almost anywhere on the loudness scale.



Why decibels?

In much the same way that each time frequency is doubled the pitch raises an octave, so each time intensity is doubled the loudness *adds* about the same effect. For this reason a special scale has been in use for some years now. Instead of referring to intensity changes of 1,000,000 to 1, as we did a little earlier, the high-fidelity jargon would refer to that as 60 *decibels,* abbreviated to *db* and pronounced "deebees."

A practical scale

This is a uniform *ratio* scale, which is what is meant by calling it "logarithmic." Every time power or intensity is *multiplied* by 10, the db scale *adds* 10. If the intensity is *multiplied* by the square root of 10, or 3.162, the db scale *adds* 5. Other ratios converted to the db scale are:

1.26:1	1 db	4:1	6 db
1.58:1	2 db	5:1	7 db
2:1	3 db	6.3:1	8 db
2.5:1	4 db	8:1	9 db
3.16:1	5 db	10:1	10 db

This is a more practical basis for measurement and comparison, because the db scale figures give a better idea of how the difference will *sound.* Remember that adding an extra 2,000 cycles to the top end of the response, when this already goes to 14,000 cycles, does not mean as much as the numbers imply. It is equally true that adding 10 watts to an amplifier that gave 50 watts does not make it sound appreciably louder.

The power question

This is better seen by making the conversion to db. From 50 to 60 is a *ratio* of 60/50 or 1.2, which is less than 1 db. If 1 db is barely detectable with careful listening, then an increase from 50 to 60 watts should not be detectable either.

This may make it look as if buying power amplifiers "by the watt" is not very profitable. This is true. If you exchange a low-efficiency speaker for a high-efficiency one, you will do as much to increase *apparent* sound power as exchanging a 10-watt amplifier for a 100-watt one!

Of course, there may be quite valid reasons why you want to keep your low-efficiency loudspeaker and use more watts to get the loudness you want. It might even be cheaper as well as more convenient. But that you'll have to figure out when you know a little more about speakers. You can get some assistance in this direction by reading Chapter 9.

This question of the dynamic range you *need* is affected by more than just the program. It is also affected by the size and decoration of your listening room, and what neighborhood you live in.

Extraneous noise

The noise that limits the bottom end may not come from your speaker, but from a nearby highway. You still want to have your dynamic range *above it*. Because of this noise, your threshold may be raised 20 or 30 db, which means your power range should be somewhat higher than needed in a quieter neighborhood, to say the least.

Of course, there are limits to this. Few high-fidelity systems will cope with the noise of a train on elevated tracks outside your window, which some people have to contend with. They just get used to leaving off listening while the train goes by.

Those specs again

Finally, what about this question of how flat? This is where the incentive to provide better advertising claims has influenced design again. As we stated earlier, less than 1-db change in level you cannot hear, even at the same frequency. So you certainly

could not be critical of a level difference of 1 db between one frequency and another.

How flat the response?

There is absolutely no purpose to be served by achieving a response flat to a fraction of a decibel over a wide range of frequencies.

In point of fact, there are differences more important than this that the claims don't mention and that we discuss elsewhere. For one, the fact that the measurements are made with the amplifier connected to a resistance load, while we *listen* to a speaker. For another, the fact that the amplifier may not hold as steady on transients as it does on the test-bench measurement.

On the trail of the grail

However good a system is made, at both the recording or transmission and playback or receiving ends, it seems as if continued critical listening will eventually disclose some departure from perfection. This is why the pursuit of high fidelity is such

a perfectionist quest. But the search for perfection should be pursued intelligently. As has been proved in a variety of ways, the perfect reproduction of the original performance is an impossibility. What we seek (and what we will always have tantalizingly out of reach) is the perfect illusion.

This means we need to ask ourselves whether imperfections scientifically measurable are, or can be, responsible for audible departure from perfection. On the basis of the quest for perfection, it would seem that holding the response at different frequencies within 0.5 db should be better than within 1 db; and

0.1 db should be much better still. That is why amplifiers with such specifications are built. It's much like the reason for extending frequency range far into the upper regions only dogs and bats can hear.

Home on the range

Dynamic range can be important, but this is seldom specified as such. If you want to figure it out, you can find the range between the hum and noise figure and maximum output power, if enough information is given. Often it isn't. But even that does not always give a true indication of *effective* dynamic range. The range of human hearing is very much less at the lower frequencies than it is in mid-range and high frequencies.

How effective is it?

Effective dynamic range will be different according to frequency. Not only will it depend on the range the system can handle at that frequency, but also on how much *of the total program sound power* appears at the frequency in question. This will vary from program to program. This means overall effective dynamic range is quite complicated to figure out — in fact almost impossible.

What can we do?

The best we can do is make sure, as we said earlier in this chapter, that we have enough headroom and a low enough ground level. What is enough is, to some extent, up to you, your choice of programs, your listening surroundings and your own musically trained ears.

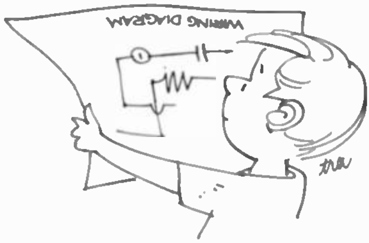
Now we've laid the foundation for getting down to the details that get discussed among fans and which we will go into, one by one, in the remaining chapters.

In short, we are finished with our bird's-eye view and are now coming down for a closer look at the different sections that go to make up hi-fi — circuits, radio, records and tape, pickups, microphones, speakers of all sorts and stereo.

chapter 4

circuits

Most high-fidelity enthusiasts never wanted to get interested in circuits, and never will get really interested in them. All they want is the necessary "electronics" – in a package, not on paper – to give them the quality of reproduction they are looking for. But this very desire is apt to find you unexpectedly "in the middle of something."



Maybe you bought some equipment that sounded good. Then along came an "electronic whiz." You asked his opinion of what you have. Or maybe he volunteered it anyway. He proceeded to tell you that the circuit in your equipment is just no good, what you need is . . .

So now you're confused about circuits. If this hasn't happened to you yet, it's likely to at any time. So, without getting involved in technicalities beyond names and a simple rundown on what some circuits do, let's have a go at sorting out the confusion.

Some background

The earliest requirement of an amplifier was to make the sound *audible* through the speaker: getting enough *gain*. Next, to get dynamic range, it needed to give the louder passages more *loudness*, so the softer ones are audible. This needed *output power*.

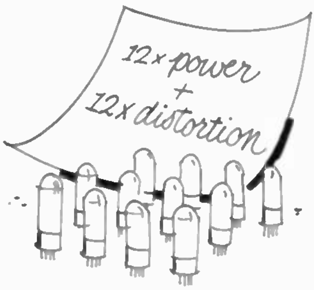
The first kind of tube used was the audion, now called a triode. With transformers between audions, a number of each would give enough gain to achieve audibility, while, if the speaker was fairly sensitive, the tubes with the heaviest filaments would give some

GETTING IT AUDIBLE

range, enough to make some louder sounds louder than some quieter ones.

Getting power

To get more range, the next step was to connect a number of tubes together for the output stage, with the idea of delivering that many times the output power. But it was soon found that simple multiplication of the number of tubes was not successful in getting the extra power without also acquiring some quite audible distortion. With, say, four output tubes together, you got four times the output power, but you also got four times the distortion.

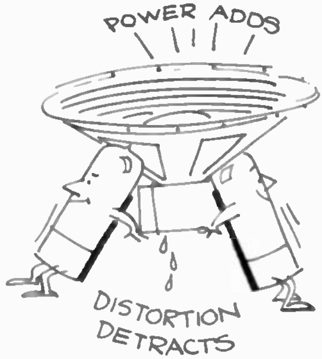


Push-pull

This invention put two tubes in a sort of back-to-back circuitry. so while their power added, their distortion (or at least quite a hunk of it) mutually subtracted. The improvement was considerable. More than twice the power, in fact, was obtained, with much less distortion, and the bass frequencies in particular were much better handled. As speakers that reproduced bass much better began to appear about the same time, this worked out quite well.

With the improvement produced by push-pull, it was discovered that the succession of transformers between tubes almost invariably deteriorated frequency response and often produced considerable distortion as well. So means were sought to eliminate audio transformers.

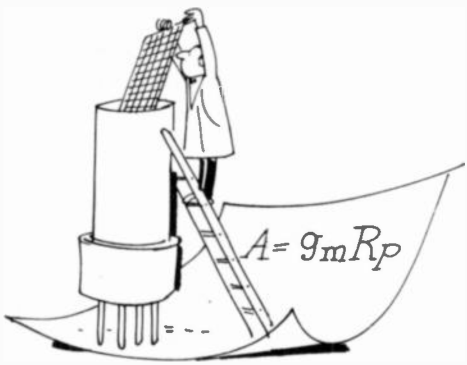
When elimination of transformers, by using R-C (or Resistance-Capacitance) coupling from one tube to the next, was first tried, some extra tubes were needed to replace the lost gain. But improved tube design soon made it possible to use R-C coupling without any extra stages.



Tube development

Had that been the only line of development, the sequence of progress would have been clear. But at about the time push-pull was being invented, tube inventors were busy devising better kinds than the triode. In the triode, which means "three-electrode" tube, the *cathode* provides electrons, the *grid* controls their flow by smaller electrical impulses fed to it, while the *plate* receives the fluctuating electron flow, in the form of an amplified version of those impulses fed to the grid.

But a lot of amplification is needed. In those days, many tubes, each one amplifying the output from the previous ones, were needed to get an audible output. Couldn't more grids, or something, be put into one tube, so the necessary amplification could be obtained with less of them? One extra grid made the *tetrode,*



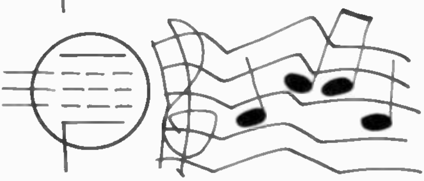
or "four-electrode" tube, and a little later the *pentode,* or "five-electrode" tube, was born.

With different mechanical variation in its construction, the pentode can do a lot more things than the simpler triode. In one form it can achieve as much gain or amplification as (or more than) two triodes, one after the other (as much as three or four of the earlier types). In another form, a pentode of given size and electrical power consumption can give much more sound power than its equivalent triode tube.

Triodes vs. pentodes

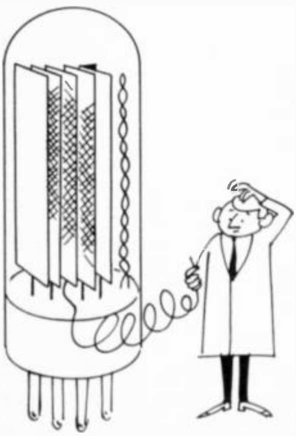
So, around the late 20's and early 30's there were already two "schools." One group, working with circuits outside the tube, had improved quality and power by inventing push-pull, while another, working inside the tube, had produced pentodes capable of more power out for less in, with more amplification thrown in for good measure.

If you really wanted the best high-fidelity quality, the push-pull boys had it but, for getting plenty of "oomph" at minimum cost, the pentode tube was it.



But this was only the beginning. It was inevitable that someone would combine the two — pentodes in push-pull. But, in itself, this did not help too much because push-pull does not do as much to improve pentode performance as it does for triodes. So the general picture remained the same: triodes for best quality, pentodes for best efficiency (most sound output for a price).

As time went by, all the bigger tubes began to be pentodes. But this did not stop the triode enthusiasts, for whom quality assumed paramount importance. Because a tube has three grids and a plate in it does not mean you have to use it as a pentode. By connecting them together outside, so you only have three separate connections, corresponding to the triode's cathode, grid and plate, the tube works just the same as a triode. So the same tubes were being used, sometimes as pentodes and sometimes as triodes.



Ultra-linear and others

This fact probably inspired one of the later developments in circuitry: Ultra-Linear. The middle grid could be connected either to the plate, to make the tube function as a triode, or to a high-voltage supply point, when it performed as a pentode. As a triode, the middle grid had the same audio "signal" as the plate; as a pentode, it had only a supply voltage. Ultra-Linear "splits" the difference: the middle grid is connected to a circuit point artificially created, somewhere between the two extremes.

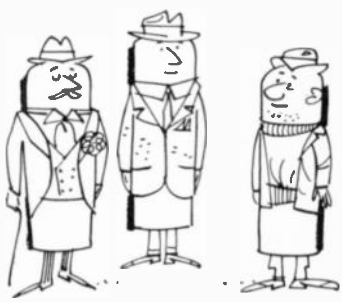
This yields a mode of operation that almost reaches the quality of triode operation, with the power output of the pentode. In fact, it came as near as we shall ever get to eating our cake and having it.

Other circuit tricks with the output tubes use different methods to get similar results. Among these are unity-coupled, twin-coupled and circlotron varieties, all of which use pentode tubes with some circuit trickery to improve their effective performance.

But all of these, including Ultra-Linear, are of comparatively recent origin. Before their advent, other variations had appeared, which added further possibilities to the permutations and combinations of circuit design. Principal of these was the group designated by *class* letters and numbers.

The advent of class distinction

This closely followed the invention of push-pull. Connecting two tubes to work back-to-back, so each tube shared in giving output all the time, was "class A." By using the same pair of tubes so each handled only one half of the sound wave — the two handling opposite halves — from two to five times as much sound power can be obtained. This is "class B."



The big stepup in efficiency made available by class B was not by any means for free. Expensive transformers were needed to make the circuit work properly, as well as an expensive supply unit to feed the tubes with current. But the overall stepup, sound power for dollars, represented an economy. A well-engineered circuit could just about match class A for quality, too — in fact, many surpassed the average (a fact little known). But, in general, the class-B circuit was more difficult to handle and did not achieve the quality of class A.

So, like the compromise between triode and pentode that was to come later (Ultra-Linear), a compromise was devised between class A and class B, called class AB. For small sound powers this worked as class A, changing to something more like class B when extra power demanded, but never going to the full extent of class B. This enabled less expensive components and less critical circuits to be used, while still achieving about twice the output formerly obtained from the same tubes working class A.

That about covers the different variations, at least as regards the output tube section – quite a variety of tangents from the original starting point. It is the permutations and combinations of these that provide most of the amplifier circuits in use today. Whichever of the better quality variations you pick up, it will have one feature in common with the others: the use of push-pull. And this entails another feature they all must have: the phase inverter.

Phase inverters

To work tubes back to back, which describes the way they handle the audio "signal" rather than any physical relationship, the feed to one of them needs turning around, the technical name for which is *phase inversion.*

By now you are probably asking yourself, "How complicated can audio circuits get?" But you are quite likely to run into the circuit whiz who will tell you that your amplifier "uses the wrong phase inverter," and it's good to know how to meet this comment.

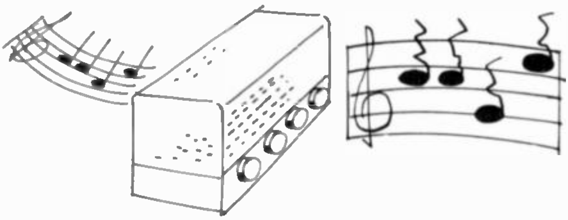
He says this because he happens to believe that a different circuit is the best phase inverter. Without attempting to go into details about phase inverters – split load, paraphase, floating paraphase, long-tailed, cross-coupled, or what have you – it can definitely be stated that no one phase-inverter circuit is best under all circumstances. It takes a man with many years' experience in circuit *design* – a much more comprehensive background than your self-styled circuit whiz – to tell which is the best for a particular amplifier design. So we shall not attempt it here.

Choice of the power end

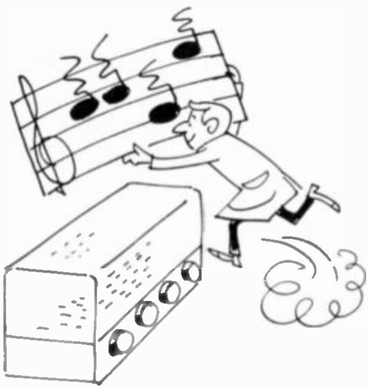
We're almost through with the variations that can occur in what the British call the power amplifier, while the American expression is more commonly basic amplifier: the output end responsible for developing sound power. One more ingredient to the circuitry:

Feedback

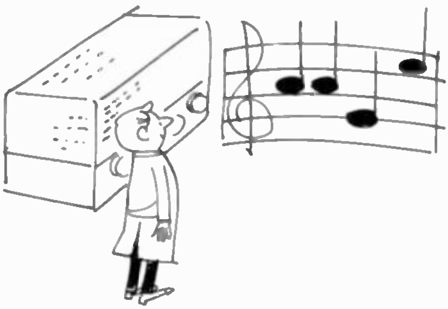
You may use (or your amplifier may) any of the circuits



so far discussed, without being considered particularly out of date, but if it doesn't have feedback, the conclusion is its designer



must have antedated Noah! This attitude of mind has grown from the perfectionist obsession for reducing distortion and extending frequency response.



Long ago, when careful design had reached its limits, with a frequency response from about 30 to 16,000 cycles (picking figures out of the air); and a distortion of about 1% or 2%, feedback was invented to take matters still further. A sample of the output is fed back through an "attenuator" which reduces its magnitude to correspond with the original input. It is then combined in such a way that unwanted distortion components trying to show up in the output are fed in at the input *the other way round,* so as almost to cancel the resulting distortion at the output. A clever idea!

Following the trend already established, of trying to reach *zero* distortion and a frequency response from zero to infinity, amplifier designers saw merit in stuffing in more and more feedback: 10, 20, 30, 40 db . . . amplifiers with as much as 80- or 100-db feedback have been built. If they had 2% distortion without feedback, they should reach .00002% with feedback! Pretty close to zero, if they really do it!

Maybe they do, under measurement conditions. But the acid test, as this book explains elsewhere, is how they handle high-fidelity program material. Critical listening has often found that these well-nigh "perfect" amplifiers do not perform so well on actual program as some with less spectacular figures.

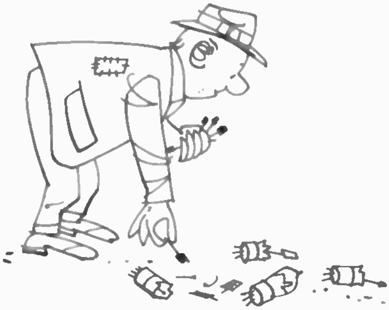
The essence of all this is that a basic amplifier may use any one of an almost infinite variety of circuits. The number of possible good ones must be very large, with so many to pick from. But there are many more possibilities for bad ones. Among modern techniques, any circuit that has been really well engineered is likely to perform indistinguishably from any other of the same rating, equally well engineered.

But being well engineered does not necessarily follow from a statement to that effect in the advertising! Your best test is whether it does what you want it to. We have given you enough of the background of circuit development to enable you to talk intelli-

gently to any circuit whizzes you may meet. If you know when you're going to meet one, a second reading of this chapter will probably make your conversation convince him you know more about it than he does! And you may well find that you really do.

The front end

Next we come to the preamplifier, or input end. The first question often asked by newcomers concerns whether a separate preamplifier or a complete amplifier in one piece is best. At one time there would have been several quality points in favor of the separates, and a budget one in favor of the one piece. Nowadays, however, the choice is a little more obvious: separates provide better, for future flexibility in expanding or changing your system, while the one piece saves cost. Performance must be judged individually nowadays — neither one is predominantly better than the other.



Cathode follower

Where the preamplifier is separate, a common feature of the circuit to get discussed is the output — whether or not this is a "cathode follower." What does a cathode follower buy in performance? It makes the unit more flexible for electrical connection to the power amplifier.

If the preamplifier is to be quite close to the power amplier, and this has a high-impedance input (as most high-fidelity amplifiers do), a cathode follower does not provide any particular advantage. But if the ampliers are to be separated by any length (more than 2 or 3 feet) of shielded cable, the cathode follower will save unnecessary loss of the higher frequencies.

Some claim that a cathode follower enables the preamplifier to be used with a public-address or other amplifier with a "line-impedance" input (500 or 600 ohms). This is not to be recommended as it causes distortion in the cathode follower, whatever advocates of this arrangement may claim.

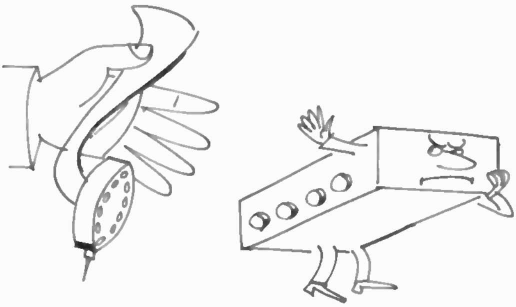
Volume controls

As far as other preamplifier features are concerned, it makes no difference whether the amplifier is separate or one piece, the functions are the same. Every system must have at least one volume, gain or loudness control. We discussed the function of loudness compensation in Chapter 1.

While some prefer an arrangement with separate gain and loudness controls as being more flexible, my own preference is for a straight, uncompensated gain or volume control, with a three-position loudness or contour switch, to suit different listening levels: loud (uncompensated), medium (partial compensation) and soft (maximum compensation). This does all that is needed with less possibility of confusion. At least, that's my opinion.

Sensitivity

You should make sure that your preamplifier has enough sensitivity to work with whatever input you have — magnetic



pickup, ceramic, radio tuner or what have you. It would be good to ascertain that its sensitivity is good enough for the particular brand of pickup or tuner you have, as they vary, and so does preamplifier sensitivity.

Equalization

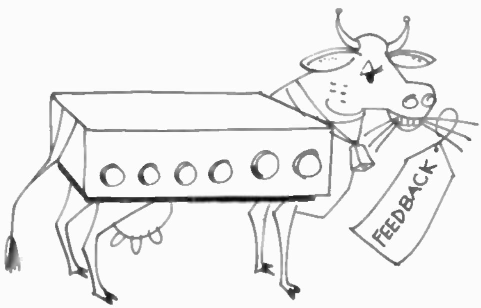
What about equalizers, tone controls, rumble or scratch filters

and features of that kind? Unless you intend to play some really antique records, made before the advent of the RIAA characteristic, more than the one equalization characteristic is unnecessary. Tone controls are a matter of choice, but I feel that some programs need a little adjustment occasionally to make them sound best in my room, although it is true the controls spend by far the most of their time in the "flat" position. Whether or not to have rumble or scratch filters depends on whether you expect to have rumble or scratch, from using a poor turntable or old worn records respectively.

So much for whether you *need* those features. Now for what type of circuit they use (any or all of them): for some time now a big talking point in some circles has been the importance of using *feedback* type circuits. What does this mean?

Sacred cow or pride of the dairy?

A preamplifier often has as many – if not more – circuits with specific functions – equalizer, tone control, loudness control, etc. – than it has amplifying tubes. Then the horrifying thought that,



if you don't have feedback, you must have distortion, means feedback's *got* to be squeezed in somewhere! But if the feedback goes past any control circuits it will neutralize the effect of the control: control circuits and feedback just cannot ignore each other.

So those who feel feedback is necessary in the preamplifier, as well as the power amplifier, to reduce distortion to a minimum, devised the idea of killing two birds with one stone. Rather than mess about with little bits of feedback connected to each amplifying tube, with the control circuits sandwiched in between – which can get very complicated – they put the controls *in the feedback*.

In an ordinary (nonfeedback) tone control, a treble boost works by letting through *more* of the higher frequencies than the lower and middle ones. When the same effect is desired with a feedback tone control, the feedback connection must feed back *less* of the higher frequencies. Then amplification of these higher frequencies is not reduced so much as the others, and the result is a boost. So the feedback circuit works in reverse. Of course, it is quite simple to arrange it so the mechanical direction of rotation in working the control is the same.

When bass or treble boost (or both) is provided in a tone control circuit or in equalization, the extra amplification has to be available for these frequencies. In ordinary tone controls, this extra gain is thrown away at other frequencies, where it is not wanted. With a feedback control, the feedback is used to reduce gain where it is not wanted. If this kind of circuit is properly used, this reduction of gain can also reduce distortion at the frequencies where feedback is operative.

It must be stressed that feedback controls need to be properly designed to bring about this effect. There is no point in using a sloppy design that causes distortion and then using feedback to clean it up. The result will have no less distortion than a well-designed nonfeedback type. Probably the tone control will not work as well, because it is also likely to be clumsily designed, and, therefore, its response will not achieve its purpose as effectively as the older simple circuit done properly.

And there is no guarantee, just because the circuit uses feedback, that distortion will be reduced by it. Feedback can be put to work in the wrong places, where there is little or no distortion, leaving larger amounts produced elsewhere without benefit of its services.

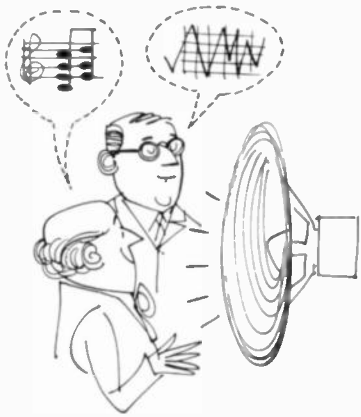
Actually, it is questionable whether feedback is really needed in a preamplifier. A well-designed one, without feedback, produces distortion so small that it is inaudible in all tests that have been made so far. However, a well-designed one with feedback will have distortion a long way below audibility, for whatever satisfaction that may give.

Enter transistors

As this book is being written, quite a few circuits are appearing that use transistors for high fidelity. These offer avenues for a whole new set of circuits, different from those that have been established with tubes. Undoubtedly, in just the same way that each innovation in tube circuits has produced its followers and opponents, the advent of transistors will multiply the variety of opinions still more.

Art or science?

We have the perfectionist attitude of enthusiasts to thank for the vast improvement in high fidelity over the last three or four decades. But the real progress has been due to perfectionists who were not too idealistic.



The audio engineer very often doesn't hear music, even when music is being played – he hears an assemblage of frequencies. And he hears the distortions his training has taught him to recognize. Anything else must be part of the music, and that he does not really listen to.

A musician, or someone with an appreciation for music, will hear something quite different. The kinds of distortion the engineer can hear, the musical ear probably misses, *and vice versa.* This is because the audio engineer listens to frequencies, while the musician or music lover listens to musical instruments.

Of course, there are audio engineers with an appreciation for music, but they are the exception rather than the rule. Generally, when a musician complains that one amplifier produces distortion another doesn't, the engineer, listening for what he "knows" the test figures mean, cannot hear the difference, and concludes the musician is imagining it. This happens so often it just isn't funny.

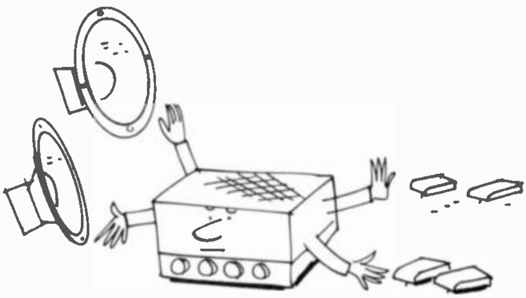
I believe what the musician hears is the more important to high fidelity; that, therefore, the engineer should pay serious attention to what he says, and take steps to educate his own hearing faculty to hear *music.*

Some work has already been done in this direction. Some of the tricks that circuits can pull when reproducing program, that do not show up with standard test methods, have been found out. Yes, the amplifier circuits were at fault, in many instances.

Audio quirks

But sometimes it isn't entirely the amplifier's fault. For example, some amplifiers function better with some types of speakers than others, while other amplifiers "prefer" different combinations. This is not so much a function of which kind of circuit is used – Ultra-Linear, unity-coupled or what have you – as it is dependent on finer points of circuit design. Some recent amplifier designs make provision to adjust for feeding different speakers.

Going to another part of the hi-fi system, some pickups will go



best with certain preamplifiers, because the level and impedances happen to suit one another more ideally. A bad match of this kind will cause distortion or poor response, even though both units are good in themselves.

When things of this kind happen, one or other of the components often gets the blame. For example, if an amplifier works perfectly well with one speaker, and changing the speaker for another type results in poor reproduction, it is only natural to blame the second speaker. Actually, the *combination* may be at fault, and the same speaker would perform excellently with another amplifier.

This kind of thing has been no mean headache to some manufacturers. Several makers of pickups whose products gave excellent quality have had to reduce their output, because some preamplifiers on the market would cause distortion not produced from a pickup with less output. A manufacturer of electric (or electrostatic) tweeters incorporates a filter network in his speaker, the function of which is to safeguard against possible bad effects with some makes of woofer and amplifier. This should be none of his business, but he knows if he doesn't do it, his tweeter is likely to get blamed for the bad effects.

Component vs. packaged hi-fi

This kind of problem is one of the penalties of the component approach to high fidelity: the manufacturer of each component has to allow that his item may be connected to any one of the available products in the adjoining items.

Although this *could* be an argument for packaged hi-fi, it seldom would be a valid one either. While it is possible for a packaged

system to have each part optimized to fit into a complete whole of maximum quality for the price (and a few really do this), it is the exception rather than the rule. Most packaged systems are designed on the well-worn principle of putting in the minimum material on which suitable advertising claims may be hung!

So the choice between a component or package approach has no simple answer. You must judge on the basis of practical performance, using the hints given elsewhere in this book for a guide. The only general remark we can make about this is that a component system of given quality and rating is likely to be more expensive, but it provides better for future changes, upgrading as better items become available, for example.

However, a new trend is slowly gaining ground. This splits the difference between the two extremes. In just the same way that pickups and tone arms (see Chapter 7) should be considered as a single entity, because they work together, there is a move toward incorporating the power amplifier as part of the assembly with the speaker unit it drives. Further extension of this approach will maintain the inherent flexibility of the component approach, while avoiding its pitfalls, and may split the difference on cost too, although this remains to be proved at this writing.

Confidentially speaking

Now you've read this chapter I'll let you in on a little secret.

That electronic whiz who may have bothered you is probably the guy who called me up (he's always doing it) and asked: "Tell me the difference between a phase inverter and a cathode follower" — or some equally ridiculous request. Patiently I started to explain the difference to him by describing the circuits, line by line, over the phone. Then he cut in, "Oh look, I'm not technical enough to understand all that! Can't you tell me in a few brief sentences? I've got to know because somebody asked me, and my reputation for knowing about hi-fi is at stake." So then I told him what you found out by reading this chapter, that they are not directly related, they are for quite different jobs.

No kidding, now you probably know more about circuits than the circuit whiz who was bothering you — or who may yet do so. He may try to convince you how much he knows by showing you schematic diagrams in a book or magazine, or he may draw some himself, feeling he is safe doing so with you, because you wouldn't know the symbol for a tube from a hole in the wall. Don't let that bother you.

You can, without taking your shoes off, count the number of people who can really judge the quality of an amplifier or other piece of equipment from a schematic diagram. Circuit diagrams are not published to show how good a system is. They are to show *how* it works, not *how well.* Any technical man can recognize an Ultra-Linear circuit, for example. But he cannot tell from the schematic whether it's a well designed one.

That's why we haven't bothered trying to fill your head with a lot of diagrams. You still wouldn't be any wiser than your circuit-whiz friend. But by giving you the background, you can now talk to him on an equal footing, without having to learn a lot of things you never need to know — and he doesn't know either. If you are interested enough to want to know more about circuits, so you can recognize them in print and have a little more background about their possibilities, I have written a whole book just on that*.

But suppose, for the sake of argument, that one circuit does have more advantages *as a basic design* than some other circuit. Even so, the latter (with good design) with all the *theoretical* disadvantages against it might outperform a poorly designed sample of the superior circuit.

***Understanding Hi-Fi Circuits.**

Take one modest circuit and mix well

In fact some manufacturers do a really good job of using what are generally regarded as being inferior circuits to get good results. They may not be the best. But by using good design, they use an inexpensive parts list to achieve performance better than some others who rely on using "the best circuits" but don't bother to see they are properly designed.

The transformer — more to be pitied than censured

This trend has been evident all through the development of high fidelity. For example, long before feedback was introduced, the impression began to circulate that audio transformers between tubes give more distortion than the tubes do. It is true they *can*. But a well-designed transformer produces much less distortion than a tube working at the same level.

Designers of professional equipment, for broadcast and recording use, know this. Some of the best equipment used has many audio transformers in it to this day. But, in the hi-fi arena, transformers got a bad name. They had to be eliminated at any cost. So the best circuits were R-C-coupled throughout — even before we had feedback. Their distortion was much higher than professional equipment using transformers, but nobody would believe this. Few took the trouble to measure the difference.

Then came feedback. The way it was promoted, adding feedback to an R-C-coupled amplifier was gilding the lily. Actually, many designs were so poor, they needed all that feedback to make them sound even passable. At this point the injury was compounded. The presence of audio transformers in an amplifier seriously limits the amount of feedback that can be used before the amplifier starts to be unstable. So the advent of feedback brought the final, overwhelming reason for rejecting transformers. For some enthusiasts, even output transformers had to go.

Here come the men with the white coats

But all this was really putting the cart before the horse. The need for so much feedback assumed the amount of distortion brought about by using a poorly designed all-RC amplifier. So, assuming you weren't going to use transformers anyway, you couldn't put them in, because that wouldn't allow you to use enough feedback to be able to do without them. If that sentence sounds crazy to you, it's because it is.

A few ornery souls, not convinced of this logic, have defied it and built amplifiers using transformers (which they had difficulty in finding by this time) with judiciously applied feedback. The specifications may not read the fantastically low distortion figures achieved by some of the "conventional" designs, but a listening comparison showed their faith was justified — even to people who didn't know which was which!

Today, the real reason for not using audio transformers is not the distortion they might produce, but a matter of cost. One reason the cost of audio transformers is high is that not many are made. Maybe, if enough people investigated their merits and used them where performance warranted it, the price would come down to the point where even the cost reason would disappear.

But who's going to start such a move? So many pioneers die before they ever strike gold. I've designed more amplifiers than I care to remember and it's several decades since I designed one with any transformer in it except the output. My client always has to compete with people who make them without transformers.

The same thing goes for making amplifiers that work well but whose specifications are "only as good as they need to be." Manufacturers who have tried that find it difficult to compete with the amplifiers whose specifications show them to be ideal for dogs and bats. An amplifier must be able to deliver its rated watts from 20 cycles to 20,000 cycles, merely because that is the way some "authorities" interpret the specifications. The mere fact that no piece of music you have ever heard or are ever likely to hear contains crescendo power at 20,000 cycles is, apparently, beside the point. Designers have to build that performance into it, even though it doesn't sound so good on music because of what they have to do!

Well, there you are. I've given you enough circuit facts to confound those circuit whizzes who want to boss you around. My advice is, trust your ears. The ultimate test is how well it can produce a hi-fi illusion that satisfies *them*.

chapter 5

radio

THE high-fidelity end of radio should not be confused with ham operations. The radio ham is typically a dxer whose sense of achievement is stimulated most by getting Tokyo or Timbuktu. For him, anyone can get the local station. The high-fidelity radio

enthusiast is content to receive only local stations, but he must do it really well. No "commercial" reception for him.

However, something of the ham atmosphere found its way into high-fidelity radio in the early days before enough work had been done in electronics to know what the score was.

If there were only one station on the air, fidelity would present no problem on radio. But there are many thousands of transmitters, crowded as close as possible, both geographically and in transmitting-frequency allocation, under proper FCC control.

Distance or comfort

For dx, you need a highly directional antenna to exclude unwanted stations using the same frequency but coming from different directions. And you also need a highly selective set to pick out the frequency of your station and exclude others near it that may be coming from almost the same direction.

The fact that the local station is so much nearer than others on the same frequency means you do not need the super-selectivity of a communications receiver.

The best way to cover long distances is by an unmodulated radio wave, called CW (continuous wave). This is just interrupted by a switch, called a "key", to transmit Morse code. Because it requires only the one radio frequency to transmit a continuous wave, the receiver can be made extremely selective – tuned so it will only respond to frequencies within a few cycles of the transmitted signal.

But for high fidelity, the full audible range of frequencies has to be transmitted somehow. They are not radio frequencies themselves, so they have to be converted in some way so they can be transmitted as radio waves and so they can be selectively picked up separate from transmissions from other stations.

Amplitude modulation

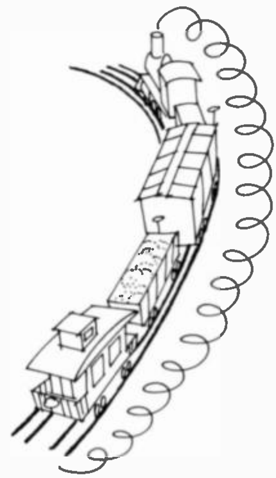
The first way used was to control the amplitude of the radio wave in accordance with the instantaneous sound pressure picked up and amplified from the microphone. Because the *amplitude* of the radio wave is *modulated* or varied to conform to the audio or speech currents, this method is called AM (amplitude modulation).

Only a steady, continuous wave can be transmitted on a single frequency, without requiring more "frequency space" in the "radio spectrum." As soon as you modulate the wave, it needs more space according to the rate at which you modulate it. If you use 10,000 cycles as the audio frequency, a bandwidth of 10,000 cycles on both sides of the radio frequency is required – a total of 20,000 cycles.

Suppose the radio frequency is 1,000,000 cycles, or 1 megacycle. To transmit an audio frequency of 10,000 cycles on this *carrier* of 1 mc requires space at 1,000,000 — 10,000 cycles, or 990,000 cycles, and at 1,000,000 + 10,000, or 1,010,000 cycles. Every audio frequency transmitted requires two such *sideband* frequencies.

Wide and narrow

Now we can see why dx operation and high-fidelity requirements conflict. For long distance, we want to use as few sideband frequencies as possible by making the receiver very selective. For high fidelity, we want to receive the full range of sidebands needed for the complete audio from our station, with uniform response, undiminished.



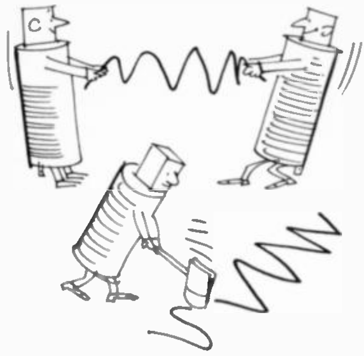
It's relatively easy to make a set as selective as you want by using more tuned circuits to cut off more and more sidebands or to use less selectivity so a wide range of frequencies come through. But a circuit that accepts the range from 990,000 to 1,010,000 cycles uniformly will also accept 980,000 and 1,020,000 cycles almost as uniformly. Maybe 900,000 and 1,100,000 will still come through without much diminution.

Under modern conditions, this range of frequencies would probably contain about 10 radio transmissions, on different frequencies, some near, some from afar. But making a set high fidelity does *not* mean you want to receive about 10 programs at once!

Those square-top coils

What we need is some means of making the radio circuit selective, not to one frequency, but to the precise *range* of frequencies needed for the station we want.

It needs to be a "square-topped" response, level through the



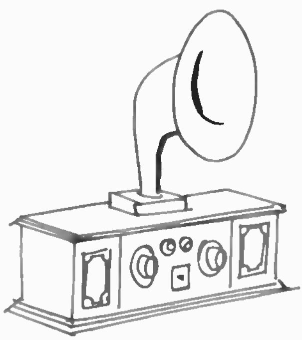
range of frequencies for one station with a sharp cutaway before any frequencies from adjacent channels get through.

If we ever wanted to pick up only one station this might be possible, by careful adjustment of a very complicated circuit. But, as there are many stations transmitting within usable range these days, most people want a set they can "tune" to different stations.

The only way to gain more selectivity is to use more tuned circuits. The more you use, tuned to the same radio frequency, the more selective the set becomes. To make it possible to tune all the circuits at once, to different transmission frequencies, "ganged" controls – a number of tuned circuits controlled by the same shaft – are used.

These ganged circuits have to be made with very good precision. When the shaft is set by its dial for a frequency of 1,000,000 cycles, *all* the circuits must be tuned to that frequency. It will not do for some to be at 1,010,000 or 990,000 cycles, which is only a 1% error, because this would destroy the selectivity of the combination. For this reason, ganged tuned circuits have seldom been made in numbers beyond four.

The only way to get wider response *and* selectivity is to use still more tuned circuits. Two such circuits in what is called "bandpass" coupling respond uniformly to a small range of frequencies and then discriminate twice as much against those beyond the selected range. But still, to get the required sharpness of discrimination in the few cycles between stations we need many more than just four circuits – two bandpass pairs.



Custom-built

At one time the real high-fidelity enthusiast would do it that way since there was no other choice. Generally the receiver used however many circuits that were needed for the job, *on each station.* All these circuits would be carefully tuned for that station and then another set for the next station wanted, and so on, up to six or eight stations. Then a selector switch, usually of the pushbutton type, was used to pick the right set of circuits for the station desired at the moment.

Such a set could be quite good, when all the many circuits were correctly tuned. But setting up the tuning was a job for an expert, and they rarely stayed in position for very long. These problems led to a search for a method of making the tuning simpler to do.

Making it easier

The superhet was the answer. With this circuit, the radio receiver first changes the received waves to different frequencies in

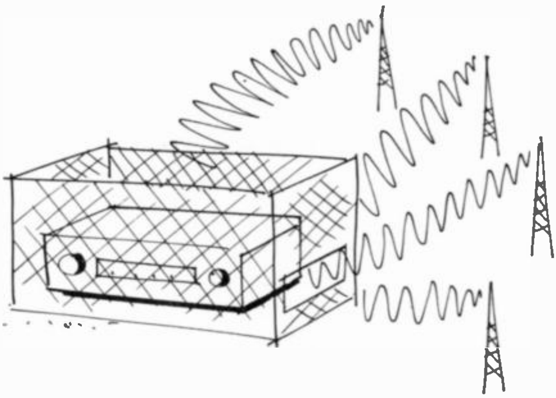
such a way that the one you want is always the same frequency, usually 455,000 cycles (called 455 kc). Suppose you want that 1-mc station. The input stage changes the incoming frequencies by subtracting all of them from 1,455,000 cycles. Now the station at 980,000 cycles becomes 475,000 cycles while that at 1,020,000 cycles becomes 435,000 cycles.

If you want the 1,020,000-cycle station, the input stage is set so everything is subtracted from 1,475,000 cycles, and so on. The advantage of this method is that many more tuned circuits can be used at this *if* (intermediate frequency) of 455 kc, because they can be tuned once and fixed. Then they serve for every station you want to receive, without any resetting.

The superhet has its problems though, which is why, for a long while, many high-fidelity enthusiasts preferred the older trf (tuned radio-frequency) set. The advantage of the superhet is that use of enough circuits can build just about the response you need to get the full fidelity from the station you want and still reject its next door neighbor. But sometimes another station would break through.

Second channel

Suppose the station you wanted used a frequency of 550 kc.



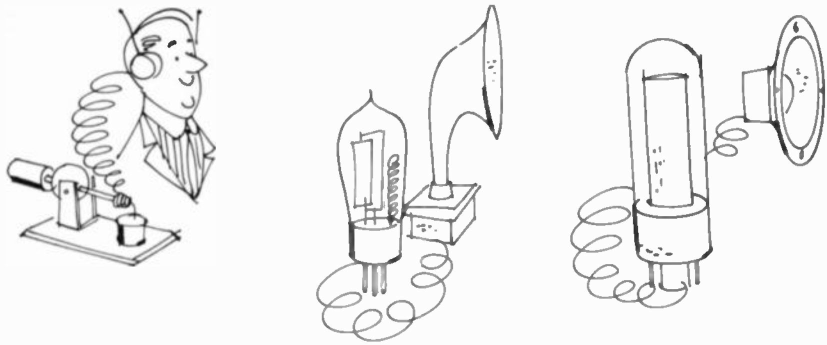
You have to subtract this from 1,005,000 cycles to get 455,000. To do this, the set makes its own frequency of 1,005,000 cycles in a *local oscillator*. But if there is a station at or near 1,460,000 cycles, the local-oscillator frequency will subtract from it and also give 455,000 cyles. So, at one tuning position, you can receive *both* 550 kc and 1.46 mc, in this particular example.

Separation between these two stations depends on the selectivity of the input circuit, which is tuned along with the frequency adjustment for the local oscillator. Early superhet receivers would have had trouble from the "second channel," as this particular spacing between stations was called. But modern design has managed to improve this situation considerably.

The detector

In any radio set, the detector is an important part. This is the circuit that recovers the audio "modulation" from the radio-frequency "carrier." Before tubes were invented, the detector was a crystal with a "cat's whisker." With the advent of tubes to provide amplification, something more reliable was sought that did not have to be "twiddled" until a lucky spot was found.

More by accident than design, the "grid leak" detector was



used for many years. It uses a tube in a circuit that was discovered by accident. Later improved circuits were developed to handle, with less distortion than did the grid leak, the range of radio-frequency amplitudes encountered.

Finally, it was found that a tube counterpart of the old crystal and cat's whisker, called a *diode,* really does the best job. In modern designs, the diode costs practically nothing because it is a little piece of metal with a connection to it, included in the construction of one of the other tubes and using the same cathode.

Wanted — more room

With the best transmitters and receivers that could be designed, it was soon found that the broadcast range of frequencies, 540,000 to 1,600,000 cycles, was limited in its possibilities for high fidelity. The way frequencies have been allocated, there is only a width of 10,000 cycles per channel, with the carrier in the middle.

If adjacent channels were also close geographically this would restrict the audio bandwidth to 5,000 cycles, poor fidelity indeed. But allocations are made to avoid this double proximity problem. Even so, it is difficult to get effective reception with a range up to 10,000 cycles in the audible frequencies. Added to this is a quite severe noise problem, due to radio interference, both atmospheric and man-made.

New territory

So a fresh allocation of frequencies was made for broadcast use, from 88 to 108 megacycles. Using this range of radio frequencies, a separation between carriers of only 200 kc (200,000 cycles) allows 100 channels to be allocated, almost as many as there were on the previous wave band from 540 to 1600 kc, at 10 kc per channel.

In this new frequency range, the radius served by a station is more limited, being not much over the line-of-sight distance, so geographical location could be arranged to provide many more stations in this range than was possible in the old one. And 200-kc separation is more than 20 times as much as before, with no possibility of adjacent channel breakthrough.

In fact, 200 kc would allow a response up to 100,000 cycles in the audio – more than five times what we need. This led to the question whether more of the 200-kc space could not be used to advantage by each channel. Using AM, the highest-frequency audio determines the maximum spread of radio frequencies needed. But if the *frequency* of the radio wave is modulated instead of its *amplitude,* this limitation no longer exists. FM (frequency modulation) does, in fact, result in better utilization of the available radio space.

Enter FM

In AM, the maximum modulation at any audio frequency (or all of them combined) is from nothing up to twice the average magnitude of the radio wave. If you try to drive the transmitter further than this, very severe distortion results. In fact, the average modulation is kept to 30% of this, to allow a safety margin.

With FM, the magnitude of the radio wave does not change, but its frequency can deviate by 100 kc on either side of its average frequency. By limiting this to 75 kc, we have some margin to prevent possible encroachment on the next channel, and still utilize the channel about five times as effectively (depending on the program to some extent) as AM can.

This means FM can improve the distortion and dynamic range, as well as the frequency response, tremendously in comparison with the older broadcast band.

Careful, now!

We should emphasize that this improvement is possible only because the space is there in the FM band. People sometimes ask whether use of FM on the older broadcast band would not improve matters? The answer is no. Changing from AM to FM does not *make* more room; it *utilizes* it, when it's there.

So much for the basic systems. A few refinements will finish the picture of radio's contribution to high fidelity. All radio transmissions suffer from a very wide variation of intensity at the receiver's antenna. This occurs not only because of different radiated powers and distances between transmitter and receiver (in which case a manually operated control could adjust for the differences) but also on the same transmission, unless it is very local, due to changes in atmospheric conditions that cause *fading*.

Holding it steady

Without some automatic control, all except the very local stations would fluctuate up and down in loudness, at least, and run alternately into severe distortion and heavy background noise at worst. So it was automatic gain controls (agc) came to be invented.

At one time they were called automatic *volume* controls, but this is a misnomer. They have no control over audio volume, only over the strength of the radio wave. This they hold to constant level at the detector, so the audio volume always corresponds with the input at the transmitter.

For FM, as well as fluctuations in amplitude, which is corrected in quite a similar way to the AM counterpart, the tuning frequency may drift, which can cause distortion. Frequency cannot change in transit from transmitter to receiver, so any error of this kind must be either at the transmitter or the receiver.

Transmitters have to be rigidly controlled, so the actual cause proves to be in the receiver.

If the carrier frequency is 100 mc and it deviates as a maximum by 75 kc, this is a matter of only .075%, a very small fractional change. If any of the components in the receiver are the tiniest bit temperature-sensitive, they could easily vary this much, resulting in complete detuning from the station.

Quality in an FM set is dependent on accurate tuning, so the average transmitted frequency falls right in the middle of the reception channel. Failure to do so quickly causes distortion.

For this reason, afc (automatic frequency control) was devised to adjust electronically the tuning of the receiver. When it was introduced, it certainly brought about an improvement, unless you were prepared to keep retuning your set about every 15 minutes.

To afc or not to afc?

But some of the more modern FM tuners do not have afc. On the advisability or otherwise of having FM with afc, there are now two stories. Which is right can be determined only for each particular receiver.

One argument says that afc cannot eliminate detuning errors, whether due to careless operation or to temperature effects within the set; it can only reduce them. So these people prefer to design a set without any temperature drift. Without afc, it's easier to tune correctly. That's their story.

The others say that absolutely zero drift is impossible, so afc will reduce whatever there is and make it better. To overcome the objection about ease of tuning, they provide a switch to disable the afc while you tune it in.

Actually, it's a question of how accurately the job is done, either way. For the first one, how well can temperature drift be completely "designed out" to make afc unnecessary? For the second, exactly what does the disabling switch do? Where is the tuning with afc disabled in relation to the variation that may occur when you switch it back in – to the middle or to one side? This is somewhat a matter of chance, as well as good design.

Buttons or knobs?

The advent of switched tuning many years ago started a vogue for pushbutton tuning. And the reason for modern pushbutton tuning is really only a vogue and not the performance question it originally was. Nowadays it's strictly a question of whether you prefer to turn a knob while looking at a dial, or to hit a button.

Interference

Noise and interference have been a severe problem that often proves more serious than restricting dynamic range. On earlier designs, some locations could be rendered useless for radio reception by various forms of interference. Modern design has made tremendous strides to improve this situation.

There are two kinds of noise and interference. One causes a steady rushing or hissing sound. Unless you are looking for real dx reception, this is mostly within your own set. The early radio stages and the tube that changes the frequency from incoming radio carrier to intermediate frequency have been the main offenders. Careful attention to tube and circuit design has worked wonders in getting it down.

The other kind is impulse noise. Rather than being at a lower level than the channel we want to receive, this is often much higher. Atmospheric storms are a natural cause, while all kinds of man-made machinery produce more local, but quite as severe, an effect. The saving grace of this form of interference is the fact that each impulse has a duration too short for it to be audible. Only because it is *big* enough to upset the set's operation for a longer period does its *effect* become audible.

This fact is the basis for all eliminators that work on this kind of noise. In effect, they shut the set off for the very brief instant, so the interference does not get amplified, then switch it back on again. Because the set is switched off only for a matter of microseconds – a period too short to be audible, only the audible sound modulation gets through. No interruption is even audible.

Early eliminators working on this principle did leave something audible, because they put a slight hole in the waveform, where the noise had a tremendous spike before. So the audibility of the interference was tremendously reduced. More modern versions will completely eliminate its audible effect, unless it is extremely severe. This they do by using a sort of "memory" circuit that holds the waveform for that very brief instant, instead of putting a hole in it. When the set goes back into action, it continues the waveform so nearly where it left off that the effect is inaudible.

Why a tuner for high fidelity?

Beginner's in high fidelity generally ask, "What is the difference between a radio set or receiver, and a tuner?" The word tuner is not, as might at first be supposed, a fancy high-fidelity name for a radio receiver. When you have a high-fidelity installation, you already have an audio amplifier and speaker to reproduce the program. All you want is the front end of a radio set, to give you an audio voltage for your amplifier to finish the job.

This is exactly what a tuner is designed to do. It is designed to a higher standard than most radio sets, to conform to the fidelity you expect. It does not give power output to feed a speaker, but just voltage enough to feed into your amplifier. You could use the front end of a radio set and connect across into your high-fidelity system. But you'd probably get a shock.

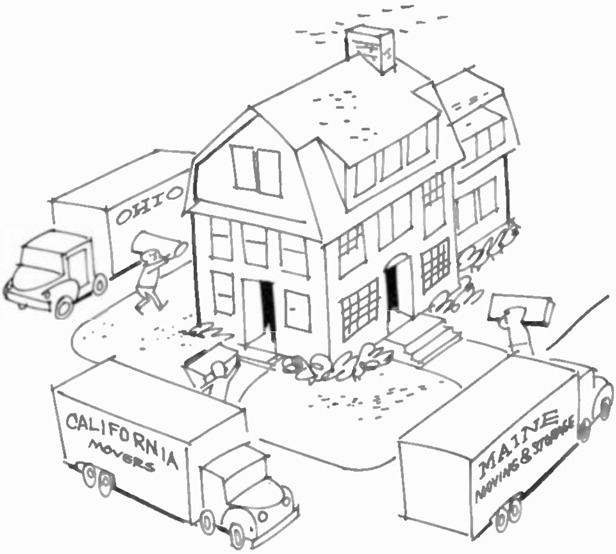
The average radio receiver is designed for lowest cost down to the last penny. Even if they put in more than one speaker, so they can advertise a "multi-way speaker system," everything else is pared down to the minimum that will sound passable when it's all put together. They may have done such a good job that you imagine connecting it into your high fidelity system is all that is needed to make it sound really superb.

That's when you get your surprise. Those cheap speakers prevented you from hearing many of the electronic "flaws" in the set design: background noise, hum, distortion. When you connect it to your high fidelity, these flaws suddenly sound about twice as large as life!

Multiplex

The latest thing in radio is a byproduct of the interest in stereo: multiplex. As of this writing, it is not exactly decided how it will be used, as several systems have been made to work experimentally and none of them has been accepted by the FCC. But whichever one is finally accepted does not materially affect the principles involved.

Multiplex was first devised as a means of sending several telephone messages along one pair of wires. The first conversation is transmitted as it is, but limited to a top frequency of 4,500 cycles. The second conversation has 4,500 cycles added to all the frequencies in it, so it occupies a range from 4,500 to 9,000 cycles.



The third one gets 9,000 added, so its range is 9,000 to 13,500 cycles, and so on. Then the same process is reversed at the other end, to get back the original voice transmitted.

What swings where?

FM seemed a natural for multiplex. You have a much wider band than you really need there anyway. To improve the utilization on both channels, for FM the second channel does not just get a fixed frequency added. Rather, the first channel leaves enough headroom for a complete FM channel to be added above it.

For example, instead of "swinging" the main carrier 75 kc on either side of its average frequency, the first channel may swing it only a maximum of 40 kc. Then a "subcarrier," whose average frequency is, say, 50 kc is added to the first-channel audio, swinging the main carrier steadily at 30 kc, to make a total maximum of 75 kc. The odd 5 kc is used as a "guard space" to make sure the transmission stays *within* its 75 kc. Then the second-channel audio is used to frequency-modulate the 50-kc subcarrier by a maximum of about 25 kc, between 25 and 75 kc.

This means the audio fed into the main transmitter consists of the first channel, plus a subcarrier of 50 kc that swings at a maximum between 25 and 75 kc carrying the second channel. This composite wave modulates the carrier its 75 kc on either side of the main carrier frequency,

just the same as the simple audio did before multiplex.

Using space

To see better how the radio "space" is used up, it is helpful to analyze this operation in terms of frequencies. With AM, the only sidebands produced are spaced from the carrier frequency by the audio frequency modulating it. With FM, more sidebands are produced. If the modulating frequency is 10,000 cycles, sidebands at 10,000, 20,000, 30,000, 40,000, etc., on either side of the carrier frequency will appear, according to how deeply it is modulated.

If the maximum swing of the carrier frequency is 75 kc on either side of its starting point, you can reckon that sidebands will be produced up to about 100 kc on either side. But, however little the modulation, it must at least produce the first pair of sidebands.

The modulating frequency

Thus if the modulating frequency is 5,000 cycles it must at least produce sidebands above and below the carrier frequency by 5,000 cycles. Using more intense modulation increases the amount of these sidebands and starts others at multiples of 5,000 cycles from the carrier. If the maximum permissible modulation were concentrated at this frequency, there would be sidebands at 5,000-cycle frequency intervals up to 100 kc on either side of the carrier — about 40 in all.

Now let's see what this means in multiplex FM. Assume the subcarrier is 50 kc, modulating the main carrier enough to produce just the first two sidebands, at 50 kc on either side. If this is modulated by audio of 5,000 cycles to the maximum, it will have sidebands at 5-kc intervals up to about 30 kc (for 25-kc deviation), producing a range of frequencies at the same intervals from about 20 to 80 kc.

This range of frequencies is then added to the main audio, consisting of frequencies up to, say 15 kc. But these will still be used at an intensity that will produce several sidebands, up to about 40 kc. So, if there is also a 5-kc tone producing full modulation here, it will produce sidebands from 5 to 40 kc on either side of the carrier.

A source of interference?

This looks as if there might be interference. The main carrier is using frequencies at 20, 25, 30, 35 and 40 kc, as also is the subcarrier. But remember these frequencies are only part of the total modulation of each. There is a matter of combined *phase* relationships. At this point the main channel is a fluctuation in amplitude, so the lowest component of the subcarrier, which is 20 kc, will not be audible. This is because it appears due to a change in *phase* or speed of the frequency that is normally 50 kc. This component of 20 kc occurs due to the fact that the frequency dips to 25 kc once every 5 cycles or so.

Because of this the 20-kc components in the subcarrier modulation are not audible — even if *you* can hear 20 kc — they cancel as an amplitude effect. Because the frequency does not dip low enough (only to 25 kc) it never becomes audible. The components of the transmission due to main audio, on the other hand, have different relationships so, when they recombine, the main audio is reconstructed.

Looked at another way, demodulation of the main carrier produces just the main audio plus the subcarrier that was with it: 50 kc swinging up and down in frequency between 25 and 75 kc as a maximum deviation. For any of one channel to break through into the other requires some distortion. This can also happen with straight FM.

Some arithmetic

Suppose a single-channel FM splits the modulation, momentarily, between frequencies of 1,000 and 5,000 cycles. Both will produce sidebands up to about 50 kc on either side of the carrier. The 1,000 cycles will produce a sideband at every kc up to 50. The 5,000 cycles will only pick every fifth, in the same range.

In this case distortion of the relative sidebands could upset the relative intensity of the 1,000- and 5,000-cycle tones at the receiving end. But this is because the relation is an exact harmonic, a 5 to 1 ratio. Suppose the lower frequency were 1,500 cycles. Now we have sidebands at the following frequencies on either side of the main carrier: 1.5, 3, 4.5, 5, 6, 7.5, 9, 10, 10.5, 12, 13.5, *15,* 16.5, 18, 19.5, 20, 21, 22.5, 24, 25, 25.5, 27, 28.5, *30,* 31.5, 33, 34.5, 35, 36, 37, 39, 40, 40.5, 42, 43.5, *45,* 46.5, 48, 49.5 and 50 kc.

Only the numbers italicized are now sidebands of both frequencies. If the originating frequencies have an even more remote numerical relationship, they may not have any sidebands in common. But notice that the interval between successive sidebands is now different. With only 1,000 and 5,000 cycles, the only sideband interval is 1,000 cycles. But with 1,500 and 5,000 cycles, the intervals may be 500, 1,000 or 1,500 cycles. This means that distortion of the sidebands can now produce 500 and 1,000 cycles, in addition to the originating 1,500 and 5,000 cycles. It can also probably produce components at 50-cycle intervals almost anywhere.

Thus distortion can produce quite high order intermodulation components. But well-engineered FM can achieve lower distortion than AM, as well as lower noise level. This means it is quite possible to keep the subcarrier components separate from the main channel.

Subchannel space

What we do need to watch, though, is that the subchannel has "room". Some are recommending the use of subcarrier frequencies nearer the limit – say 67 kc, with only 5-kc deviation. This means that only the first subcarrier sidebands can be used, corresponding to quite small frequency modulation, compared with the main channel. Also the highest frequency that can possibly be put on the subcarrier is 5,000 cycles.

Right now this is causing some argument. Some claim that it's quite possible to use the main audio channel for left and the subcarrier for right (or maybe vice versa) without any "loss of stereo". This we will discuss more fully in Chapter 12. Others want to use some tricks to "splash" some of the high frequencies from one channel to the other at the receiving end.

Keeping distortion down

As far as radio is concerned, the use of multiplex in any way at all means that the need for keeping distortion low, both in transmission and reception, becomes much more important. We cannot help feeling that the simpler the system, and the relation between what goes into the two channels, the better the likelihood of achieving consistently good performance. In other words, the fewer times the frequencies get switched around, the better the possibility of avoiding getting unwanted ones, in the form of distortion.

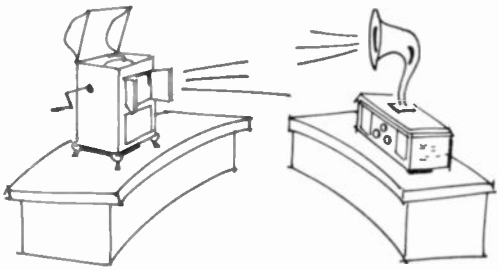
The figures we have used are arbitrary and may not be the ones decided upon for final use. But that's the principle. The remaining question is: Now we have a way of putting two channels onto one radio transmission, what will these two channels be: left and right, sum and difference, or some other combination? This is strictly an audio question and not directly connected with the *principle* of multiplex. So it will be discussed at greater length in Chapter 12.

chapter

6

records

A QUARTER of a century ago, a popular subject for school and college debates was "The Phonograph vs. Radio for Home Entertainment." Sides were often taken, each of which took the view that one would defeat the other because of its "inherent superiority" as a means of entertainment. A few had the foresight



to realize that both had their proper place and they would ultimately prove to be complementary rather than competitive forms.

The same thing happened later between movies and television, and it has been happening to quite an extent also between disc and tape as a recording medium. While a few already see disc and tape as continuing, complementary forms of recording medium, many still hotly argue in favor of one or the other, as if they were contestants.

An age of progress

We can find examples of this in other fields. The horse cabbies fought the electric trolleys quite vigorously as a threat to their

livelihood. Today the horse cabbie, or his son, drives a motor cab, while the trolley driver now operates a bus.

Hand compositors fought the advent of the composing machine in printing. Today many times more are employed in the trade as a result, because of the vastly greater use of print, and the hand compositor still makes up visiting cards and wedding invitations at the corner print shop.

The dispute between different media, where it exists, usually starts from someone employed in the "old" — in this case, disc recording — who sees a threat to his livelihood in the "new" — in this case tape. Something new, rather than presenting them with a challenge to accept new frontiers and methods, threatens their feeling of security.

"Canned" music

All this may seem rather irrelevant in a book on hi-fi. It is much less so than you may think. When recording first got started, many musicians resented it highly for this reason. It would do them out of a job. It certainly has changed the pattern bv which the average person "buys" his musical entertainment.

In the transition, a great many musicians have taken up other work because their "old" employment was no longer available to them. Only top talent was wanted. But today recording companies spend a lot of time and money looking for talent.

The fact is that a great many musicians failed to accept the new medim as a challenge to accept new frontiers.

Who are the top talent but people who accepted the challenge and made the transition with success? Where one way of thinking accepted the advent of recording as a means of enlarging the musician's audience, the other saw it as the way of dispensing with a great many musicians.

The latter view is based on a fictitious notion of saturation. There were only about enough places to accommodate the active musicians of the day so, if records are to be used *instead,* many of the musicians were going to be out of a job.

This was not necessarily true. Many of us still prefer a restaurant where background music is "live" rather than recorded. And the biggest market for recorded music is *in the home,* where paid musicians never had an audience.

Early beginnings

Recording got its start without the aid of electricity or electronics. Both the recording machine and the player for early phonograph records were acoustic. Older musicians still remember the sense of romance in making those early recordings. Often the orchestra or musical group had to rearrange itself completely for the purpose. At least they had to arrange themselves around the "trumpet."

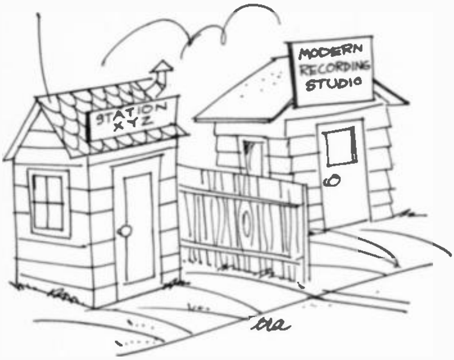
In those days, there were no microphones or booms. The musicians just "aimed" their sound into the large trumpet. The recording machine was just like the old trumpet type acoustic phonograph, working in reverse. Instead of using a hard, already engraved pressing, the sound waves actuated a cutting stylus that made the "master" on a disc whose surface was soft wax.

This was then hardened and a mold made from it, from which copies were pressed in a hard shellac material. The process was much like the way a dentist makes an impression of a mouth to make a set of false teeth.

The big thrill to musicians was that of being able to hear their own performance as an audience would. This had never been possible before.

The first controversy

In those early days, radio and phonograph were completely divorced. Microphones and headphones, which were later replaced by speakers, were appurtenances of radio. The phonograph used a trumpet, both to record and reproduce its "mechanical" groove. This separateness was part of the hostility between the entertainment media.



Electrical recording

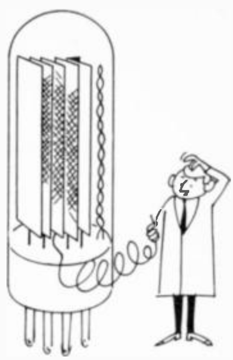
Then electrical recording arrived. To some, this was "stealing from radio." The microphone, with an amplifier, was used to provide an electrical, rather than purely mechanical, drive for the cutter. This made possible a tremendous advance in recorded quality.

It was no longer necessary to crowd the performers round a single trumpet; microphones and booms could be used as they are for radio. And the electrical drive for the stylus made much better frequency response possible than the trumpet and its mechanical linkage could give. Even when the final record was played on the still prevalent acoustic trumpet phonograph, much of the "trumpet" sound was gone. It had a fullness and quality that was a vast improvement.

Pickups

If records could be electrically recorded, why could they not be electrically reproduced? This, naturally, was the next step. Pickups, the forerunners of the ones we use today, were introduced as an add-on kit, to attach to the phonograph tone arm (that carried the old acoustic "sound box"). The pickup could then be wired into the grid of the first audio tube in your radio set and, voila, you have electrical reproduction. The records play over your radio speaker.

For the old phonograph, a turntable speed of 78 revolutions per minute had become standard. This was still used for electrical recording, to keep things compatible. But about this time the movies were beginning to make talkies, and the first talkies used records, with or without some attempt at synchronizing, to carry the sound.



Wanted — more time

For this purpose they needed a disc that would carry more than 4 or 5 minutes of program per side, which was the limit with 10- or 12-inch discs played at 78 rpm. So they introduced the "transcription" disc, which was 15 inches in diameter and ran at 33-1/3 rpm. This played for about 15 minutes each side. The transcription disc also came to be extensively used for "canned" radio programs, for much the same reason.

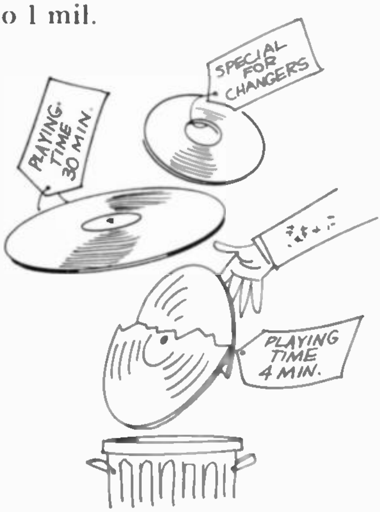
With the advent of moving-coil speakers that improved bass response and tweeters that improved the high-frequency end, radio achieved quality much superior to records of the time. Records were pressed in abrasive-loaded shellac, purportedly to keep the needles sharp, but this made them noisy. There was more background noise on disc than on the local radio station. Pickups, too, were clumsy, compared to modern ones, and this restricted frequency response, even compared with AM radio. And a record carried only about 4 or 5 minutes of program a side.

There was plenty of reason for dissatisfaction with records. For the people who wanted a longer program, without having to keep getting up to put it on, the record changer had been invented, that would play many records automatically. But for high-fidelity purposes, this only complicated matters. A record changer was clumsy and needed a rugged pickup that could not be made to give the good quality that a few specialized makes were capable of doing on an ordinary turntable.

Record sales were not expanding, and many record companies had become defunct or merged with bigger ones. It began to look as if the prognosticators of victory for radio might be right. But many of these problems had been built into the phonograph by the haphazard way the whole thing had grown. Two developments of the late 'Forties proved the radio victory prognosticators wrong.

New things

Chief of these from the fidelity viewpoint was the LP (long play), a disc that ran at the old transcription speed, 33-1/3 rpm. More than this, it used a microgroove and was pressed in a new material. Where the old 78's had become standardized, over the years, on a groove with a radius at its bottom of 3 mils (thousandths of an inch), the new microgroove scaled this down by 3, to 1 mil.



This achieved several advantages. The old transcription discs had a poorer frequency response than the 78's. But this was because the stylus was too big for the speed. Scaling it down to 1 mil actually achieved a response better than the 78's ever could, right from the start.

By using a new plastic, vinylite, without any abrasive, for the LP pressings, the much smaller groove needed much less width to "wiggle" in, the noise from the groove was very much below the previous 78's, and between two and three times as many grooves could be squeezed into an inch.

The result was the discs that are now familiar as LP's, with a playing time of up to 30 minutes per side and performance possibilities far superior to the old 78's.

At about the same time, RCA introduced their EP's (extended play). Based on a slightly different objective, they had arrived at a somewhat different answer. Included in their objective had been an improvement in both performance and cost of the record changer. The old 78 changers had been costly and took quite a while to change from one disc to the next.

The new EP disc was designed to go with a special changer that could be produced at low cost and that changed much more quickly than the old types. Additionally, better quality in the "space" was sought, by using also a 1-mil stylus and new pressing material. Thus the familiar large-center-hole, 7-inch diameter, 45-rpm EP was born.

Between the LP and EP, the old 78 was quite dead by a decade later, but the phonograph industry was going great guns again. Many new companies have been formed and high fidelity is a booming business, largely on the basis of records.

Forms other than disc

The original 33, if not in its present LP and microgroove form, started life as a development for professional use. For the same reasons that 78's had not been altogether satisfactory for home phonograph use, professional people were not satisfied with the transcription disc, and other media were being investigated.

The movie industry particularly wanted to avoid the synchronization bug, whereby a good picture could be ruined by having the sound out of step with the projector. So they wanted sound they could put on film. This led to the optical sound track, which had plenty of problems of its own, but did make the motion picture producers' job much easier, not to mention the local theater projectionists'.

The radio business also explored the possibilities of optical sound, without too much satisfaction. A few used an adaptation of it for a little while, in the form of film about 1/4 inch wide on which a transparent track was scribed by a stylus cutting through an opaque coating on one side.

Steel tape

The real breakthrough for professional radio use was the advent of magnetic recording. A German invention called the Blattnerphone recorded program by magnetizing steel tape.

For a while some broadcasting companies used this quite extensively, notably the British Broadcasting Corp., since they could afford it, being a Government-subsidized company. But it was extremely costly, and the tape, like a clock spring, could prove difficult to handle.

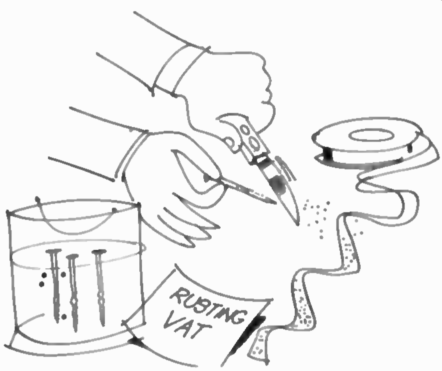
Wire

A saving in cost, but not so much an improvement in handling ease, was the wire recorder that used a thin steel wire. Both of these media require a rather high traverse speed to achieve reasonable quality, because of molecular limitations in the material. With the wire, the "amount" of magnet is so small, due to the thinness of the wire, that background noise is rather high. The wire recorder did achieve considerable popularity for dictograph use, for which it was quite adequate and convenient because the same wire could be erased and recorded over and over again.

Oxide tape

Real headway in magnetic recording awaited the advent of magnetic tape, a magnetic oxide coated onto a plastic film base. From the first it showed improvement over the metallic steel tape, and progress in processing the oxide, bonding it to the film and microfinishing its surface have resulted in a medium that has many advantages.

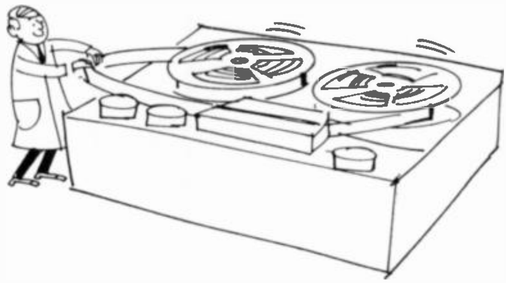
For professional use, ease of editing and the fact that any length of program can be recorded with complete continuity, coupled with availability for immediate playback without any processing whatever, make tape unquestionably the only medium for handling program material in its production stages. For many years now, all phonograph recordings have been taken on tape first. They

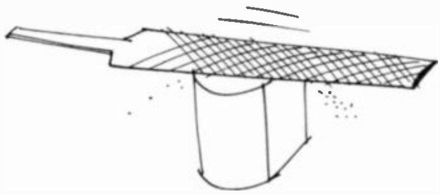


can then be edited, rearranged, copied, mixed and worked on in various ways to get a satisfactory product before transcription into disc form.

A great advantage for any use is the fact that no moving parts are involved in the recording and playback heads, which makes it relatively easy to achieve an almost perfect frequency response. This was the great appeal for high-fidelity use and, as improved technique made it possible to work at as low a speed as 7½ inches per second with ¼-inch tape with quite good quality, more and more of the hard-core enthusiast group took to tape as a means of recording high-fidelity program.

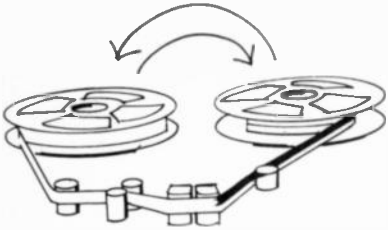
At first they took their program off the air, or recorded it from new records before wear had spoiled them. Later the recording companies, as well as some new companies, started putting out prerecorded tapes as an alternative program medium to discs. This was where the real possibility of competition began to show.





Squeezing it down

Tape continued to go forward, due to improvements in tape manufacture and in head design. Where at one time first-quality professional tapes used at least 30 inches per second, no studio uses more than 15 these days. At first 15 inches was considered the minimum for even "possible" music quality. Now prerecorded tapes are available at 3¾ inches, while 1⅞ inches per second is the speed mostly used for dictation machines.



Both ways

The width of track has also been reduced, even further stressing the tremendous improvement that has been made. Early tapes carried only one track that occupied the full ¼-inch width of the tape. Next half-track was introduced, allowing the tape to be recorded "both ways," giving two sides. This saved the rewind problem, if you wanted to hear both sides in sequence, because you finish up with the tape "the same end out" as you started.

Enter stereo

When stereo began to gain popularity, tape was most convenient as a medium. It had already been reduced to half-track. Instead of recording two sides, both tracks were recorded going the same

way. This was where the disc people saw tape as offering some real competition, unless they got busy to meet it. There had to be stereo on disc.

Actually, the basis for stereo on disc dates back to the days before electrical recording. The first recordings made by Edison and other pioneers had used stylus motion that went up and down in the groove — vertical recording. Later it was realized that lower distortion could be achieved by making the stylus move from side to side in the groove, so the groove would have a constant depth. Modern monophonic records for many years have used this "lateral" recording.

Going after the impossible

Even at this time, some of the early pioneers proposed to record both ways at once, thereby putting two recordings in one groove at the same time. But in those days such a thing was a mere theoretical possibility, like the idea of shooting a rocket to the moon. The practical limitations seemed insurmountable.

The ordinary phonograph recording was not without its distortion problems. Many of these were minimized considerably when feedback (see Chapter 4) was applied to the cutter. The improvements in pickup design following the advent of microgroove, went much further. So some "screwball" types still went on experimenting with the idea of putting two recordings in one groove.

Gradually they proved it was possible. The early attempts didn't do it very well, but well enough to encourage the researchers a lot and other people a little. Most people said it would never be good enough. But the enthusiasts kept at it, and they had quite a task.

The monophonic phonograph had made its improvements by easy stages. Electrical recording made better discs available, then better pickups improved the listening facilities, so the recording people improved their cutters to make still better discs. Finally a complete change of system, from 78's to LP's, resulted in a further series of improvements that have now reached a really high standard.

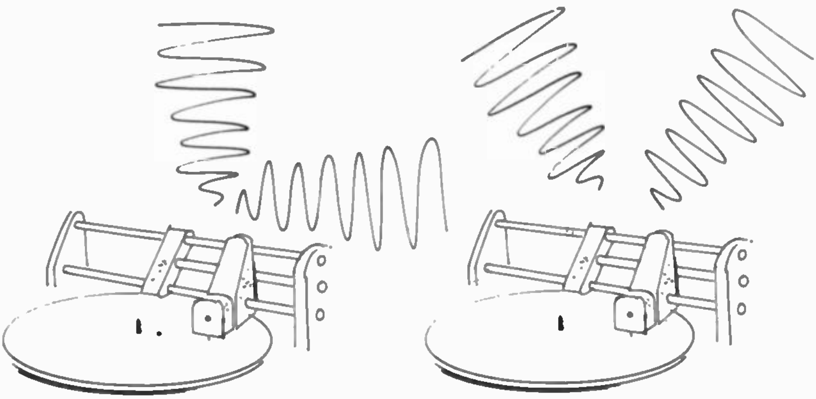
Which came first?

To switch to stereo, we need something that's as good in two channels (still in one groove) as the LP's are in one—or pretty near it. This means we need a recording cutter, some means of pressing discs and a playback pickup, all of this quality at once. We don't have a stereo pickup as good as the previous mono ones to check how good the experimental records are, and we don't have a stereo record as good as previous LP's to check the stereo pickups on. It's a "which comes first, the chicken or the egg?" problem.

By 1958, sufficient progress had been made in this experimental work so that it was possible to give demonstrations whose quality was very close to that of the monophonic records of the day. With well arranged stereo material, the illusion *could* be achieved that it represented an improvement.

Which way?

Most were not satisfied. They wanted both records and pickups to be *as good* as monophonic before they would release them. Also,



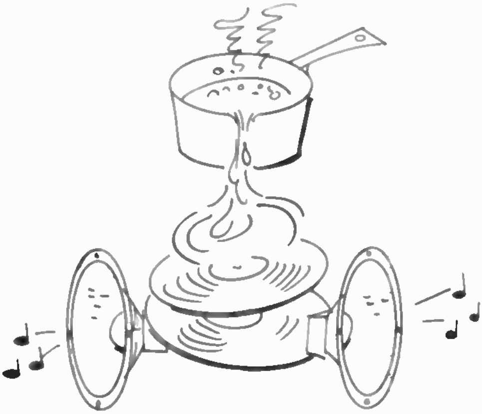
by now, two possible "standards" had appeared. In one, the two stereo channels were represented by the classic vertical and lateral directions of stylus motion. In the other, which aimed at avoiding the unequal distortion of this method, the two motions were at an angle of 45°. This posed another problem for deliberation.

But to a few this seemed like waiting till travel to the moon is as comfortable as first-class cabin cruiser, before shooting a rocket up there. So one record company and one pickup maker did what their competitors called "pulling a fast one" or "jumping the gun."

Out of the pan

This forced the situation somewhat. For good or bad, we had settled on 45/45, rather than vertical/lateral. The makers all had to come out with what they had, good or bad. At first there was not too much that was good, in comparison with existing mono standards.

Because of this, while many stereo systems were sold in 1958. and the market continued to climb in 1959, there were still many



people who preferred to stay with monophonic, because, said they, stereo represents a step backward in quality to get a "spectacular" effect.

Stereo vs mono

Superficially, it must be admitted, this was true. Much of the stereo material put out that first year showed quite inferior quality and relied on the spectacular left/right effects to put it over. Maybe this was inevitable. In any event it was not very unlike what had happened with LP's about a decade earlier.

Although the first LP's proved that quality in microgroove, with speed reduced from 78 to 33, was possible, many LP's then were far inferior to the better 78's of the day. However, nobody would compare a modern LP with an old 78 for quality.

Undoubtedly the same thing is happening with stereo. We shall finish up with stereo representing an all round improvement in high fidelity, discs as well as other media.

Tape vs disc

When discs made this step forward, the relatively new tape stereo, which had been the exclusive source of recorded stereo, looked like it would lose quite a bit of the market. Using 7½-inch

per-second, two-track, one-way tapes, the cost for a given recorded time is much more than from discs.

So further research went forward to improve tape still more so it could become competitive with disc. Quite reasonable quality is obtainable from 3¾-inch speed, using four tracks on ¼-inch tape, two for each direction. This gives four times the playing time from a given amount of tape, and again avoids the necessity for rewind after playing.

Now which is better?

This brings us back to the argument we mentioned at the beginning: which is best, tape or disc? In favor of disc is ease of handling and production.

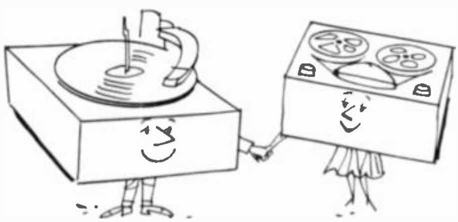
The machines that press discs can stamp them out by the thousands, with relatively simple quality control, as is well established by now. Tape needs to be copied by recording individually on each new tape. True this can be speeded up and operated in multiple, but it is still difficult to equal the simplicity of production with discs.

Discs still are much easier for the average person to handle. To offset this, several tape loading cartridges have appeared that avoid the threading problem. But tape can never achieve quite the convenience of disc.

True, tape *can* achieve better quality than disc. It must be able to do so, because discs themselves are all made from tapes, and something, however small, must be lost in the process. But there is a big difference between professional tapes, run at higher speeds and using wider tracks, and the versions issued prerecorded.

Peaceful coexistence

This relative merit question on a comparable cost basis is extremely doubtful. Both media are constantly being improved, so the right answer today might be wrong tomorrow. It seems to us that the basic answer is much simpler.



The two media are complementary. For recordings you go out and buy, and in which you value ease of handling, disc holds the indisputable advantage. Tape scores in making your own records, either of family and friends, or off the air from radio. People who want to do both will have both disc and tape, and there is no competition.

Those are the basic differences. There are some practical aspects that often influence matters too: how well these systems can be made to work. From this viewpoint wow, flutter, hum and rumble enter the picture.

Wow

The simplest cause of wow (but fortunately rare these days) is having the record with the hole not quite central. This means the pickup follows an eccentric groove, going along it faster when further from the center and slower when nearer. The groove was made with the cutter going at fixed speed, so this change in the speed of the pickup along the groove (or if you prefer, the groove past the pickup) causes a rhythmic change in speed of reproduction.

The change in tempo is scarcely noticeable. What you will notice is the change in pitch. If there are 260 cycles in a length of groove that should be traversed in exactly 1 second, but it is actually accomplished in slightly less, the frequency will be slightly higher than 260 cycles. This cyclic variation in pitch of the whole reproduction is called wow.

It can be caused by a slight tightness in a bearing that occurs only at one point in the turning. Or by an irregularity in the gearing, belt or pulley that transmits the drive from the motor to the turntable. As well as affecting disc records it can disturb tape recorders. Similar deficiencies in mechanical manufacture can cause an irregularity in tape movement.

Flutter

Sometimes the defect causes a much more rapid fluctuation in speed. Maybe the motor, which is geared down by a big ratio to the main turntable or tape capstan has a slight irregularity in speed every time it rotates. This will cause the speed fluctuation to occur many times for each revolution of the turntable or capstan. This, too, results in a variation in pitch, more like a tremolo or vibrato effect that shouldn't be there. It's speed is often similar to that at which metal blinds flutter in the breeze. So this form of pitch variation is called flutter.

Both a disc turntable or record changer and the drive for the tape in a tape machine need precision manufacture to avoid these effects. Either one is susceptible to them, so the best of each will prove satisfactory, while the poorer models will detract from the fidelity due to these effects. It is not possible to say either one is *inherently* better in this respect.

Hum

Also a problem to both disc and tape is hum pickup. Modern pickups and tape playback heads alike need a lot of amplification to make the small electrical output enough to drive a speaker. Everything connected to a 60 cycle power line is radiating 60 cycles. So the wiring associated with the pickup or playback head may well "find" some of this stray 60-cycle-radiation and pass it to the amplifier along with the minute program currents.

Some types of pickup, and practically all playback heads, can also get this radiation directly in their body, as well as by means of the connecting leads. This means extremely well constructed shielding is needed to keep this hum out, and also precautions are necessary not to get the radiating sources too near the pickup or playback head. Such sources are the motor that drives the turntable or capstan, the power transformers in the amplifiers and tuners, and almost any electrically operated device.

Rumble

Again, it would not be right to say that either one is more susceptible to this than the other. But one more thing has to be watched. This is rumble. If any vibrations from the motor reach the turntable, they may make it vibrate. From the pickup's viewpoint, it makes no difference whether the stylus responds to vibrations due to undulations in the groove as it goes by, or whether the vibrations occur in the whole groove, transferred from the turntable. The electrical output would be the same. So such vibrations get amplified and produce a rumbling sound.

The only remedy is extremely good isolation of mechanical vibrations from the turntable. Tape is much less susceptible to this kind of trouble, but by no means exempt. Taken by and large, many inexpensive units, both for disc and tape, do a very good job these days.

Trouble is where you find it

Tape recorders, record changers, record players and turntables are mechanical devices. Assuming that these are always in top-notch condition when brand-new (an assumption that isn't always justified) there is no reason to suppose that they will always remain as good as new. Unlike a new pair of shoes, the mechanical components of hi-fi do not "wear in" to provide a more "comfortable" brand of audio. If your hi-fi system shows some evidence of wow, flutter, hum or rumble you can be sure the situation will get progressively worse — unless you do something about it.

A little care saves repair

From the viewpoint of musical enjoyment, no hi-fi system can be said to be expensive. Considered as a momentary means of distraction for little toddlers, you can very well find it to be the most expensive toy you have ever purchased . . . . when they get their hands on it. A hi-fi system is a sophisticated setup, requiring faithful adherence to the instructions of the manufacturer. This is beyond the ken of little children — and some adults as well.

It takes two to hi-fi

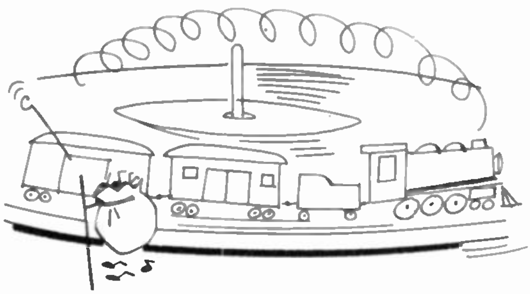
Although both media, hi-fi and tv, are generally located in the same room, they evoke different responses. The reaction of a viewer to tv is passive. The viewer is there to be entertained, and if he is not, the switch to another channel isn't long in forthcoming. But the approach to hi-fi is quite different. The amount of enjoyment is in direct proportion to active participation. This doesn't mean dial twiddling or knob turning. Active participation means listening both aurally *and* mentally; it means appreciation of good musical technique; it means knowing the music well enough to listen for particularly enjoyable passages; and it also means knowing the difference between music to listen to and dinner music – music as a background for conversation.

chapter

7

pickups

IF any component of a high-fidelity system can be regarded as more vital than any other, it is probably the phonograph pickup. True, in one sense it is not even essential — you could use tape

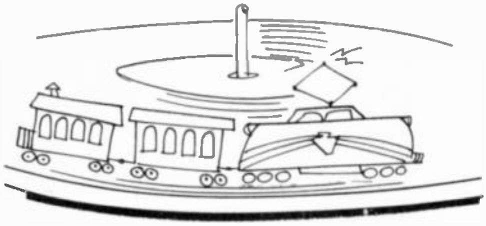


instead of records, in which case no pickup is needed. But, assuming you use a disc as your source material, the pickup probably can do more to make or mar the reproduction — at least as far as your end is concerned.

It is one of several "gateways" in the system. If the pickup does not transcribe what is in the record groove accurately in every way, nothing will correct the deficiency thereafter. Once you have the signal from the groove, by means of the pickup, you can use any amount of equipment — preamplifiers, amplifiers, crossovers, speakers — to convert it into sound waves. But you have to rely on one simple pickup to retrieve the program from the record groove.

Electrification

In the early days, the pickup was a fairly crude attachment that clipped onto the tone arm in place of the existing "sound box,"

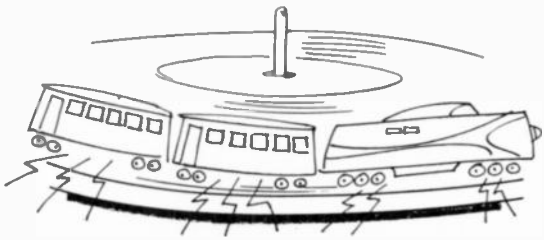


to convert the acoustic phonograph for *electrical reproduction*. The tone arm at that time was a twisted pipe that carried the sound (or "tone") from the diaphragm in the sound box to the acoustic horn that radiated it to the air. The use of the word "tone arm" today is really a carryover from those early days. The arm now carries electrical currents, rather than acoustic tone.

At the time, almost any pickup resulted in noticeable improvement compared with the acoustic version. Transcribing the waves contained in the record groove into electrical impulses that could be emphasized and reproduced over the radio loudspeaker was better. While the result was better than that possible from the old acoustic sound box, it definitely was not high fidelity. There was often quite severe distortion by modern standards. But because more of the original frequency range was present, the sound was characterized as being "fuller".

Integrated design

It was only a short step to make the tone arm exclusively for the electrical pickup. This permitted paying closer attention to



getting the needle angle right and to achieving optimum *tracking* at all points across the usable part of the disc. We'll have more to say about tracking later.

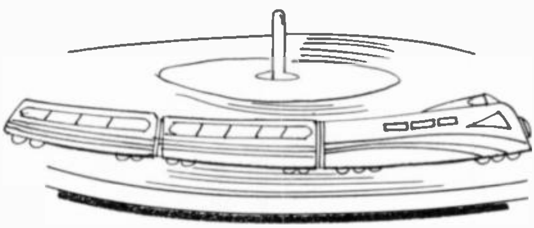
Those early pickups used the magnetic principle. The needle was attached to a magnetic armature that controlled the amount of magnetism around a coil, which produced the electrical output. The needle was invariably held to the armature by a small thumbscrew, protruding from the front of the pickup. This type of pickup remained in vogue with little change for at least a couple of decades.

Those magnetic pickups were good enough for the rest of the system then available, the shellac 78-rpm discs and the best speakers of the day. The improvements effected by more modern types, dramatic as they now appear, would probably not have been noticeable then — if we had the chance to compare.

As recordings improved with the aid of better cutters, eventually using servo feedback, and as speaker response was extended further in both directions, it became evident that these pickups were not all that could be desired, either as regards frequency response or distortion.

Modernization

Frequency response was quite irregular on account of the various mechanical resonances. One of these was the needle screw,

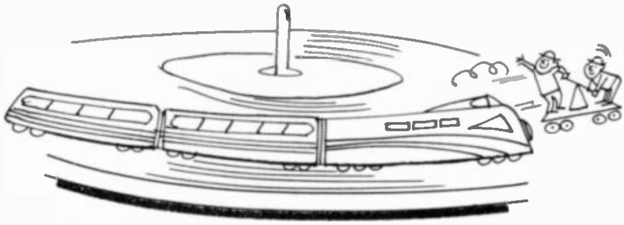


which invariably had a long thin threaded shank, finishing in a large (about 1/4-inch) knurled knob to be turned with the fingers. This caused a very high peak somewhere about 6,000 cycles. Then the mechanical coupling between needle and armature, and the parts of the armature itself, all added more irregularities to the response.

Distortion was due to two causes: The magnetism whose changes produced electrical output also pulled the armature to one side or the other, and this pull tended to accentuate movement at some parts of the wave and minimize it at others. So the wave was not proportionately reproduced. Then, too, to control this side pull, rubber centering pads were used. When new they were nice and resilient, but as they aged, they perished and lost their smooth resilience, causing harsh movement and harsh sound.

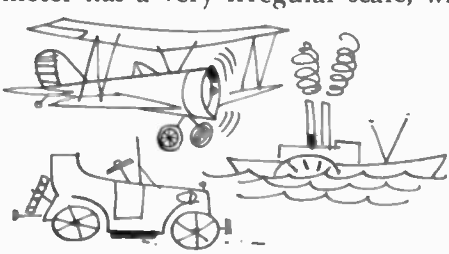
Showing the way

Many improvements in those days can be credited to the do-it-yourselfers. They would file off the knurled knob and use a small



square key for fixing needles, to get rid of the worst resonance. Some of the more ambitious would change the shaping and construction of the armature to smooth off the response some more. Then they would file back the magnet pole pieces, so the magnetic pull was drastically reduced, and replace the rubber pads with more resilient ones, to reduce distortion to a much lower amount. These home-improved pickups produced much better quality, but at the expense of output. This dropped to about one-tenth of that from the original, undoctored version. More amplification was needed, which brought up the noise more, and emphasized the need for more dynamic range.

Along other lines, different people were trying other forms of pickup. They felt that the distortion was inherent in the magnetic type, so they wanted a type that did not have "built-in" distortion. By analogy with meters and the new speakers, many favored the possibility of a moving coil for pickups. A moving-iron meter has a very irregular scale, while a moving-coil type has an almost perfectly uniform one. It was felt that this principle could be applied to pickups to get virtually zero distortion.



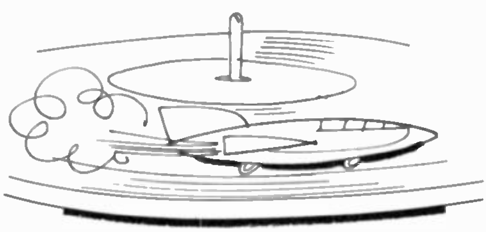
Other ways of getting there

As well as moving iron, or magnetic, and moving coil, other possibilities were investigated, including, variable capacitance and even ribbon. Most successful, though, was the use of a piezoelectric crystal called Rochelle salt. When a thin wafer of this crystal is bent or twisted, it produces a voltage between its surfaces. By appropriate cutting and mechanical coupling arrangements, a pickup with quite large electrical output, compared with other types was possible.

So, for a long time the choice rested between moving iron, or magnetic, and crystal. The others never got further than the experimental stage. The magnetic could be made with reasonable output and not too good quality, or with low output and quite reasonable quality. The crystal had both better output and quality than the high-output magnetics, but, however carefully it was made, it could not match the *quality* of the better, low-output magnetics. It did have the advantage of low cost, so was used in all the inexpensive phonographs, in which the quality was further degraded by their inferior components.

New ideas help

Just before World War 2 a new approach to pickup design



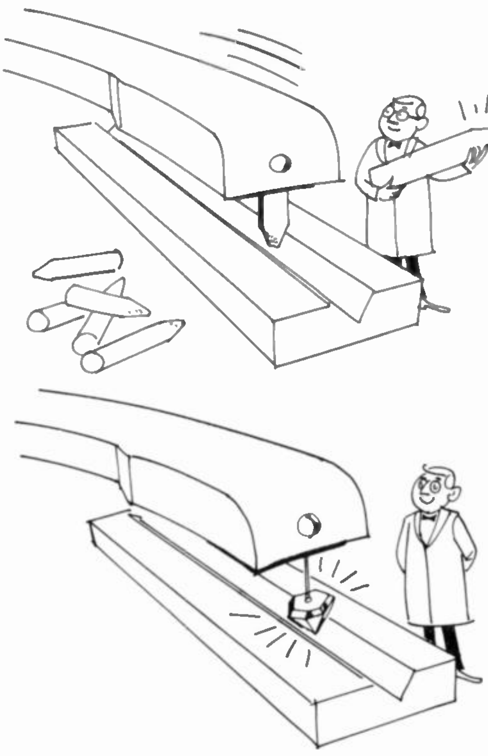
had appeared, dispensing with the hitherto replaceable needle. Up till this time, whichever type of pickup you selected, you used a separate needle, clamped in by the needle screw. There were various grades of steel needles to choose from that contributed their own peculiarities to the overall quality, while some preferred fiber needles, made from thorn or bamboo.

None of the replaceable needles would retain an adequate point for longer than the playing of one record side. The fiber ones did not hold up for even a whole side (about 4 or 5 minutes!). Some of the later needles were of harder steel, claimed to last for 10 playings. But they were in very bad shape by then. People who wanted the best quality and also to preserve their records against undue wear, used these "semi-permanent needles" but played only one or two sides with each one.

There were two reasons for the short needle life and excessive record wear: the needles were not hard enough; and there was too much downward pressure on them. This was necessary with those early pickups, to make the relatively stiff needle mechanism follow the groove. So the new approach attacked these causes.

Expensive, but permanent

The newer types of pickup began using permanent styli, first with sapphire points and later with diamond. They reduced both stylus and groove wear even more by reducing the downward pressure necessary to keep the stylus following the groove. This was possible only by using a much scaled-down mechanism, with very light parts in a much more flexible mounting.



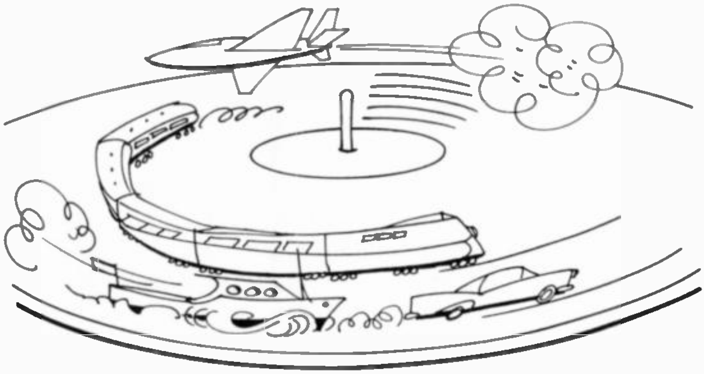
Among other things the tone arm came in for more attention. The pivot now had to be much more free, or, due to friction, it would stop the pickup following the groove across the record.

These new pickups, like the earlier doctored variety, had much less output than the high output magnetics but their quality was vastly improved. The deficiency in output was made good by use of a preamplifier, for which better tubes had by now been developed, so the noise problem was satisfactorily handled.

The advent of LP's, with their microgrooves, was accompanied by further pickup improvements. Now the stylus tip had a radius of 1/1,000 inch instead of three times this, which had become standard with the 78's. With the whole stylus scaled down in size, the mechanism underwent a further miniaturization process, to result in still smaller pickups with even better quality.

What they needed

This is where some of the other types began to become workable possibilities. Several moving coil types have now appeared,

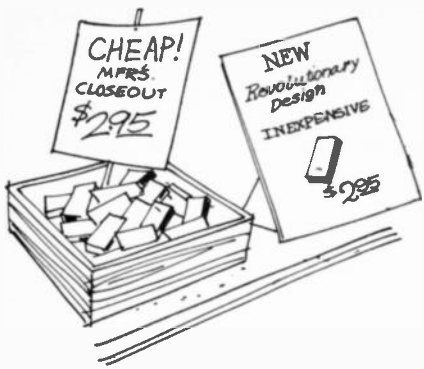


and there is also a capacitance type that uses the variation in capacitance to frequency-modulate an oscillator whose output is then handled by a simplified FM tuner.

This was when the battle of the types started, a battle that continues yet. The moving-coil advocates are still maintaining that theirs is the only inherently distortionless pickup. But that word "inherently" has a rather theoretical significance. *If* a pickup could be made which was perfect in all other respects, then the moving coil would have no distortion and the magnetic's would be very small. But that "if" involves too many things for the theory to have any really practical significance.

We can say right here that the performance of any type must be judged on each model's individual merits. Any type may perform well or poorly, according to how well it is designed and made.

For some idea of the improvement, the weight or tracking force needed to keep the stylus following the groove is a good indication. In the early magnetic types, this was measured in ounces. An extremely good pickup of those days needed a weight of an ounce.



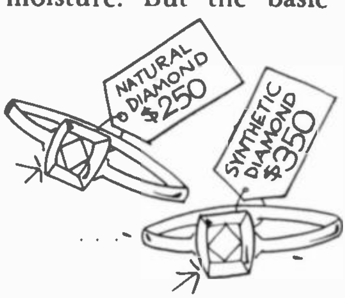
The latest monophonic pickups developed for microgroove, before stereo arrived, needed an average of about 4 *grams* (there are 28.35 grams to the ounce). The best managed to achieve 1 gram, while crystal types ranged from 6 to 10 grams.

Cheap or inexpensive?

So even the "cheap" crystals of modern design need less than one-third the tracking force of the best magnetic designs of a decade or so earlier. But improvement has not stopped at this.

There is always a tendency to couple "cheap" with "nasty." If an item is cheap, it is expected to be inferior. Calling it inexpensive may eulogize the situation for those with restricted budgets, but the people whose quest is quality tend to pass up anything inexpensive as unworthy of further attention.

But the crystal is inexpensive because it is simple. And simplicity in design is a basic requirement for quality when it comes to pickups. True, the earlier crystal pickups had been much too stiff, and their frequency response was poor and distortion bad. Also crystals can deteriorate because Rochelle salt is apt to absorb moisture. But the basic simplicity continued to attract some designers.



Natural or synthetic

New ceramic materials were developed with piezoelectric properties. Completely resistant to moisture absorption, they do not deteriorate, have slightly less output than Rochelle salt for the same mechanical drive in the same arrangement, but are more readily constructed to close performance tolerances.

Just substituting a ceramic element in a pickup cartridge designed for crystal results in an improvement. But designing the entire pickup specifically for ceramic results in considerable improvement. Some of the ceramic pickups achieve satisfactory operation with tracking forces of only 2 or 4 grams. Their performance almost equals that of the best magnetic or moving-coil types. There is no basic reason why a really good new design should not outperform anything else, as well as beating the price of most.

Really, the question of which kind makes the best pickup is, as we have said, a little beside the point. More important is whether, whichever kind, they do the job *right*. A perfectly good transducer, by which we mean a device for converting the mechanical motion picked up from the groove into an electrical waveform, can have its performance spoiled in several ways.

Doing it right

Most important is that the various pivots allowing the stylus to move, of which there are usually three, combine to control that movement strictly at right angles across the line in which the groove itself moves where it passes the stylus. If, for any reason, it moves at some other angle, then the sideways movement caused by the wave in the groove will get distorted.

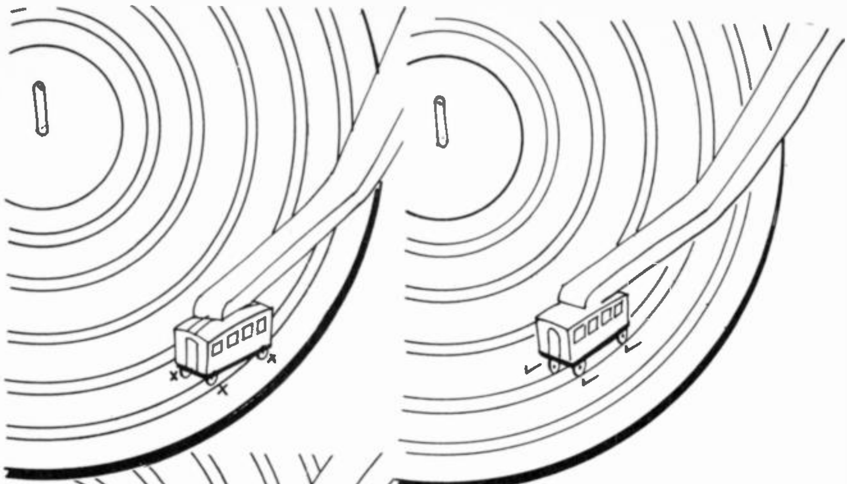
The three pivots found on most pickups, including their tone arms, are: (1) the one allowing the whole tone arm to swing from side to side across the record; (2) the one that allows the pickup to be raised and lowered into the groove; and (3) the one in the pickup itself, allowing the stylus to move according to the waveform in the groove.

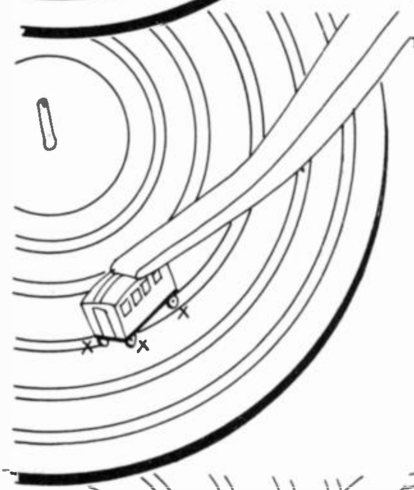
The first two are sometimes located at the same point, using a ball or unijoint construction to allow movement in two directions at the same pivot. Sometimes the pivot allowing the pickup to be raised and lowered is moved along the arm nearer to the pickup.

Tracking

Let's assume first that the movement of the stylus is correctly aligned with the pickup body. This is the assumption made in discussions of tracking. Records are made with a cutter that travels in a straight radial line across the grooves, driven by a slow screw-thread. But pickups are usually mounted on hinged arms, leaving them free to move in a radius across the record.

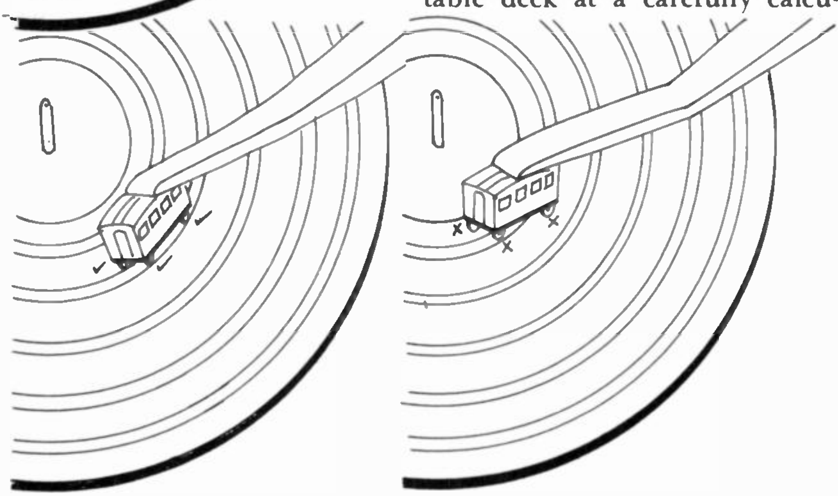
By using carefully calculated positioning of the arm mount on the turntable deck, with a correct offset angle, the stylus movement can be made absolutely correct at two points in the radius range of the disc, and very close to it all the rest of the way. The





offset angle is the tilt between the line of the pickup head, which should be *along* the groove where the stylus touches, and a line from the arm pivot. This is usually fixed by the arm maker, either by making an angle mount for the pickup, or putting a bend or curve in the arm.

Correct mounting for this is usually fixed, either by setting the mounting point on the turntable deck at a carefully calcu-



lated distance from the turntable spindle, or by adjusting this mount so the stylus overhangs the spindle by a carefully specified measurement as the arm is swung across it.

The angle of error that occurs when an offset arm of this type is *correctly* mounted reaches a maximum of about 2°. It is more for shorter arms and less for longer ones. But a 2° error will cause much less distortion than some other things.

Same distortion, other causes

We assumed the stylus movement was correctly aligned with the body of the pickup. But this assumption may not be justified.

The little arm in the pickup that carries the stylus is only a fraction of an inch long. So a deflection to one side of only a very small amount would result in a 2° error. This could happen, for example, due to a side "drag" that might occur because of friction in the tone arm pivot, tending to hinder the pickup from following the inward spiral of the groove.

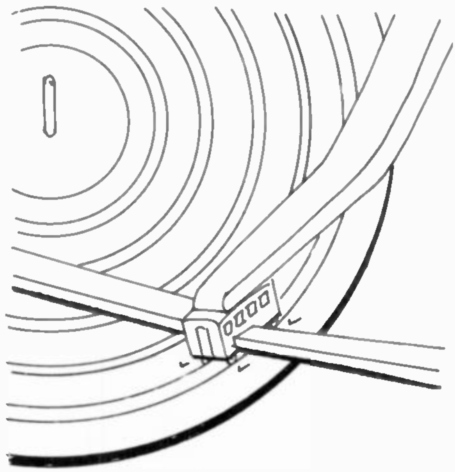
Some have been very concerned because this 2° can occur even when the arm is correctly mounted, and have tried to design pickup mounts with less error, thereby hoping to *eliminate* some distortion.

Radial arm

One method is the radial arm, a straight rod on which the pickup mount slides, rather than hinging in the usual way. This copies the movement used in recording. The problem is that it is much more difficult to get friction in a slide as low as can be made in a pivot. If there is appreciable friction in the slide, even though it is not enough to need more force to hold the stylus in the groove, it may be enough to cause more error in the pickup stylus arm than the little bit objected to in the pivoted arm.

Multiple arms

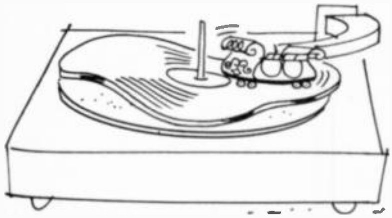
Another method is the rhombic or double arm. This substitutes two almost parallel arms that tilt the pickup head around as it



swings across the record. In theory, this can get perfect tracking at three different playing radii, and the maximum error is down to a fraction of a degree. But again, to avoid throwing away this advantage and more, the four pivots needed, two at each end of the double arms, must have less total friction than the single one on a simple arm.

Until some means is found of guaranteeing that all other possible forms of distortion are less than the residual due to the tracking error of the simple, single pivot arm, we can expect to see this remain the most popular version.

Rollercoastering



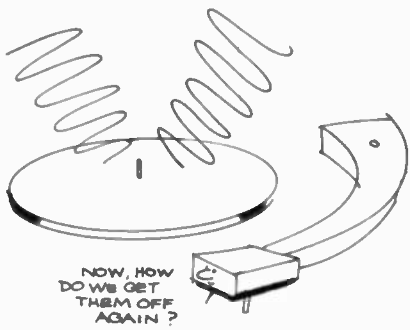
The pivot that allows the pickup to be raised or lowered deserves some attention. If the record is perfectly flat, this pivot has nothing to do while the pickup is playing. But some records get warped. They may be good records that you do not want to throw out just because of this. So the pivot should allow the pickup to ride up and down as the record spins, without causing trouble.

Making the part of the arm that moves up and down shorter will make it lighter and more readily able to follow the warp without needing more downward pressure. But it is also important that the pickup move up and down *vertically*. If the up-and-down movement is at an angle, it will cause some distortion on a warped record. This means the pivot allowing this motion needs to be down as close to the record surface as is practical.

This pivot, too, needs to be very free, with as little friction as possible, because friction will hinder the weight from following the warp down, and necessitate more pressure on the stylus to keep it in the groove.

Enter stereo

The latest development in discs is the stereo record, which puts two channels of program into one groove. Pickups to play stereo records need to give the stylus freedom to move in two directions, sideways (as always) and up and down. Ideally, the up-and-down part of the movement, both in the recording cutter and the playback pickup stylus, should be strictly vertical. Any other angle will produce some distortion, however small.



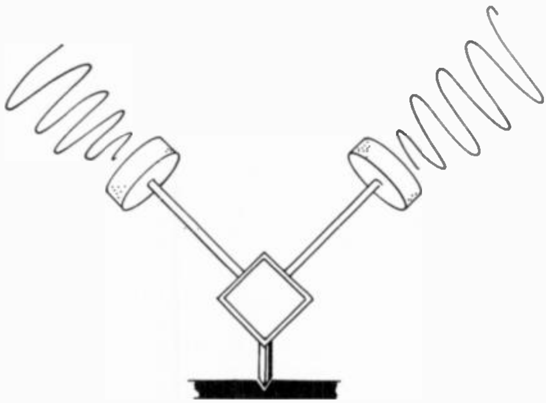
The usual way of mounting both the cutting and pickup stylus is on a pivoted arm. The alternative is to use some arrangement of slides or guides, that seems inevitably more cumbersome than a simple pivot. If a pivot is used, it has to be above the record, however close to it. This means the up-and-down motion of the stylus must have an angle, the same as the line from the stylus point to the pivot must be at an angle to the record surface. The arm may not be straight, but it is this line that matters, whatever shape the arm may be made.

Because practically all designs use a pivoted arm mount for the stylus, the existence of an angle here seems inevitable. If there has to be *some* angle, then lowest distortion results from both the recording stylus and pickup stylus using the *same* angle. Using different angles results in much more distortion than using some angle, but having both the same.

Sorting out the channels

Another problem for the stereo pickup is that of separating the two programs picked up from the groove. Motion at one 45° angle has to give output into one channel, with nothing in the other, while motion at the other 45° angle reverses the situation.

This separation can be achieved right at the stylus point, in whatever device converts the mechanical motion into electrical



output, or somewhere in between. Either extreme can be made into a good stereo pickup, but an in-between idea presents very severe difficulties.

When the separation is made right at the point, the stylus is mounted on a pivoted arm that controls its movement in the proper directions. Then separate mechanical couplings are also taken right from the point to each transducer element.

When the separation is made in the transducer part, this becomes integral with the stylus arm pivot. Movement one way produces one output, movement the other produces the other output.

Problems

With monophonic pickups, so long as freedom for the stylus to move was properly aligned to be from side to side, it did not matter too much where the stylus arm was pivoted. But the advent of stereo has made this critical. This means that what was an excellent mono design may be very difficult to translate into a stereo design.

On the other hand, the fact that the required components of movement are neither vertical or lateral, but at two 45° angles, means types that were not too readily adaptable to first-class design in mono may do an excellent job for stereo.

Fools and angels

Because stereo came in with such a rush — we believe it had to — most of the early pickups were strictly adaptations of the same manufacturer's previous monophonic models. As such, they un-

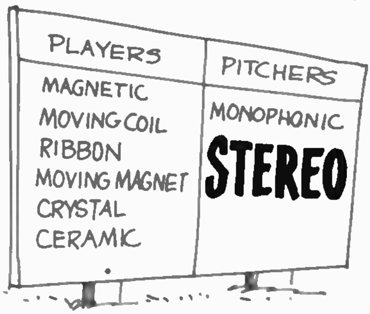
doubtedly did a creditable job in many instances. But the way they had to be made restricted them from doing the best possible. This is undoubtedly one reason why early stereo discs did not get a good reception.

However, new pickups appearing since the manufacturers have had time to *design* them for stereo give much better performance. And the pattern of things, as to what makes a good pickup, has definitely changed since monophonic days.

Changing the score

Just before the advent of stereo, moving-coil pickups had enjoyed a slight margin of superiority over magnetic, but a good ceramic did a *very* good job for the money. A very recent arrival at that time was the moving-magnet type. This differed from the moving coil and the conventional magnetic, in which an armature, rather than the magnet itself, moved. This new moving-magnet pickup gave a really serious challenge to the moving-coil types. People who had tried them asserted they were better. But then came stereo.

Early stereo designs resulted in a definite upgrading for ceramics. Still inexpensive, they no longer merit the implication in being called cheap. Magnetic types showed drastically variable performances. Early ones, at least, seemed very critical of adjustment and could not be relied upon to give *consistently* good performance. But, correctly adjusted, they were good. Moving-coil designs really had their problems. The advent of stereo was most unkind to them. But the moving magnet shows promise of holding the high place it had only started to establish for itself when stereo hit.



As this book is being written, the second round in stereo pickups is only just beginning. We may expect some dramatically new designs to appear. One or two already have, some good and some not so good. In general, the picture has not changed appreciably from the previous paragraph — *yet.*

But we believe it will. Stereo has given the back-room boys of high fidelity new inventive incentive. The criticism, still leveled at stereo by some, that its quality is inferior to monophonic, may be justified today. But we are confident that the future of high fidelity will take stereo in its stride. After working out the technical wrinkles, both quality and realism will take another very positive step forward.

chapter

8

microphones

FEW high-fidelity enthusiasts have any serious interest in microphones, yet this is really where any high fidelity begins. To the high-fidelity consumer, program material comes as a *fait accompli:* radio transmission, disc or tape recording. He may take some cognizance of quality, commenting that a specific program

item is good, bad or indifferent, in a vague sort of way, without knowing just why he gives it such a rating.

He may even go further than this and discriminate between the musical performance and the recording technique. But, if it happens to be a good performance poorly recorded, very few people can tell you in what way the technique was at fault, largely because it just doesn't happen to be their business. The record companies seem to have encouraged this indifference by maintaining a kind of screen over their activities since the early days when they left off using the acoustic trumpet.

Why the secrecy?

If technique is mentioned at all in the record literature or on the jacket, it tells the reader that the recording is "a superb performance, captured with precision by the most advanced techniques known to science" or some such blurb. Occasionally, they may go a little further and give some fancy figures about the frequencies included, and some names of microphones (usually with a German flavor) or cutters (sometimes with an English flavor).

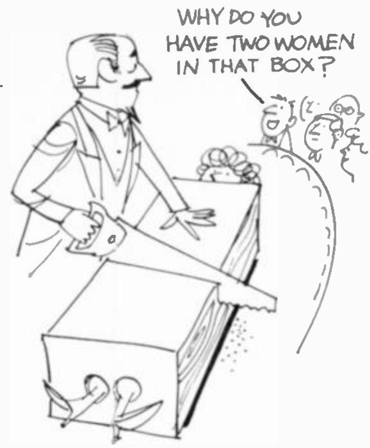
For the absolute layman who says, "I don't want to know anything about what's in it, just sell me a hi-fi," such an attitude is understandable. If each make claims to be best, he has no means of knowing which is, other than by a very superficial listening with his own ears. And the conclusions he reaches are liable to be completely wrong because so many different things can be involved, about which he can know nothing. But it doesn't really matter. Everybody's happy.

However, the reader of this book has good reason to take an interest in microphones, even if he never gets to see one, at least of recording or broadcasting standard. And who knows, with the industry expanding as it is, the reading of this book may be your first step toward becoming one of the people who *make* high fidelity rather than just listening to it.

"Don't spoil the illusion"

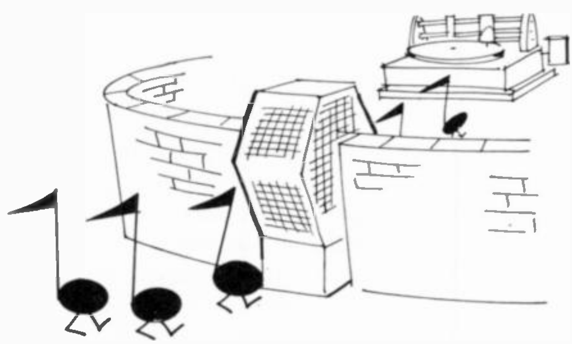
There is probably a reason for the secrecy about recording method. Most of the definitions of high fidelity, and especially stereo, tell you that the realism is such as to transport you so you imagine you are really *in the best seat* of the concert hall. After such a definition of the realism they expect you to imagine, a knowledge of the actual microphone techniques used would be disappointing, and probably puzzling, to say the least.

The layman naturally visualizes the microphone being placed in the position occupied by this fictitious "best seat" so as to pick up exactly what he would hear if he sat there. The fact is you only have to try that method of making a recording just once to discover that it just doesn't work. The result would be very unprofessional, for reasons we shall see presently.



The main gate

In the previous chapter we referred to the pickup as being a vital gateway, if you use phonograph records as source material. But whatever source you may use in your home, the microphones and the technique with which they were used, are a vital gateway between the original performance and the source as you know it. If the microphones fail to get a



satisfactorily balanced, clean transcription of the sound from the original performance, it takes a lot of work to make the final result sound even presentable.

Microphones need, like any other transducer in a hi-fi system — cutter, pickup or loudspeaker — to handle the program faithfully, without distortion. The sound waves reaching the microphone must be faithfully transcribed into fluctuating electrical currents. Microphones can be subject to all of the problems dealt with in the first three chapters of this book, and a few more of their own.

The gate keeper?

It is generally assumed that any microphone used for broadcast or recording would not be poor in any of these measurable respects

– frequency response, distortion or dynamic range, and yet some are. Very few recording studios bother to measure the performance of their microphones. Everything else they measure, but the mikes they take on their reputation. If it bears a good label (preferably German) and responds by giving the appropriate output when you blow into it, it's good!

Maybe a later edition of this book will change this statement (we hope so). But it's a fact at this writing that makers of better domestic microphones, or even other imported ones, have difficulty selling them because of this. Many studio people just are unwilling even to make a simple side-by-side comparison, much less take any measurements.

If a microphone becomes noisy or develops some other extremely obvious defect, it will be replaced and returned for service. But provided it works, it is accepted. One domestic company making ribbon microphones even withdrew this type from their professional line because they found that 90% of the ones actually in use were working, but not anywhere near their specified performance. They may have deteriorated suddenly or gradually, but nobody ever bothered to check them – they worked!

So the high-fidelity consumer is not the only member of the hi-fi community without too serious an interest in microphones. Not that progress in this area has lagged behind that in speakers and pickups, but there has not been such active interest in it.

Early progress

The early microphones were very crude compared to their modern counterparts. The carbon contact mike was very early developed into the carbon granule or button mike, still used extensively for telephones. Because it uses a polarizing current, it really acts as an amplifier, which is why it can send enough speech power from a telephone mouthpiece to actuate a distant earpiece. This was its favorable feature in the early days, before satisfactory techniques had been developed for low-level electronic amplification. In those days any attempt to use other types of microphone was accompanied by very severe noise problems.

Nowadays the situation has reversed. When good electronic amplification became available, along with improved designs in moving-coil and other microphone types, the noise due to polarizing current flowing through the granules of the carbon mike left it much the noisiest, and quite unsuitable for modern high-fidelity work.

In the early days, speakers would sometimes be used as improvised microphones. They are quite sensitive used this way, but have severe quality restrictions. This is chiefly because a good microphone should use a diaphragm (or whatever picks up the sound wave) that is small compared to at least most of the wavelengths. A speaker, naturally, has to be bigger because it has to *radiate* sound, not pick it up.

When the higher-quality moving-coil speaker arrived, a scaled-down version of it formed the earliest type of moving-coil microphone. For the time, it wasn't bad – much better than anything else to date.

Early contenders

Ribbon and capacitor (then called "condenser") mikes had quite an early start, not as an inversion of a kind of speaker as was the moving coil, but designed specifically as microphones. Because of this, they were then regarded as the only really "professional" types. There were two camps: those who favored ribbon or those who favored condenser.

The ribbon has an intriguing directional characteristic that has always proved a useful feature. Ideally, it is sensitive to pickup in both front and back, but completely insensitive on the sides.

Compared with another type, which picks up unwanted sounds from three or more directions, as well as the desired sound from one, the ribbon picks up unwanted sound from only one direction.

For some uses, both its pickup directions are used, in which case it picks up virtually no unwanted sound.

But the early ribbons were no more ideal than their brothers of other kinds. The frequency response was poor and unreliable, and the ideal directional property was not realized because of the cumbersome design.

Condenser microphones were the first to get scaled down to size that enabled superior performance (suitable for early concepts of high fidelity) to be built into them. This was probably because this development did not depend on the innovation of new materials, as was the case with other types. Sensitivity was achieved by the use of a polarizing voltage and progressively miniaturized head amplifiers, built right onto the microphone itself.

This undoubtedly led to the reputation for superiority that they still enjoy in some circles. But any actual test shows that other types have caught up.

Later entrants

Two more types of microphone have been developed at least to the point where modern samples would make any of the earlier so-called high-fidelity microphones look poor in comparison. But at present they rank behind the best in quality. These are the magnetic and crystal, or ceramic, types.

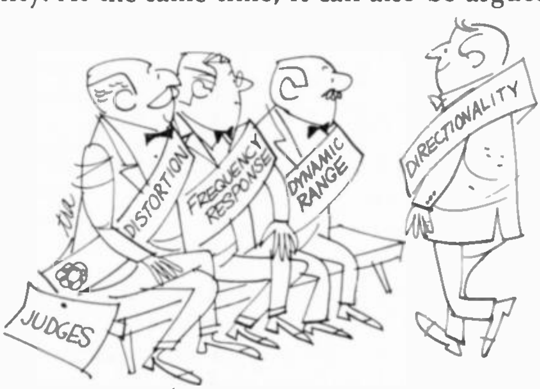
They have the advantage of bigger output, which makes them eminently suitable for use with inexpensive equipment. Basically, too, they are more rugged then the "better" types, which makes them suitable for nonprofessional handling. With the current interest in high-fidelity — and particularly stereo — recording at home, a really superior microphone in this class, at low cost, would find a big market waiting.

The newcomer on the judge's stand

So much for "the field," and their progress according to the "big-three" quality standards: frequency response, distortion and dynamic range. But there is another important feature about microphones: directionality.

Directionally speaking, there were two kinds at first, the ribbon, which is bidirectional (just front and back), and all others, which are essentially omnidirectional (picking up sounds from all directions). This raised the question as to why — or whether — a microphone should have a directional response. It was argued, following that "best-seat" notion, that a microphone should "hear" what a human ear at the same location would hear.

On this basis, it might be argued one way that human hearing is not directional — it receives sounds from any direction with equal facility. At the same time, it can also be argued that human

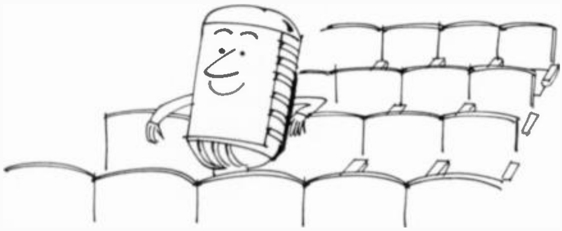


hearing is very critically directional. Sound can be very accurately pinpointed by careful listening. So which should we have, a directional or nondirectional (or omnidirectional) mike?

Mikes for monaural listeners?

Further analysis shows that a person having two good ears must derive his sense of direction from a subconscious comparison between what his two ears hear. Each ear, by itself, is practically omnidirectional. Not completely so, it must be admitted, because a person deaf in one ear can still perceive direction to a degree. This is largely due to quality differences in sounds from various directions, brought about by the obstacle effect of his own head.

If we are to take this as a basis, then a microphone should be an element as small as the hole leading into the human ear, mounted in a facsimile of the human head. While this would pick up sound in a way that would be realistic to people deaf in one ear, it would



not be right for the rest of us. This is why intelligent usage prefers "monophonic" to "monaural" as descriptive of the kind of program that antedated stereo.

Directional listening

Human hearing, by using this two-ear pickup, is able to direct its *attention* quite specifically. You must have done this sometime when in a crowded room where several conversations were taking place at once. You can pick out one person's voice from the babble of others to such an extent that the rest become only a background. You may use lip reading to an extent. But you are also dependent on directional attention and an ability to single out one voice's characteristics from many others.

This no microphone can do. It just picks up sound waves. It does not analyze them. It has no means of finding that certain components of sound came from a violin while others came from a flute. Neither can it tell that certain components of sound come from the instrument itself, while others are due to the room it is in — reverberation. This your hearing faculty can do — *if you are there*. Subconsciously, unless you happen to be deaf in one ear (and that means stone deaf, not merely

hard of hearing), it is by comparison of what your *two* ears pick up that you can do this.

Mikes for binaural listeners

How does all this affect desirable microphone directional characteristics? Probably the answer to this will depend on whom you ask, because there are several ways of tackling it. While some recording people have a very definite idea about the "right" way to do it, the majority have developed techniques of their own, both for mono and stereo, that achieve recordings *they* feel do what is expected.

But, however you approach the problem, it appears to be desirable to control a microphone's directional characteristic. Excluding sound coming from unwanted directions permits placing the microphone farther from the sound source, whether this is vocal, an instrument or an orchestra, without getting too much extraneous sound or reverberation. So a microphone was needed that picked

up in the front only, and excluded the back as well as the sides.

A new pattern

This led to the development of the cardioid pattern. By combining the output from both kinds of microphone this new pattern is obtained. In front, the two outputs add, giving twice as much voltage. At the side, only the omnidirectional part (not the bidirectional part) works, giving normal output, which is half that from the frontal direction. At the back, both portions again give equal outputs, but this time so that one subtracts from the other, leaving nothing.

The result is the cardioid pattern — a much wider area of sensitivity in front than the bidirectional, but with complete cancellation at the back. This forms a third basic directional pattern. Very often, though, it seems that something between the two is desired, so a combination is made in which the two parts are deliberately not quite equal.

Make your own pattern

Suppose, for example, the bidirectional component has twice as much voltage output as the omnidirectional part. In front the two will add, giving a maximum sensitivity from this direction. At the sides only the omnidirectional part is working, which will now give only one-third the output voltage obtained in front (or 10 db less). At the back, there will again be subtraction, leaving an output equal to that from the omnidirectional part, although most of it comes from the bidirectional part. So it is again one-third that from the front. At some position toward the rear, there will be complete cancellation, but not right at the back.

A pattern such as this proves to be more practical for many purposes and it is called the "ultra cardioid." It gives a bigger

average ratio between wanted pickup from the front and unwanted pickup from sides and back.

Whether the cardioid or ultra cardioid pattern is wanted, it can be produced in more than one way. A microphone does not have to be a ribbon to be bidirectional — or at least to have a bidirectional component. Any microphone placed where the sound has access to both front and back of the diaphragm will exhibit directional properties (not omnidirectional).

So by controlling *how much* of the sound wave can get to the back, it is easily possible to make a single microphone *unit* part-and-part sensitive, so it has either cardioid or ultra cardioid response, as required. Using this method, almost any kind of microphone can be made into a directional microphone, with whatever directional properties (within the possibilities) may be desired.

Using your patterns

Two opposing views have grown up as regards microphone technique. One favors using as few microphones as possible, not only because microphones happen to be expensive, but because they believe that is the best way to do it. The other favors using as many microphones as may be necessary to cover individual items in the program. This would be one per instrument or group of instruments, and one for overall reverberation effect, at least.

These views are often expressed without any reference to microphone *types* involved. That is why you encounter quite conflicting opinions. For example, one "authority" may tell you that microphones cannot be operated too close together, say within a few feet, without running into undesirable phase effects, while another may say it just has no effect. Both can be true in different circumstances.

If you use directional type microphones, particularly bidirectional or ultra cardioid, then they should not be put too close together and connected to the same circuit. There *will* be phase interference patterns. But if you use all omnidirectional types, you can put them as close together as you wish, provided you make sure they are *correctly* phased, without this kind of problem.

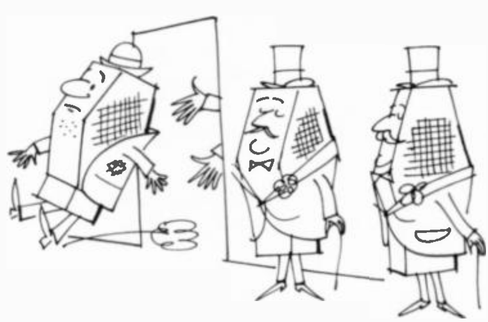
Which is best?

This might look like an inherent advantage, one way or the other. But it isn't really. It means only that placement technique must be coordinated with the type you use. With directional microphones, they *can* be placed farther part, because you can use their directional sensitivity to maintain balance and keep the wanted sounds enough above reverberation and other background effects. With omnidirectional microphones, you will probably have to put them closer to the sound sources, which may also mean more of them, closer to each other.

By and large, the overall picture seems to be (aside from possible prejudice one way or the other) that a good selection of omnidirectional microphones — as many as you may need — is the *easiest* way to get acceptable effects without any special care. But if you want to do the best possible job, by exercising a little more care, a fewer number of appropriate type directional microphones will do a better job.

Transients here, too!

Whatever type of microphone is used, it must have the smoothest possible frequency response. Not only should it be smooth as

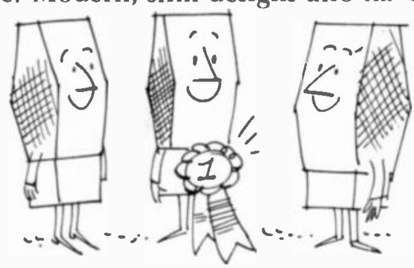


measured by the standard method, it should have good transient response. And, if it is a directional type, the response should be as smooth as possible in all directions. Expressed differently, the ideal directional properties should be effective as nearly as possible, at all frequencies.

A cheap directional microphone may have quite a good directionality, according to its design, over the middle range of frequencies, and then lose it at higher (or maybe lower) frequencies. This would be useless in a night club, for example, where the high-frequency components of the background (noises from cutlery and platters) would not be kept in the same proportion with other parts of the background. The same thing is true in a studio. Poor consistency in directional response can give quite a coloration to the reverberation.

The winner?

Generally speaking, this fact offers a slight edge to the bidirectional microphone. On the average, this maintains better uniformity in its directional response. Modern, slim designs also have an overall frequency response second to none. But you do have to be little more careful about placement than with the cardioid or ultracardioid patterns. They are ideal for dialogue or when there are two or more sound sources. They can be put between the two sources, enabling them to be spaced very well apart and still use only one microphone between two.



But you do have to watch that the performers do not creep around to the side, where they are "dead." Sometimes this dead area can be useful, too. In a small musical group, it will often be found that the brass (if you have any) is much too strong, compared to other parts — even more noticeably so when you try to "mike" it than just listening to it. You can put the offending brass almost on a dead spot, as well being a little further back, and quite good balance can be achieved. It will also bring out more of the "character" of the brass than using more conventional techniques.

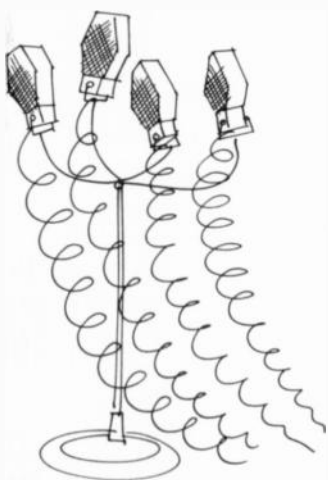
Bistereonauralphonic

For stereo, all this conflict about microphone methods has multiplied. There are people who still maintain (or did until very recently) that you should not use more than two microphones, or maybe four (two on each channel), for very big pickups such as orchestras. These are the people who felt that the use of many microphones, even on monophonic recording, with a separate but more distant one, for picking up reverberation, was "phony" on principle.

But when it came to recording stereo, these same people had to admit that "there is a lot of luck" about the business. Out of all the recordings you take, they tell us, very few are really good, most are mediocre and quite a few "smell badly" and have to be thrown out. Yet they are unwilling to concede that their principles might be partly responsible for this situation. It must be some of the many things "we still don't understand" about stereo!

As we shall discuss in the chapter on stereo, the early conceptual ideals are not altogether reliable. Acceptance of this fact enables more progressive people to adapt themselves to the needs of the new medium more effectively.

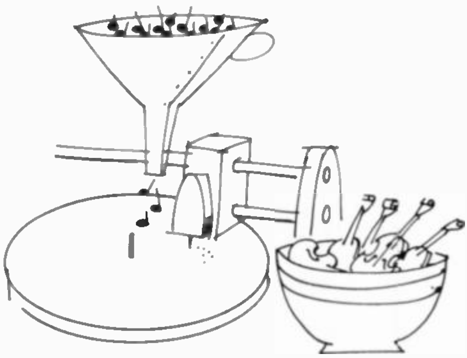
And intelligent listening will improve your appreciation of stereo, give you an insight into methods and help you set up your "receiving end" to get the best results. So your patient interest in mikes will pay off.



Add strings to taste

One thing does favor the use of many microphone channels—as many as you may need according to the number of component sound sources. Since the advent of tape recording, it is quite possible to make a master recording of as many channels as you wish, at least professionally. The home recordist is more or less limited to two. But this results in two advantages for the professional recordist.

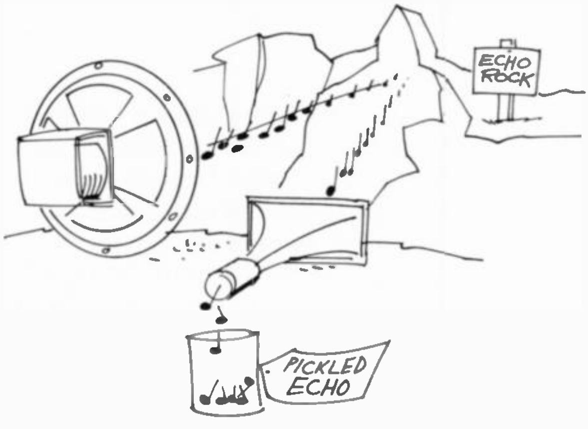
Musicians' time, as well as that of other performers, is expensive. Also, to get their best *performance,* more than one "take" may be necessary. With the simpler microphone techniques, it might be necessary to try over different mike positions in successive takes, as well as getting slightly different performances each time.



By using multiple recordings of each take, the best performance can invariably be used, without having to bother whether "we also picked the best mike placement of it." And the best can be made of any particular take by remixing into final release channels, long after the musicians have gone home.

. . . And echo

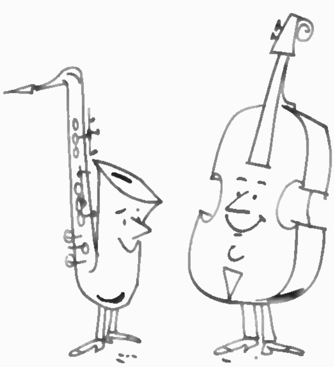
Closely associated with microphone technique is the matter of reverberation. As well as covering it in the ways already mentioned, some recording companies avoid making appreciable reverberation in the studio, and then use special "echo chambers" to add reverberation "to taste."



An alternative to this is to use electronic reverberation generators. In their simplest form, these are just tape recorders that record the sound and play it back a moment later for recording. More elaborate variations include filters and multiple rerecording facilities that simulate building reverberation more accurately and with more precise control.

Musicians are people

But the use of artificial reverberation has one serious drawback. Natural reverberation has a useful function in the original performance. Musicians play with a "feel" for the sound they create. Much of this is due to reverberation effects. Very few musicians like working in an "over-dead" studio, where they get the impression their instruments are not responding to their touch normally.



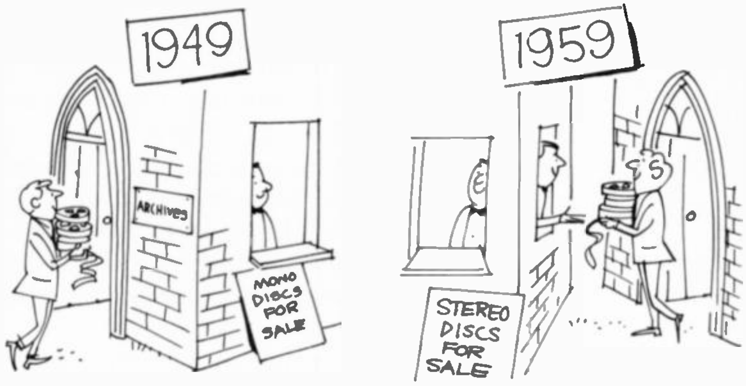
This fact emphasizes that studio acoustics and other aspects of the surroundings at the original performance can be important not only to the recordist. They can actually affect the performance itself, by their effect on the performers, as well as affecting the way that performance is transcribed into a recording or transmission.

In some instances, bad as the acoustics may be to the recordist, it may be well to use a studio that is poor from the viewpoint of acoustics, because it has the right "atmosphere." Maybe even the tradition of the place will help the artist to do his best, if it happens to be the Met or Carnegie Hall.

Stereo for nothing

Following the principle of multiple-track master recording has carried another bonus for many of the leading record companies. For some years before stereo on disc became an actuality, they had been making multitrack master recordings. At the time these were mixed to make up good monophonic releases. Then, after stereo became a fact, the same originals could be remixed to make stereo releases.

Of course, this put companies that had not been doing it on the spot. They either had to get frantically busy making new stereo recordings or fake some of their existing library. Nobody has owned up (publicly, at least) to such faking, but this probably is another reason why so many "questionable" records appeared in the early days of stereo.



Home recording

"But what microphones should I get to make stereo records at home?" is a common question these days. To some extent the answer to this depends on what kind of program you want to record. But there are some important principles.

Single channel recording of quality quite good enough for most people could be made with the inexpensive mikes that used to come with the recorder. For stereo, the quality of the microphone is very much more important, to get acceptable results. Most important, whatever other characteristics it may have, is a flat, smooth frequency response. This is more essential in making recordings in the average living room, than it is in many professional studios. If you are ambitious enough to want to do your own stereo recording, it will be worth your while to get microphones from a manufacturer who supplies an individual frequency response with each one he sells. Don't expect it to be *perfectly* flat – such a mike doesn't exist. But the fact he takes that trouble means you can be sure it's a better mike than one where this trouble was not taken.

It would be nice to see a good, inexpensive cardioid made for this purpose. Undoubtedly before long some manufacturer is going to "break the ice" and make one. Then all the others will follow suit. Meanwhile, the slim-shaped ribbon microphones are the best buy for this job.

Many have been able to achieve acceptable results with the better quality microphones using either dynamic or piezoelectric transducer elements, and designed to have what is termed a "semidirectional" pattern. This usually means they favor the front pickup of sound progressively more at higher frequencies.

But this type of microphone has to be placed considerably closer to the performer (s) than the ribbon or cardioid types. The trouble with this can be that the performer may not be aware of the importance of "staying put" and might move a little bit during the performance, spoiling the recording. Many musicians cannot do their best working in an imaginary straight jacket, either. This is why the cardioid or ribbon is so advantageous. Then you can have the microphone a "comfortable" distance away, so the need for holding a precise position while playing is not critical.

You may be helped in group recording by remembering that bass notes are not particularly directional anyway. A string bass transmits most of its "body" sound through the floor. This can cause unnatural pickup if the microphone is too susceptible to vibrations transmitted from the floor, via its stand. You may need to place the stand on a heavy mass of sponge rubber or some other form of vibration isolation. Or you could suspend the microphone on long elastic cords.

Having ensured that you only pick up sounds via the air, the same as you hear, you can then give due attention to instrument placement. Although these low frequencies themselves do not convey any sense of directionality or location, the same instrument's transients can.

chapter 9

speakers

ONLY when someone humorously refers to a baby's wailing as a speaker, or when we realize that the French *"haut-parleur"* is a more or less literal translation of "loudspeaker," do we have cause to reflect on progress that has been made with this particular component. The quest for perfection has been stressed so much, and so much progress has been made, that we are now more aware of the concept of transducing electrical power faithfully into sound waves than we are of achieving a loud sound. Or are we? When one attends an audio fair, one wonders!

Some of us remember when radio reception meant sitting in a group round a pair of headphones laying in a cake pan, connected to a carefully tuned cat's whisker-and-crystal set. With rapt attention the group listened to the amazing phenomenon. Then tubes came in, and the bigger triodes, giving all of 25 milliwatts audio output power (and flattening rechargeable A batteries at about two per diem), could produce enough sound to be heard without sitting in a group round it, from the new *moving-iron speaker,* an overgrown headphone mechanism driving a large paper cone.

Progress

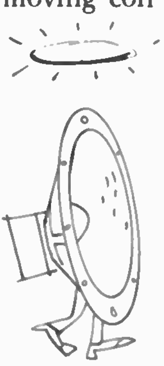
The early speaker certainly was not a reproducer. The object of a modern one is the reproduction of sound waves in as close a facsimile to the original as possible. But the crude term, deriving from the days of crude beginnings, still persists— loudspeaker.

Where all the distortion came from that made those early loudspeakers more like loudscreechers, we probably shall never know. It wasn't measured in those days. Most of the low-frequency amplifiers (not yet dignified by the title "audio) used two or three transformers, which added to the loudness and also the distortion.

But R-C coupling, pentodes, push-pull and all the other refinements discussed in Chapter 4 started reducing distortion from the set, and it became evident that the speaker was not exactly perfect!

Enter the moving coil

The moving coil was the magic formula that started the high-fidelity trek among speakers. Instead of using a big coil, fixed to the magnet structure (which controlled the pull of the magnet on the moving-iron armature to which the cone was attached) a small moving coil carried the output current, suspended in a circular space between the magnet poles. It could move freely over a wide range of frequencies and was not subject to the distortion produced by the stiff armature mechanism.

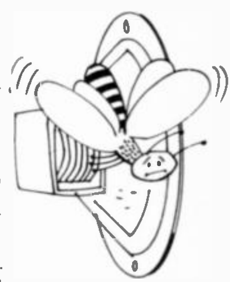


At first, the use of a smaller coil meant the sensitivity or efficiency was reduced. You needed more power to drive a moving coil to get equivalent sound from it. But when you got it, it sounded so much fuller and *cleaner*. By modern standards, those magnets were terribly weak, not much better than primitive lodestone.

You had two choices: permanent magnet (PM) or energized. Energized used an electromagnet, with a big coil to push more magnetism into the thing. This coil soaked up more electrical power from the A-battery, the B-battery or wherever the power came from. With the later indirectly heated tubes and eliminators to obtain the A- and B-supplies from the power line, we could better afford the energized type and most of the sets of the era were characterized by energized speakers. Very often they used the speaker "field" coil as a choke to help smooth the B-voltage after rectification, thus killing two birds with one stone.

The hum bug

This led to another trouble: hum. This could either come from the set, because the supply was inadequately smoothed out, or from the speaker, because the hum current in its energizing coil got into its magnetism. Careful supply design got it out of the set, and a hum-bucking coil eliminated its effect from the speaker.

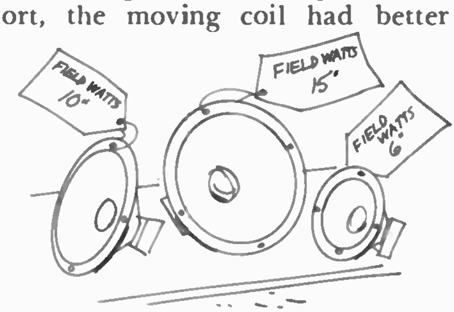


This worked by using two coils connected to the set. One was the regular moving coil, while the other was a similar but fixed coil. This picked up the hum magnetism, and was connected so its effect in the moving coil bucked that due to the magnetism.

An energized speaker with a hum-bucking coil was only half as efficient as one without it, but it still managed to beat the PM type, which didn't need a hum-bucking coil.

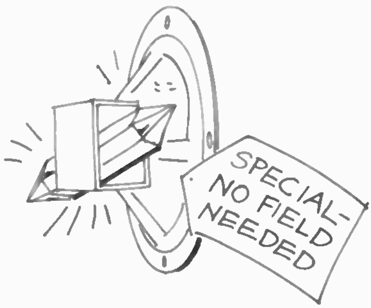
Field watts

Using an energized magnet rather than the much weaker permanent type had another advantage: better frequency response and damping. In short, the moving coil had better control over the cone movement when the magnetism was stronger. So the quality of a speaker was judged very largely by how many watts you put into its energizing coil. The more watts, the stronger the magnetism and the better the performance.



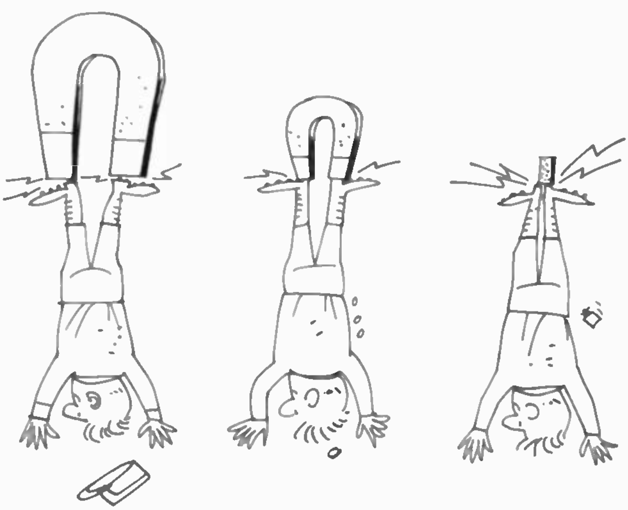
Crystals — no field needed

About this time, the high efficiency of Rochelle salt as a piezoelectric material became known. Phono pickups and microphones had been made using it, and larger pieces went into a piezoelectric speaker, driven through a lever system, to get the big movement needed, with less stiffness in the drive. This was more efficient than all except the very heavily energized moving-coil speakers, and promised to take over, at least for the low-cost market. The piezoelectric did not meet the quality of the moving coil.



The wonder magnets

Then the metallurgists saved the day for the moving coil by developing some fantastic permanent-magnet materials. These produce between 5 and 10 times the magnetism from a magnet of comparable size or weight. This development tipped the scale the other way. Now the permanent magnet is undisputably better than the energized. If you want more magnetism, just use a bigger magnet. And no hum-bucking coils are needed. With this development, the piezo speaker enjoyed a very brief life



before complete dominance was regained by the moving coil.

Because moving coils have held dominance, with little competition, for more than three decades, many look on them as the *only* kind of speaker. You can have large coils, small coils, thin coils, thick coils, but moving coil it has to be. You can have large or small cones, round or oval cones, thin or thick cones, straight or curved cones, or dural diaphragm in a pressure-driven horn, but it is driven by a moving coil.

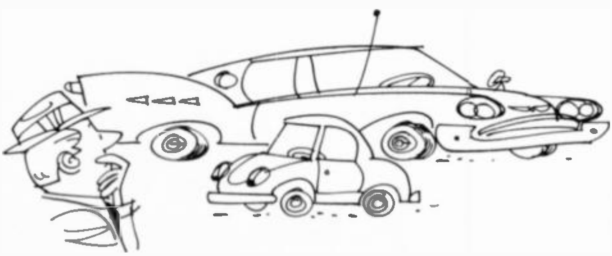
Top place forever?

A long unbeaten setup like this can have two effects. Any other type does not get serious thought from most people, by force of habit. And the moving-coil type becomes rather "worked out." All the possible variations have been tried, it seems. An "epoch-making change" will consist of substituting a new plastic for the previous paper cone!

However, the moving-coil speaker has had, and continues to have its problems. It is far from being the ideal speaker. Let's trace the course from audio power delivered by the amplifier to a sound wave in the room, and see what these problems are.

Motor efficiency

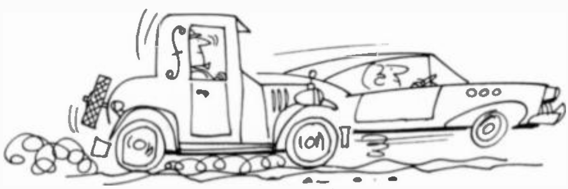
The power is delivered to the cylindrical voice coil, held in the magnetic field, which should transform it from electrical into mechanial power, driving the coil back and forth along its axis. Here we encounter difficulty one: to get maximum efficiency we want maximum current flowing round maximum turns in the coil, working in a magnetic field of maximum intensity.



To get this the cylindrical coil must work in a gap between magnet poles with practically no clearance. If it becomes only the slightest bit out of round, or leaves its proper centered position, it will rub.

A smooth ride

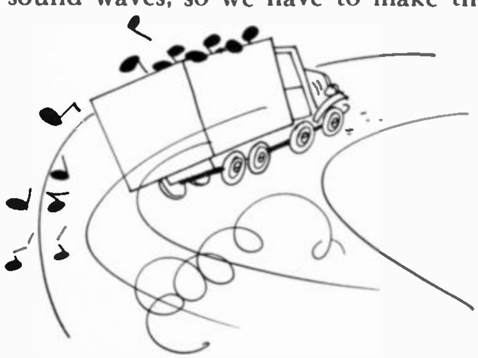
So we need an accurate centering device to hold the coil in position while allowing it free movement the way it's supposed to move. This is difficulty two: to get a centering device (called a spider, because earlier speakers used a device that looked like one) that holds the coil central while allowing it free back-and-forth movement, without causing distortion.



Another cause of distortion at this point can be due to the way the magnetism distributes itself between the poles. Maintaining uniform transfer of all parts of the waveform from electrical to mechanical requires that the coil move through a space where the magnetism does not change, to interfere with the efficiency *at different parts of the waveform.* That's difficulty three.

Cornering

Now we come to difficulty four. We have the movement transformed from the electrical power and going back and forth in the cylindrical coil, we hope. But we want to move air with it, to make sound waves, so we have to make the coil drive a cone or other shaped diaphragm, at an angle approaching a right angle. The mechanical sound wave has to be transmitted around this almost-right-angle bend, where the cone joins onto the voice coil.



Difficulty five is the cone itself. It should move back and forth as an entity (or piston, as the experts say) so as to move the air solidly. If it starts to flop, one part will be moving forward while another part is moving backward, causing an erratic wave to be formed, out of proper control by the driving cone. The cone needs to be rigid, to avoid this breakup, as it is called, without being heavy, so it is difficult to move at all at higher frequencies.

The road ends

Difficulty six is what happens at the outside edge of the cone. If it is just left open, the air will rustle round the edge, losing low frequencies and causing spurious "wind" noises. So it needs a "surround." This is usually an airtight but flexible set of ringed

corrugations, not unlike those often also used for the modern "spider" or centering piece. But this ring must allow free movement of the cone at all frequencies, without interfering with the waveform with which it moves.

From there on we have an airborne wave, which we still have to get out into the room effectively and without distortion. This we will come to in following chapters. Meanwhile, let's explore a little more fully what's been done with the moving-coil mechanism itself.

The obstacle course

All that can be done about the first difficulty is to use materials and construction methods that make the coil precision round and center it accurately in the space between the poles. The degree of precision is determined by how much you want to spend to have it done (broadly speaking), so the efficiency at this point is fairly well related to cost.

The centering device needs careful manufacture to prevent its having a tendency to an "oil-can" effect or to "flip-flop" through the middle part of the coil travel. It is usually made of plastic-impregnated fabric, pressed into a series of circular corrugations, so the inside is rigidly centered by the outside mounting but can move back and forth freely. The quality of the plastic impregnation is important, to avoid causing distortion through harsh movement or erratic response to different frequencies.

The third difficulty is most often ignored to obtain best possible efficiency. But sometimes, when the coil is allowed quite a long travel — $1/8$ inch or more — special precautions are taken: a small coil can move in a big magnetic field, or a long layer of turns can move through a short magnetic field. Either way, one occupies only part of the other and, because of this, the part remains constant throughout the useful range of movement.

Either way also reduces the possible efficiency of the speaker drastically. If you want to retain good efficiency, it is best to design the rest of the speaker so less movement is needed at the coil. This we will discuss in the next chapter.

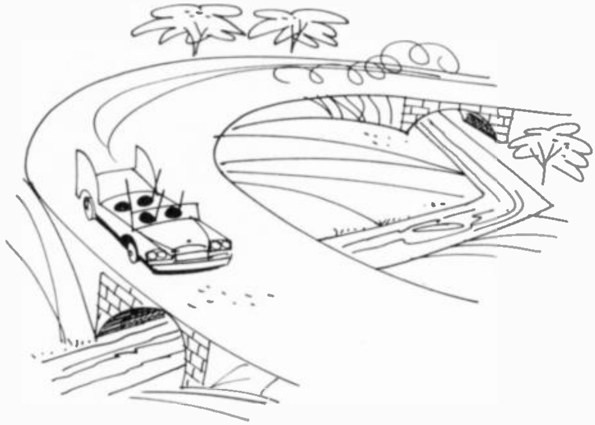
This is where a practical, rather than a perfectionist, attitude is best. By trying to eliminate one problem completely, you usually make the others harder.

Grading the curves

The fourth difficulty can be nicely circumvented by using a curved diaphragm instead of a cone. This makes for smooth transition of the wave from lengthwise movement in the coil to sideways movement near the outer edge of the diaphragm (often miscalled "cone," which it is not in this case).

But this construction is not so rigid in maintaining concentricity as the true cone, so it makes the first difficulty more severe. If you use this type of diaphragm, either you need a much more carefully made centering device or the precision dimensions of coil and magnet must be opened up, losing a certain amount of efficiency.

Difficulty five is also helped by the curved diaphragm instead of a cone. Although it is less rigid in maintaining concentricity, it is more rigid in maintaining its radial contour. So what you lose one way, you gain in another.



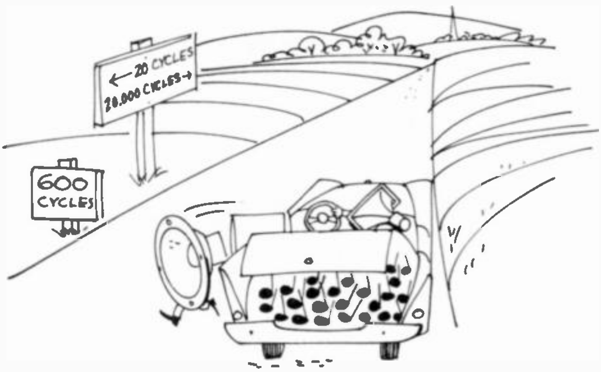
The sixth difficulty is quite often ignored, as a matter of convenience. But the better units use special plastic or other impregnation of the corrugations in the surround, or else a separate surround of different material cemented on, to maintain uniformity of response and freedom from harsh control, over the entire frequency range.

All in all, speaker design proves to be a line of compromises. The designer usually decides on a line of attack — which way he can best hold the performance to what he wants — and tries to design some of the problems out of his unit, while he takes pains to control the others within limits. A well designed and constructed speaker can be made in almost any combination of the variables we have discussed.

Is it easier to divide the run?

But, whichever way you tackle it, it's a tough task. Probably the biggest single problem is that of handling the full frequency range. To handle the low frequencies, which use big sound waves, the speaker has no alternative but to move a lot of air somehow or other. This means a big cone with small movement or a small cone with big movement — in most cases, at least.

To handle higher frequencies, this big cone must move rigidly, without breakup. Making it rigid tends to add weight. If it gets



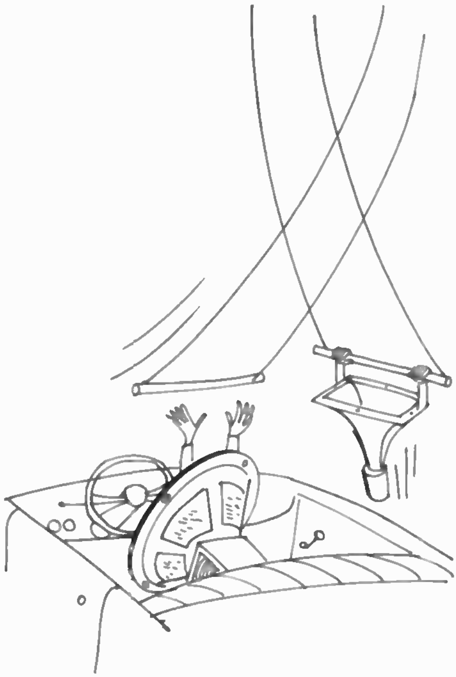
heavy, it will not handle the higher frequencies because it cannot travel back and forth quickly enough. So, when you've made a speaker big enough to handle the low frequencies, it's difficult to get the same unit to handle the very high frequencies.

This led to the use of separate units, woofers and tweeters, to handle low and high frequencies, so each could be designed for its own part of the overall job. As well as making the handling job easier, the use of separate units helps solve a distortion problem.

Harmonic distortion of low frequencies is relatively unimportant. They usually possess a number of overtones anyway, and a slight rearrangement is not noticeable. What causes noticeable trouble is the intermodulation. The low frequencies cause distortion to completely different notes in the middle and upper registers. Using separate units for each frequency range avoids distortion of this kind caused in the speaker, because they don't use the same unit.

A running changeover

But using separate units creates another problem: that of arranging for each unit to receive only the frequency range for which it is intended. A unit may be designed to handle only 600 cycles and up; but then we have to make sure it *gets* only 600 cycles and up, or our trouble will have been in vain. This is achieved by electrical circuits called crossovers.



The speaker designer decides on the frequency range each unit is to handle and then hands his problem to the filter designer, who makes the crossover. But there is no magic about these filters. We talk about a crossover frequency of 600 cycles, for example. But this does not mean, as it is sometimes explained, that *all* frequencies below 600 cycles go to one unit while *all* above that point go to the other unit.

In any crossover unit, at the crossover frequency both units get just half the total power. A little below that frequency, the low-frequency unit gets *most* of it but the high-frequency unit still gets some. The rate of transition from one unit to the other is a matter of crossover design.

How sudden?

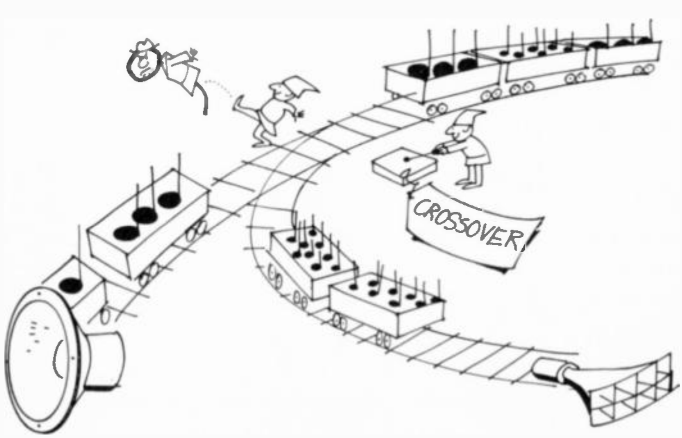
The simplest type of crossover, like any other, splits power equally, 50/50, at the crossover frequency. At one octave above or below the crossover frequency, the power ratio between units is 80/20 or 20/80. The next most complicated type, using just twice as many components, makes the 80/20 or 20/80 split at half an octave above or below crossover.

A crossover using three times as many components as the simplest type can make the 20/80 or 80/20 split at one-third of an octave above or below crossover. Then the most complicated crossovers ever made, using four times as many components as the simple ones and requiring quite critical adjustment, make the 20/80 or 80/20 split at a quarter of an octave above or below crossover. At an octave, this crossover makes the split 1.5/98.5, which is fairly complete.

So far, it looks as if the effectiveness of the crossover is merely a matter of how good a unit you can afford. But the crossovers that give "better" frequency separation do not improve performance in every way. They create other problems. As far as handling *frequencies* are concerned, our assumption would have been quite correct. But music is not a composition of continuous frequencies. We must not forget transients, remember?

What about transients?

The better crossovers, for separating frequencies, prove to be



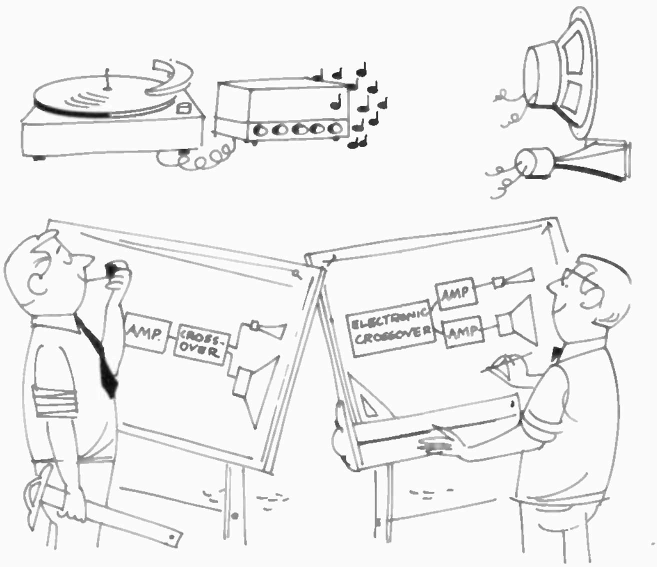
inferior at handling transients. There get to be quite serious time differences in the transient components of the music coming in the separate frequency ranges, when handled by the more complicated crossovers. So it's best to stay with the simpler types.

If speakers can be designed so their behavior, say for an octave beyond their nominal range, is still quite good, then the simplest crossover should do a good job. For speakers that do not have quite such a good "spread" beyond their nominal range of frequencies, it may be advisable to go to the next more complicated type, which gives a 16/1 separation an octave above or below crossover. This is as much separation as we would recommend, if you want to keep good transient response.

Where to divide

There has been quite a move recently toward using electronic crossovers or frequency separators. These work ahead of the basic or power amplifier. This means separate amplifiers are needed for each speaker unit, after the frequency separation point.

If the amplifier contributes any of the audible intermodulation



distortion, separating the offending frequencies before the distortion occurs will prevent it. But a good amplifier will produce no *audible* intermodulation distortion. It's true that many amplifiers with specifications that suggest their distortion *should* be inaudible do produce audible distortion. But this is due to some of the effects discussed toward the end of Chapter 4. If you get a really good amplifier, you needn't bother about this.

Some "reasons"

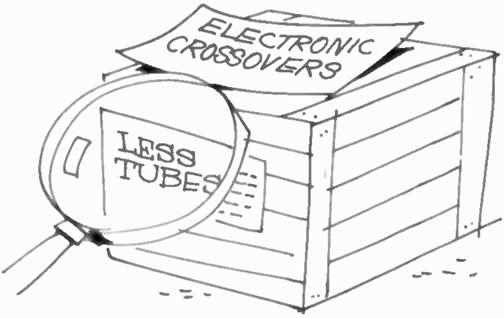
Other arguments put forward in favor of the electronic crossover concern possible distortion due to the crossover itself. Two factors are mentioned. The first is that a power crossover, in the speaker output connections, has to use coils as well as capacitors. The argument runs that, if these are plain coils of wire (air-cored), they are inefficient and cause losses; if they use a magnetic iron core, they are more efficient but cause distortion.

It is true that the magnetizing current of an iron core does include a distortion component. But with good design, the distortion this reflects into the main circuit can be as small as any other form of distortion. This argument has been grossly exaggerated. If people who objected to this were consistent in objecting to every little bit of distortion, audible or not, they would never have any reproduced sound at all!

The other objection is that using speaker units with crossovers can cause the *amplifier* to produce more distortion than feeding a simple speaker would. This, too, *can* happen. But it is also very easy to design an amplifier to avoid it – and certainly much cheaper than buying a whole extra amplifier, plus an electronic crossover.

What they don't tell you

On the other side, there is a much less often heard objection to most electronic crossovers. It is perfectly possible to design an electronic crossover that gives any desired separation. But most of them don't give what they claim. This is not due to willful deception, but basic ignorance of design principles beyond the scope of this book. Most crossovers are a compromise between very poor separation or quite severe loss of frequencies covering about an octave above and below the crossover frequency – neither of which is desirable. An electronic crossover with correct performance is rather more expensive than most of these compromise versions.



Whether to divide

Most American manufacturers in the high-fidelity field favor the use of at least a two-way system, whether you use the output circuit crossover or an electronic one. Many of them recommend three-way systems, with woofer, mid-range and tweeter units, on the basis that splitting the frequency range into three parts will better gain the advantages of splitting it than will just two.

But some English[1] and European manufacturers, prefer to use one good unit to cover the whole range. As soon as you divide the frequency range, whatever kind of crossover you use, there is an extra transient problem. The frequencies belonging to a single

sound come from different places, as well as having any time difference the crossover may cause. A single unit, well designed to handle all the frequency range, avoids this problem.

The next best thing, and one that produces quite superb results,

[1] Notably, Mr. H. A. Hartley.

is to use more than one unit but mount them concentrically or coaxially, so the sound at least seems to come all from the same place. Mounting the tweeter horn so it comes right through the middle of the woofer unit, to let the sounds merge together there, makes a very good composite or unitary speaker of the multi-unit type.

How good can a reproducer be?

We have quite a variety of ways to tackle making a complete reproducer. For a long while the objective has been the faithful re-creation of the original sound wave. But the original sound wave arrived at the microphone as a pattern, not just as a complex combination of frequencies but also with components arriving from different directions — which human hearing at the same location could appreciate. When the sound was transcribed into electrical currents, there was nothing to indicate individual directions so these can be re-created.

The best a speaker can do is to reproduce all the component frequencies in proportion and with minimum interference with transients. Directionality of radiation will have no relation to

the original. But it may be able to help or hinder realism. So a satisfactory compromise is needed.

Directionality can alter the effect of the radiated sound wave in much the same way that a directional microphone can alter the way the original is picked up. Stereo is the latest step forward that introduces a means of controlling directional effects properly or intelligently. This we shall discuss in Chapter 12.

The electrostatic makes its entrance

A recent entrant that challenges the well-established moving-coil speaker unit is the electric (more commonly called "electrostatic" but there is nothing static about reproducing sound waves). For some time, tweeters of this type have been available, but now wide-range ones, extending into the lower frequencies, have begun to appear. (Electrostatic speakers aren't *new*. They were used with American receivers over thirty years ago!)

They are encountering resistance, mostly because moving coils have held the field for more than three decades. They have troubles of their own too. The very high voltages used for polarizing and driving them are apt to cause electrical breakdowns. But it has been reported that this problem is solved.

From the reproducer point of view, they offer interesting possibilities. The basic construction is a large sheet or layer, with only enough framework to support it in shape adequately. It can be made flat or curved, as desired, and the whole surface radiates

sound as well as receiving the electrical "drive." Thus several of the problems associated with the moving coil are missing.

Thus there is a basic difference between the established moving-coil type and the incumbent electrics: the moving-coil cone is most often smaller than the waves it is called upon to radiate; the big diaphragm of the electric can reverse this situation. With the moving coil, the cone needs "outside help" at lower frequencies in moving enough air (discussed in the following chapter). Only at the highest frequencies are the waves comparable in size with the cone pushing them.

The electric can be made large without being made bulky.

This means it is possible to have it almost as big as the waves belonging to the lowest frequencies. Then the higher frequencies have waves small enough for the electric's diaphragm to accommodate many of them.

This means the problems encountered in the use of the electric principle, as well as in making it work, are quite different from the ones discussed in this chapter. Moving coils may not be easy to make, but at least designers know what the problems are!

While it is still a new thing in the wide-range reproducer field, new developments can be expected that will exploit this greater flexibility in design more fully.

Problems exist in matching it to an amplifier, because the electrostatic's impedance is even more awkward than that of the moving-coil type. But the new trend toward basic amplifiers integrated with the speaker they are built to drive may well solve this part quite simply.

We have certainly come a long way since the day of headphones on the table. But there is still a long way to go before we can think of making the word "fidelity" an absolute one instead of speaking about "high fidelity."

chapter 10

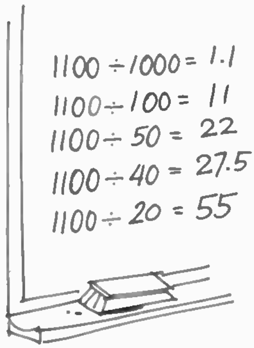
woofers

STRICTLY, the name woofer applies to a unit designed to handle the lower audio frequencies, with correspondingly big waves that need a lot of air to be moved, compared to the higher frequencies with much smaller waves. But in this chapter we are not concerned only with low-frequency *units*. Rather we will deal with how speakers are made to handle the low frequencies. What we

say here will equally include the features of an extended-range unit that are important specifically for the low frequencies.

Some arithmetic

Sound waves travel at approximately 1,100 feet per second in air. This means each individual wave must travel this distance in 1 second. At a frequency of 1,000 cycles, 1,000 complete waves will leave the starting point during 1 second. The first one will travel 1,100 feet, the last one will only just start. So the 1,100 feet traveled by the first one will contain 1,000 waves. This means each wave will measure 1,100 feet divided by 1,000, or about 13.2 inches.



Sound waves of 100 cycles travel at the same speed, so 1,100 feet will be occupied with only 100 waves of this frequency, and each wave will be 132 inches, or 11 feet long. Following this reasoning, we find that wavelength is *inversely* proportional to frequency, or the lower the frequency the bigger the wave. A 50-cycle wave will be 22 feet long. A 40-cycle one will be 27.5 feet long, and a 20-cycle one is 55 feet.

Big sounds

Musical instruments that produce these low frequencies are invariably big. The long, heavy strings on a piano produce the bass notes. They don't radiate them directly to the air, but rely on the whole sounding board of the piano vibrating to give out the sound.

The big double bass is quite a big instrument but, even so, it is quite ineffective unless stood on its peg, so the vibration is transmitted to the floor. A double bass played on a concrete floor is ineffective. It has to have a wooden floor that can vibrate with it.

The big bass drum is just that. The low notes of an organ use the long, big pipes. In this case, a large radiating surface is not used, but the holes from which sound emerges are quite large and the air flow quite considerable — enough to extinguish a match quickly.

We have said enough to show that bass notes require the movement of a lot of air in big waves. It is not surprising that acoustic engineers deduced that any speaker has to be big to produce bass. It either needs a big unit, or the unit must be placed in a big box. For a long while it seemed as if this was a fundamental principle that could not be avoided. There are still people who maintain that *good* bass can come only from a big unit. Some of them even restrict this possibility to one kind of enclosure.

Baffles

After it was found that the new moving-coil units were capable of reproducing much lower frequencies than the earlier moving-iron types, it was noticed that they wasted the very low frequencies — below about 200 cycles — by vibrating at large amplitudes but producing very little sound. This was because the air they moved shuffled round the edges without getting pushed out to form a wave.

The remedy for this was to stop the air shuffling around the edges, by mounting it in a baffle board. This really got some bass from the new moving-coil units. The bigger the baffle board, the lower the bass would go.

But once the novelty of hearing those lovely bass notes so beautifully had worn off, we realized that the speaker mounted in a baffle board was no longer doing justice to all the other frequencies. It did not give *balanced* reproduction. If you used a smaller baffle board, the bass did not go so low, so there was no way of getting it to give really wide range, balanced response.

Doing it properly

Putting a baffle board on a speaker unit was just a way of improvising with what we had. Others felt this was not the way to do it. What you had to do was investigate and determine the *need*. There is no doubt that, among musical instruments, the type that gives the most sound for the least effort input (excluding anything like the organ that needs a mechanical blower) is the horn or trumpet.

This is because the sound wave is gradually expanded from a high-pressure vibration, that can be made in a confined space, to a big, low pressure vibration, suitable for continuing in space as a

sound wave. The baffle board helps control the air movement at the lowest frequencies, but at frequencies where the wavelength is comparable to the speaker's own size, the baffle does not do anything.

The cone or diaphragm is always much heavier than the air it has to drive, so it wastes its energy moving itself instead of the air. But by enclosing its front and allowing air to escape through the throat or narrow end of a large horn, it can use much more of its energy working on the air. The wave gets expanded by the horn into a large low-pressure wave that has more sound power than the same speaker unit could produce without it. This improves the efficiency of the speaker, not only at the lowest frequencies, but over a whole range of frequencies.

A matter of size

But how big does a horn have to be? This depends on the lowest frequency you want it to deliver. Its mouth has to be big enough for it to deliver a fully developed wave of the lowest frequency. The mouth dimensions need to be at last half a wavelength each way. So a horn to reproduce down to 50 cycles needs a mouth 11 feet across each way! If you will be satisfied with 100 cycles, the mouth needs to be only 5½ feet each way.

This is not all. The rate at which the wave can be expanded depends on the wavelength too. It has been found that a workable rate of expansion allows the area of the horn to double every 1/18 wavelength. Let's see what this means.

Assume we use a speaker with an 8-inch diameter cone. This would need a horn throat about 5 inches in diameter. Doubling this means the area is quadrupled. So, starting from the 5-inch diameter, successive doubling of area will give diameters of 7, 10, 14, 20, 28, 40, 56, 80, 112 and 160 inches.

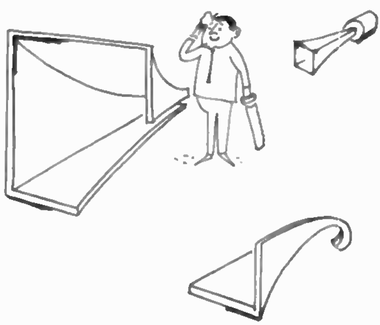
If we want 50 cycles, the mouth diameter needs to be between the last two figures, 132 inches (11 feet), which represents 9½ times the area that has to be doubled. Each doubling of area needs a length of 1/18 of 22 feet, or about 15 inches. So the horn has to be 9½ times 15 inches, or about 12 feet long, as well as having a mouth 11 feet in diameter.

If we accept 100 cycles as the limit, the mouth needs to be only 5½ feet, or 66 inches, requiring only 7½ area doublings. And the doubling can be twice as fast, requiring only 7½ inches. So the length needs to 7½ times 7½ inches, or about 4¾ feet.

The cost of getting lower frequencies is consequently a very considerable increase in size. On the same basis, a horn to go down to that favorite figure of 20 cycles would need a mouth 27½ feet across and have a length of almost 40 feet. Where in your living room would you put it?

A piece will do

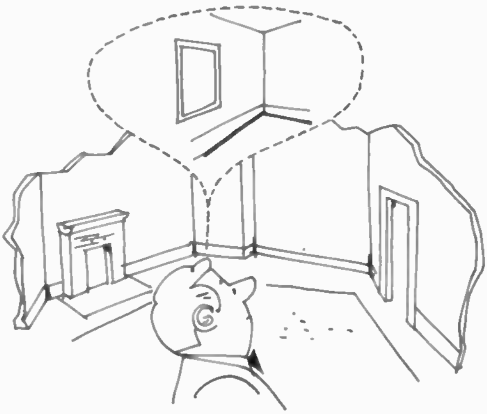
Actually these dimensions can be improved somewhat by some tricks. If you cut a horn in half lengthways and place it against a wall that continues beyond the mouth, it works just the same as the whole horn, because the wall produces a mirror image of the real half. If you put it in a corner where two walls meet, a quarter will do. Finally, if you arrange so that the emerging sound coincides with a three-way corner, corresponding with two walls and the floor, one-eighth of the mouth area will do.



With this adaptation into the form of a corner horn, a mouth area of about 12 square feet will do for a 50-cycle horn. The length too reduces by the equivalent of three area doublings, or about 3¾ feet, to 9¼ feet long. By clever design, the narrower section of the horn, nearer the throat, can be folded behind the frontal area so the whole thing can sit in the corner of the room. This is the way a corner horn is designed.

Wanted — a perfect corner

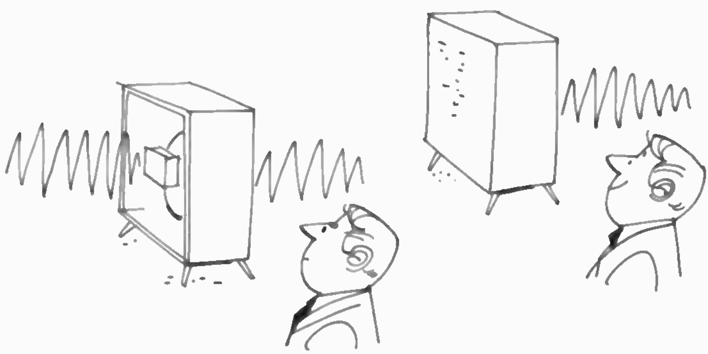
Corner horns are complicated structures, and they are very dependent on being put in a corner to get the right results. Not only this, but the corner should be a good one. A corner that has a door



or entry a few feet away is no longer a good corner from a horn's viewpoint. All these things being considered, although a corner horn is a wonderful way to get a good low-frequency speaker, you may need your house rebuilt to be able to use it. At least you must decorate to suit, which many people do not always want to do.

Boxed-in back

This is why other types of speakers were invented. Starting from the same idea that suggested the simple baffle board for stopping

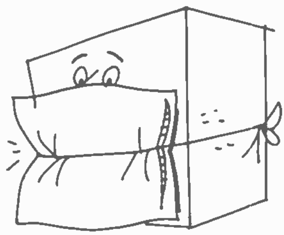


the air shuffling round the edges, one way of avoiding the need for such a large object is to box in the back. This also finished the thing off so it became an acceptable piece of furniture, which you could hardly say a moving-coil unit is.

Boxing in the back prevents any sound wave from radiating from the back at all. Only the front of the unit has access to the room to radiate sound. But the back of the cone has to move just as much as its front. Confining the back in a small box, while the front feeds into a large room, means there will be a much greater sound pressure inside the box than is radiated into the room.

Air-cushion muffling

In short, boxing in the back reduces the efficiency. The smaller the box, the less efficient the speaker. This particularly applies to the low frequencies. If you push a piston in and out of a closed cylinder, the enclosed air acts like a spring or compliance. This is exactly how the air in the box reacts to the movement of the speaker diaphragm.



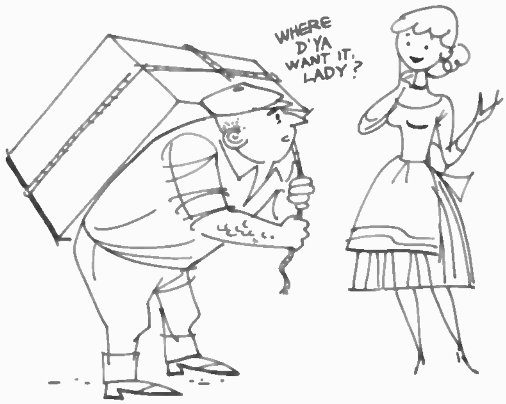
Normally, a speaker's main resonance is similar to a weight supported by a spring. The weight is that of the diaphragm and all the parts attached to it and including the air in contact with it that has to move with it. The spring is the combined "stiffness" of the spider and suspension, tending to restore the diaphragm to its starting point. At some low frequency, usually somewhere in the range from 35 to 150 cycles, the spring-and-weight resonance effect allows the diaphragm to reach a maximum movement, more than other frequencies.

At frequencies below this resonance, the combined stiffness of the parts tends to restrict movement while at frequencies above it the weight (or mass, as it is called) hinders the movement.

When the back of the speaker is enclosed, the air inside acts as an additional spring, increasing the total stiffness. This raises the resonant frequency and also increases the restriction to movement at frequencies below resonance. For this reason, making a smaller box results in progressively greater, and quite drastic, loss of the low frequencies.

A big enough box

Usually the so-called *infinite-baffle* type consists of a totally enclosed back unit in which the resonance unmounted would be, say, 35 cycles. Then a box size would be chosen which would add not more than a stiffness equal to the mechanical amount already in the unit. If the added stiffness is the *same* as that already there, making the total just double, the resonant frequency will go up to 50 cycles. Below this frequency, the response will fall off fairly rapidly.



Although such an infinite-baffle speaker is not critical about placement to the extent the corner horn is, it is still a pretty big piece of furniture, occupying about the same amount of room. Also it is relatively inefficient, because the major part of the low-frequency energy has to be spent inside the box. So a more efficient means was sought.

What bothered some designers was the fact that, in the infinite baffle, the radiation from the back of the diaphragm is suppressed. If only it could be somehow reversed in movement to aid the radiation already coming from the front, instead of shuffling air around to almost cancel it!

Bass reflex

That's exactly what the bass-reflex design did. One way would have been to "pipe" the back radiation on a journey equivalent to a half-wavelength, so as to reverse it. This is the approach used in the acoustic-labyrinth speaker. For a 50-cycle bottom limit, the labyrinth has to be 11 feet long, folded inside the cabinet. But the bass-reflex method offers an even simpler solution.

If you have two equal weights on opposite ends of a spring, there is a resonant frequency at which the two weights move in opposite directions at every instant, aiding one another in alternately stretching and compressing the spring. This is just what the bass reflex does. As with the infinite baffle, the air inside is the spring. The diaphragm is one weight, while the air in another hole on the front of the box is the other.

When both work together to compress and expand the air in the box, which is what happens at resonance, the air in the opening moves forward at the same time the speaker diaphragm does. This enables the desired reversal to be achieved without a pipe 11 feet long. It's quite a simple construction and has been widely used because it gets good bass from a smaller piece of furniture than the other types. It has the added advantage of being more efficient and using less cone movement than the corresponding infinite baffle, although it may not be quite as good as a properly placed corner horn.

Back-loaded horn

A compromise between the horn and the bass reflex and having some of the advantages of both is the back-loaded horn. In the normal folded horn, the medium frequencies get "mixed up" making the turns at the folds, so only the lowest frequencies get the benefit of "proper horn design."

So the unit is mounted on a front panel to radiate middle frequencies directly without obstruction. Then the back goes through a small space, just enough to absorb all but the lowest frequencies, and feeds a folded horn whose mouth finishes up in front for the low frequencies. This kind beats the bass reflex for efficiency, but still needs a corner placement.

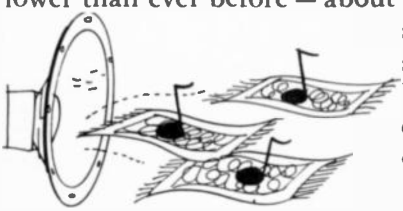
But still there was an insistent demand for *smaller* speakers. Not everyone has room to spare for the large units needed to handle the

lowest frequencies. Cannot smaller speakers be made and still retain the low frequencies?

Acoustic suspension

The first answer to this was the *acoustic-suspension* principle. With the infinite-baffle type, to get a low resonance to start with, you need a big diaphragm. Then, to avoid raising the resonance too much by the box, you need quite a large box. The acoustic-suspension system starts with a diaphragm somewhat smaller than the average infinite-baffle type and then uses an extremely "floppy" suspension.

In this way, the resonance unmounted comes out very much lower than ever before — about 5 or 10 cycles. This allows a much smaller box to be used, with a stiffness that brings the resonance back to about 40 cycles. This has certain advantages compared to other types.



Sometimes — in fact, usually — the spider and surround contribute some distortion because their stiffness does not act uniformly. But making this very floppy, so there is hardly any stiffness, means the air in the box — a pneumatic spring — forms the major stiffness, and this is quite uniform and linear. So we get less distortion than, say, an infinite baffle or bass reflex.

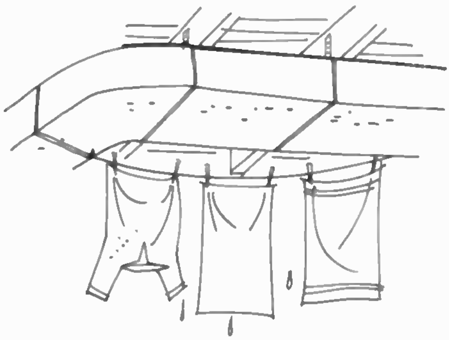
So we have got the size and distortion down too. What have we traded for these advantages? Efficiency. And the coil has to travel much farther for corresponding sound output. The floppy unit has to have much bigger clearances in the "motor" to avoid possible rubbing, and the magnetic field needs a long coil to avoid drive distortion. So the acoustic suspension unit needs about 50 watts to give as much sound output as another unit might on 5 to 10 watts.

Still the acoustic suspension unit has the same basic disadvantage, compared to the bass reflex — the radiation from the back has to be bottled up. Cannot something similar be done to improve the bass reflex? It can and has.

Duct loading

Before the advent of acoustic suspension, a version of duct-loaded reflex had appeared. Instead of having a simple hole or port in the front, this was extended to form a duct of movable air that communicated the pressure inside to the air outside. This allowed the radiation from the extra opening to be more nearly equal to that from the front of the speaker itself at the lowest frequency. This raised the efficiency of the bass reflex to its maximum.

The newer duct-loaded reflexes borrow an idea from acoustic suspension and take duct loading a stage further. Instead of making the radiation about equal from both places, the new ones make most of it come from the duct by using a much longer duct or tube, so the column of air is "heavier." In this way, the speaker diaphragm moves only a little while the air in the duct moves quite a lot at the lowest frequency.



In making this change, the unmounted resonance does not need to be shifted down as far as it was with the acoustic-suspension type. The stiffness presented by the air to the back of the diaphragm is still more than that due to the speaker's own suspension. But this extra stiffness is not wasted. The duct-loading arrangement converts it into a mass movement of air in the duct, so it provides useful radiation — in fact, the biggest part at these frequencies.

Big bass from small speakers can get lost

All these new "small bass" speakers have one problem not encountered to the same extent with the earlier infinite-baffle and bass-reflex types. The low frequencies still have the wavelength assigned them by nature. If this is being radiated in *any* way from a small box, the effect is going to be somewhat dependent on where the box is placed.

Stood on a small table somewhere in the room, the sound wave has to fill the entire space surrounding the box. Stand the table back against a wall, and it has to fill only the space in front. Stand the box on the floor against the wall, and the angle it has to fill is even smaller. Each

of these steps enables the small box to build up a better low-frequency pressure in the space it has to fill.

The duct-loaded variety has one advantage here. The speaker can be built on legs integral with its design, and the duct can face the floor. Now the position of the major low-frequency radiation is controlled by the furniture design, much more than by where you put it in the room. After the radiation has filled the controlled space between the box and the floor, the sound can spread out into the rest of the room.

Oompah, oompah

So far we have not mentioned room size. But this too is important, for all types of speakers, at the lowest frequencies. If your room is a 14-foot cube, which is a pretty average cubic content for a living room, it will just contain a half-wavelength of 40 cycles in each direction. With doors and windows closed, the job *any* speaker has to do at this frequency is to pump air alternately in and out of the room!

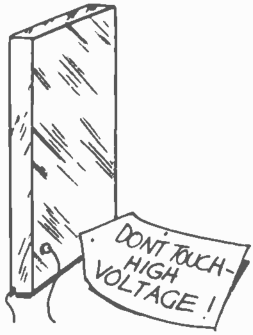
Open the door and you change the story completely. This is true even for the corner horn. Actually, the fact that the rooms in

which we listen are usually of limited size is a saving grace for any speaker we use. It helps down in the frequency region where the speaker is not big enough, according to theory.

Electrostatic speakers

Recently, some full-range electrostatic loudspeakers (see page 169) have appeared. By their nature, the diaphragm cannot move very far, as the moving-coil can, but also they require practically no depth. They can be made like a picture to hang on the wall, or in almost any other suitable shape. One makes quite an ornamental imitation fire screen. (We don't advocate using it to double as a real fire screen.)

Where the moving-coil woofer unit is still much smaller in dimensions than many of the waves it has to radiate, and so has to use *some* kind of baffle or enclosure to achieve effective radiation, the electrostatic can be made large, compared to most of the wavelengths at least. This changes the problems encountered.



Everything in acoustics is relative. For example, consider using a sealed box of fixed dimensions with different size moving-coil units. A small piston, working in a small hole in the box, will find the air offers less stiffness to its movement than a larger piston working in a larger hole. The effects are very strongly related to area. So when you make the diaphragm really big, as in the electrostatic unit, it becomes almost impossible to think of enclosing the back.

On the other hand, the large diaphragm usually requires quite a stiff piece of insulating material to support it, so the air stiffness does not represent so very much more. But this can depend on design too. Some of the newer plastic materials for insulating can give quite good resilience, which again alters the picture.

Shaped to suit you

Shaping of the diaphragm, along with the fixed "plates" to which the voltages are applied, can be important for the higher frequencies (which includes mid-range when the unit is this big) because a large *flat* sheet tends to beam all frequencies for which its size represents more than one wavelength. Using a convex curved surface will help spread the beam over a controlled area.

Of course, too, the fact that the radiating surface is so different from the moving-coil type, to which most people have become accustomed, means the reproduction is bound to sound different, even if frequency response, distortion and dynamic range are identical.

Electrical matching to the amplifier is another problem. The moving-coil unit has an impedance that deviates considerably from its nominal or rated value. But, taken over the frequency range, this deviation stays within certain limits, that again have become the norm.

A matter of impedance

The electrostatic unit has a much different impedance. As it is

really a capacitor, it would be more informative to give its capacitance value than to rate it with an impedance. This is equally true, whether or not it comes with a matching transformer. A capacitance possesses a reactance or impedance inversely proportional to frequency.

If the impedance is 6,000 ohms at 1,000 cycles, it will be 12,000 ohms at 500 cycles, 24,000 ohms at 250, 60,000 ohms at 100, 120,000 ohms at 50 and so on, going downward. Going the other way, it will be 3,000 ohms at 2,000 cycles, 1,500 ohms at 4,000, 1,000 ohms at 6,000, 600 ohms at 10,000 and so on. Assuming a unit may handle frequencies from 50 to 10,000 cycles, its impedance will change between 120,000 and 600 ohms over that range.

It is impossible to connect it through a transformer that makes

it "look like," say, 16 ohms. If it is 16 ohms at 10,000 cycles, it will still be 3,200 ohms at 50 cycles.

The best way to handle this kind of situation is to design the electronics to go with it, because no ordinary amplifier is intended to feed this kind of impedance. As a number of moving-coil speakers have already been designed with their own integral power amplifier, this approach, a sort of integrated package idea, is consistent with the newest developments.

This trend got started because the performances of amplifier and speaker cannot ever be completely isolated one from the other. This is most true in the low-frequency range. Whatever is done in the way of enclosing the speaker in a box, with various features, vents, acoustic damping, etc., alters the response of the speaker. But in so doing it also changes the electrical impedance into which the amplifier has to feed its power.

This means that changes in speaker design, or the way its enclosure is made, can affect amplifier performance: frequency response, distortion and the way it handles transients. Because of this, many amplifiers produce different quality according to what speaker they are connected to – quite apart from the difference in the way the speakers sound "normally."

Designers have worked hard to make amplifiers as independent as possible, so their quality will be uniformly good, whatever speaker they are used with. But it's a tough job, and one is seldom sure that some unexpected arrangement will not upset the amplifier and spoil its performance.

This led to the idea of building the two to go together. When an amplifier is constructed this way, the designer does not have to worry about how some unknown speaker may upset its performance. He designs it to work well with the speaker he puts it with and that's an end to the matter.

Although none is commercially available at the time this is written, several people have worked on the idea of using feedback to reduce the distortion in the speaker as well as in the amplifier. This is an extension of the trend toward designing the two together. One way this has been tried, quite successfully, is to use a speaker with two coils.

One of these coils is used in the regular way to drive the cone. Then the other picks up a voltage dependent on how much the cone actually *moves*. So when the cone resists movements at one frequency more than others, the feedback is reduced and more drive produced so the movement is equalized. When it tends to move too much, as at resonance, this produces more feedback that holds the drive back.

With proper care in the design, this method can also reduce distortion. To achieve this, the waveshape produced by the feedback coil must be a faithful replica of the movement of the cone. This means more care is necessary in placing the feedback coil so its electrical pickup is accurate than in placing the drive coil so its drive is uniform. The nonuniformity in drive can be corrected by the feedback. But if the feedback is wrong it will make a false correction, causing the drive to be worse instead of better.

There are problems in making this system practical. It is not so simple as making each work by itself, but once it is in operation, the results are well worth the effort.

So, although the advent of the moving-coil speaker several decades ago represented the initiation of good response at the low bass frequencies, the present decade is seeing tremendous advances by way of perfecting reproduction at this end of the sound spectrum. In spite of so-called theory that says good bass has to come from big speakers, some very good small ones have appeared.

Work is going forward all the time, in this as well as other areas of high fidelity, to bring what can be done a stage nearer to the perfection it will never completely achieve.

chapter 11

tweeters

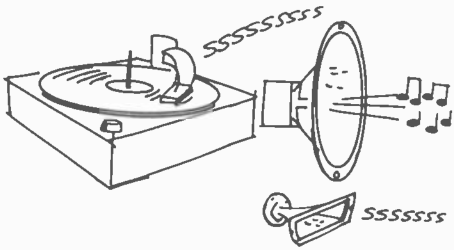
THERE is a strong feeling in some quarters that tweeters got their start in high fidelity through the ad men's need for sensationalism. The previous innovation in speakers was the mov-

ing coil that permitted better bass to be heard. So what was more logical than to extend the other end? There was also a more substantial foundation. Measurements had shown the speakers *were* seriously lacking in frequencies much above 5,000 or 6,000 cycles.

The combination of these circumstances led to an introduction of tweeters not exactly congruous with true fidelity objectives. Many who jumped into the tweeter manufacturing market just made units that "responded" to frequencies above 5,000 or 6,000 cycles—somehow. They made an audible difference to reproduction. Mainly you could hear a lot more hiss from tubes or needle scratch. This you were missing if you didn't have a tweeter!

Hearing a tweeter

But the more serious people worked a little harder on making units that gave a smooth frequency response in this region. Unfortunately, reaction to them at the time was even more disappointing: you could hear less hiss than the cheaper variety gave, and still practically nothing was added to audible *program*. There were two reasons for this.



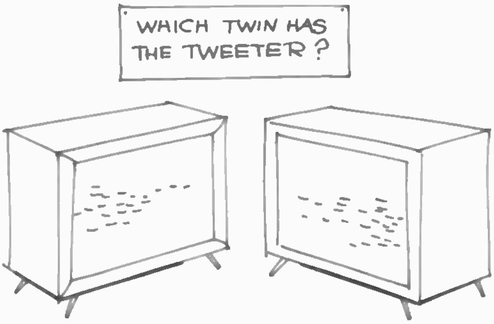
Most important was the fact that little program material of the time *had* anything up there to hear, although there should have been. Much of the realism attached to original sounds that got lost in reproduction was due to loss of these higher frequencies. But up to that time, broadcast had used only AM, with little above 5,000 cycles, and records were no better.

Second, even if special demonstration material was obtained, with the higher frequencies present, the difference they made was much less obvious than had been expected. When the latest units had made bass more audible, the difference was quite astounding. Remove the bass, and the program went quite tinny. Put it in, and all the bass instruments suddenly appeared. They had not been there before. But the frequencies added by a good tweeter take much more listening for to *hear the difference.*

Telling the difference

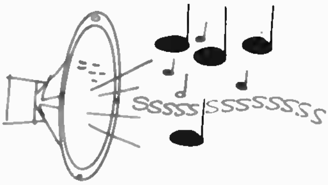
Subconsciously, we were listening for sounds that had not been there, based on our experience with bass. But if you have a frequency response up to 5,000 or 6,000 cycles, you *can* hear all the sounds in practically any program material. Adding response above this point does not make any *new* sounds audible.

It requires much more critical listening to *tell* the difference, although it may have quite a profound effect on our subconscious. It can make the difference between our thinking the reproduction is the real thing, and being able to tell the difference between original and reproduced sound. But it is difficult to *identify* what this difference is. This is because these higher frequencies do not come as separate sounds on their own, but as relatively minor parts of composite sounds whose major parts occupy lower frequencies.

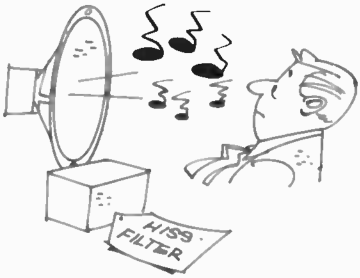


Believing your ears

Putting in an electronic filter that cuts off frequencies above, say, 6,000 cycles often makes the program sound distorted. This is not what the engineers expected. Taking some frequencies out should make it sound as if something were missing but should not make the lower frequencies left behind sound distorted. So they did not believe their ears.



Of course, the effect *is* an illusion, and it *is* produced by the hearing faculty. But it's due to loss of those high frequencies from the sounds we *expect* to hear. A quite similar effect can be produced by using a tweeter that overemphasizes some of the higher frequencies by resonance effects.



When there is really good dynamic range, such as on an FM transmission or one of the better LP recordings, a high-quality tweeter (one that adds higher frequencies with uniform response) makes the whole program sound clear, distinct and undistorted (provided all other forms of distortion are low).

But in those early days when almost any program source had quite a high background noise, it was almost impossible to hear this change. The background noise prevented the hearing faculty

from picking up these components and utilizing them. In fact, the only way of knowing whether the high frequencies were there was by the amount and coloration of the hiss.

Cone tweeter

Three popular approaches have emerged as tweeter designs. The simplest and easiest, and first in the field, was the small cone direct radiator, just a scaled-down version of the regular moving-coil speaker used for lower frequencies. Some people advocate these as the best.

Their argument is that you should use all speaker units of the same type. If you use a horn for low frequencies, then use one for the highs too. But, as most people use a direct radiator in some kind of enclosure for the lower and middle frequencies, they argue they should also use a cone type tweeter.

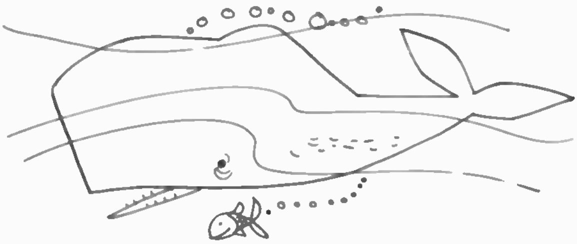
How many pieces — what size?

Let's see what this argument means. Suppose you use a 12-inch unit to handle from 40 cycles up. First let's take a two-way system and assume we'll be satisfied with 16,000 cycles tops. Then, to split the range equally, the 12-inch unit should handle from 40 to 800 cycles, while the other unit will take from 800 to 16,000 cycles. Each takes a frequency range of 20 to 1. But this means, to keep things proportional, that the high-frequency unit should be scaled down by the same ratio. We need a ⅜-inch tweeter unit!

Suppose instead we try three-way, and make each span a frequency ratio of 8 to 1. Convenient figures might be 37.5 to 300, 300 to 2,400 and 2,400 to 19,000 cycles. If we pick an 8-inch unit for the middle range, to be consistent we need a 64-inch woofer and a 1-inch tweeter! If we come down to a 15-inch unit for the woofer, we finish up with a 2 inches for the mid-range and a ¼ inch for the tweeter. These don't sound like practical sizes.

A different breed of fish

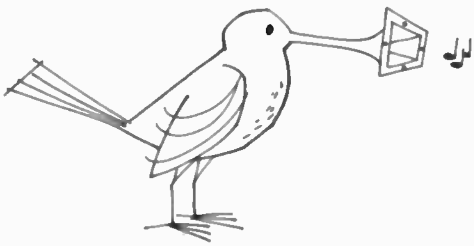
What this little calculation shows is that we cannot keep to the same kind of radiation over such a very wide frequency range.



At the low frequencies, whatever we do, we have the problem of making a diaphragm radiate wavelengths very much bigger than its own dimensions. Correspondingly, at the high end the diaphragm has to radiate wavelengths that are inevitably a fraction of its own dimensions. This fact alone calls for different techniques.

At the low frequencies, we have to resort to methods of making what seems like a small diaphragm move a lot of air in big waves. At the high frequencies. the problem is to make what seems like a big speaker radiate very small waves uniformly. Very different problems.

This does not mean that direct-radiator tweeters are not feasible. Some good ones have been made, and it is an inexpensive way of making a tweeter. But the design factors in making a good one are quite different from those for getting good performance from a woofer. So it will *not* be just a scaled-down version of the same thing. This, of course, destroys the argument that it is the best type *because* you are using the same kind of radiator at all frequencies.



A tweety horn

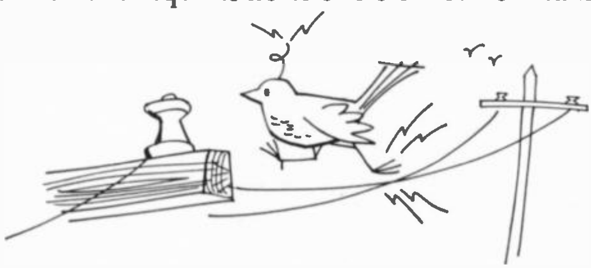
Those figures we came up with a few paragraphs back are really not as crazy as they seem at first sight. A good horn type tweeter has a throat not much more than 3/8 inch in diameter, and the diaphragm that develops the sound pressure to feed it will have a working diameter in the region of 1 inch. And for very much the same reasons that led us to these figures. This avoids breakup at the highest frequencies handled and insures a smooth response.

The horn is needed because such a small diaphragm or aperture could not hope to radiate, say 800 cycles where the wavelength is about 16 inches. The mouth should be at least 8 inches across. In a three-way system, for the tweeter to go from 2,400 cycles up would need a mouth about 3 inches across.

The horn does make a fairly simple method of covering almost any specific range in the upper region. The throat end is designed to handle the highest frequency, and the flare rate and mouth are designed to handle the lowest. Then the response in between is assured, because proper acoustic matching to the diaphragm follows automatically.

The electric bird

The third kind of tweeter is commonly called the electrostatic. We prefer the term electric. The essential part is a flexible sheet or diaphragm resiliently spaced and insulated from a fixed solid plate. The fixed one requires holes or slots — some means of allow-

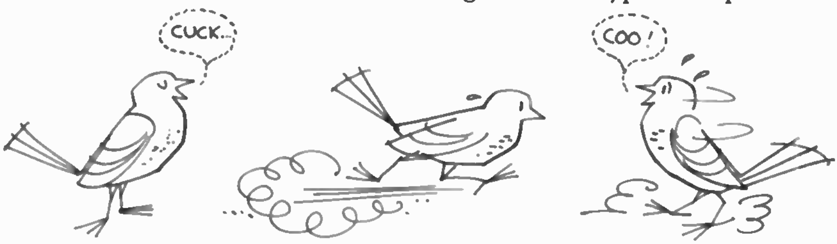


ing the air to move in and out, allowing the flexible diaphragm to move. The simplest uses one fixed plate. The movable one is fed with a combination of a high dc polarizing voltage and relatively high audio voltages (but always less than the polarizing voltage).

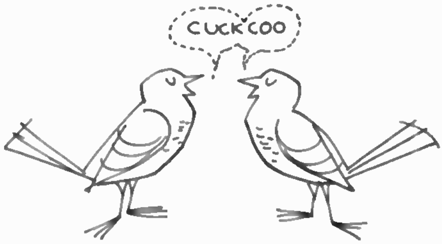
A claimed disadvantage of this *single-ended* type is that it produces inevitable second-harmonic distortion. For this reason, most people prefer the push-pull variety which uses two fixed plates with the movable diaphragm between them. The sound has to get out through holes in both of the fixed plates. A push-pull audio voltage is fed to the fixed plates, while the high dc polarizing voltage is usually connected to the movable diaphragm.

Single-ended or push-pull

The push-pull operation eliminates the second-harmonic distortion — or it *can*, with correct design. This elimination depends on completely balanced construction, both mechanically (or acoustically) and electrically. Failure to achieve this can result in as much second harmonic as a single-ended type can produce,



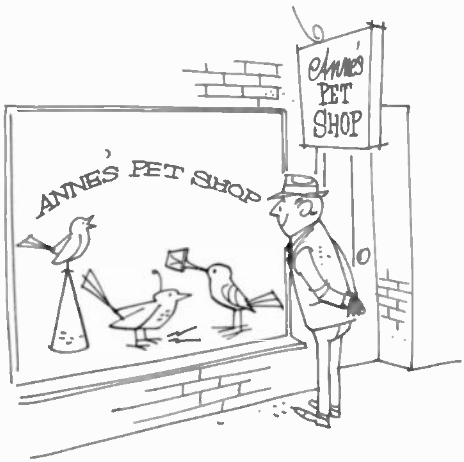
because the control arrangement is not so well able to offset it:



But it is not impossible to make a single-ended electric unit virtually free from distortion. The force corresponding to the applied voltage follows a relationship which causes a second harmonic in the *driving force*. But the diaphragm has to move both air and the supporting resilience. If this resilience is designed so it too *needs* a distorted driving force to get undistorted movement, the resulting movement can correspond quite closely to the *electrical* input that caused the distorted force. The two effects cancel. And that is not so difficult as it might at first seem.

Take your pick

So now we come to the usual question, "which is the best?" To



listen to any of the adherents of a particular type, one would conclude that their favorite is virtually the only one that works. The direct-radiator group has a persuasive way of telling you that your entire radiating surface, for all frequencies, should be of the same material. The horn group says theirs is the only one capable of precision design to fit the requirements. And the electric (electrostatic) group says theirs is the only one that avoids three-stage, electro-mechanical-acoustic coupling. It couples the electric force directly to the same diaphragm that produces the acoustic output.

Is the speaker a musical instrument?

On the first one, the analogy is often drawn to a musical instrument, a violin, trumpet, organ, piano, etc. According to the argu-

ment, each instrument radiates all its frequencies in the same manner or from the same surface. Correct. Therefore, a speaker should radiate all its frequencies from the same surface or, next best, from the same *kind* of surface.

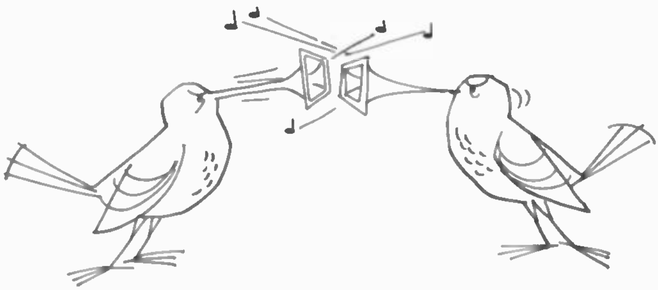
Is this argument intended to suggest that this will enable the speaker to simulate the original instrument more closely? If so, it should surely be argued that the speaker should have the same kind of radiating surface *as the original instrument*. It certainly is true that a speaker with a metal diaphragm seems to reproduce brass music better than other types, while one with a phenolic diaphragm closely resembling wood does a good job with woodwinds.

But this would seem to argue for an impractical proposition: how do you have a speaker that uses different radiating surfaces, wood, metal, gut, etc., according to the instrument it happens to be reproducing at the moment?

If there is anything to this, it would seem that it might be more practical to give some attention to *how* the surfaces radiate, rather than what they are made of. But more of that in a moment.

Counter horns

People who object to the horn, usually because they have another preference, complain that a horn type unit always makes



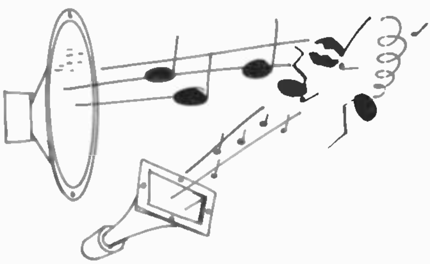
the sound "horny." If the horn is made of tin, the sound is "tinny." It is true that the material of which the horn is made *can* have resonances of its own that lend a character, which may sometimes be unwanted, to the reproduction. But it is possible nowadays to use a casting, either in metal or plastic, that is quite free from any such characteristics.

The peculiar "edgy" quality noticed with some horn type units is due to poor throat design that causes resonances in the upper part of its response. A well-designed precision-made horn tweeter produces a smoothness in response that is second to none.

Quality of the electric unit, too, is very much subject to design. Because of the almost uniform electric drive and acoustic takeoff, it is possible for the relatively large surface to produce a uniformity in response not to be found in moving-coil driven-cone radiators of comparable size.

Tweeter contribution to distortion

While any of the tweeter types can cause distortion, this is much easier to design out of any of them than in the case of lower-frequency units. What is more difficult is obtaining a uniform frequency response and also uniform distribution of dif-



ferent frequencies in space, which can be as important to good overall sound at these frequencies as distortion can at lower frequencies.

In short, then, we have a slight preference for the horn tweeter, but any one of them is only as good as the particular model is designed and made. None is *inherently* better than other types.

Directionality

Directionality of radiation begins to become important above

about 500 cycles. At 40 cycles, the wavelength is so big that it amounts to alternate inflating and deflating of the room air pressure. At 120 cycles, common room dimensions hold an almost exact number of half wavelengths (multiples of $4\frac{1}{2}$ feet), which explains why hum at this frequency can bother you at one spot and be inaudible a few feet away. But, above 500 cycles, the waves are small enough to get reflected around the room like complete waves.

This means directional properties about a speaker's radiation can contribute some effect to the reproduction up here. Some reverberation, caused by reflection of sound waves, is needed in any room to make the sound seem natural. A padded cell or an acoustic "anechoic" room where all the walls absorb sound almost 100% has too little reflection to be natural. An indoor tiled swimming pool has too much.

To suit your listening room

Most rooms used for high-fidelity listening will come well within these extremes, but there will be quite a difference in individual cases. In a room that has fairly absorbent walls, floor and ceiling, some reverberation can be forced by using speakers that deliberately beam the higher frequencies to some extent. This way the sound has to bounce around somewhat to fill the room.

On the other hand, a typical American recreation room with smooth walls, plenty of glass windows, hard floor and ceiling surfaces and nonabsorbent furniture, is quite a bit too reverberant for normal high-fidelity methods. This is where some careful dispersion can help. An acoustic lens, or some design that distributes the sound uniformly in horizontal directions but avoids the floor and ceiling, can do wonders to reduce the confusion effect.

Each of the kinds of tweeter is adaptable to different directional treatment. The direct-radiator types vary according to cone angle and can further be controlled by putting deflector or reflector surfaces in front of them. To restrict radiation further toward floor and ceiling, a number arranged in a vertical line, carefully connected so they all "push and pull" together, will get very good distribution for quite low cost.

The horn unit can be made directional or otherwise at will, by design of the expansion in vertical and horizontal directions. Alternatively, an acoustic lens can be used in front of either of these types to get the radiation you want.

The electric type tends to beam because of its large surface compared to wavelength. But this beaming can be readily controlled by shaping the surface. A flat surface beams in a relatively narrow angle. A curved surface, or a number of flats arranged like a curve, will distribute in the direction of curvature, leaving beaming in the vertical direction.

Mid-range

Most of this chapter has been devoted to tweeters. But we have not left a chapter for mid-range units. So we will squeeze them in

here. Actually, all that needs to be said about mid-range has been covered (maybe in different proportions) in the last chapter and this.

There was a trend, in going after response to the bass and, later, the upper treble (with tweeters and super tweeters), to overlook the old standby mid-range frequencies, the ones we have been listening to all along. Any old unit can reproduce those, seemed to be the philosophy.

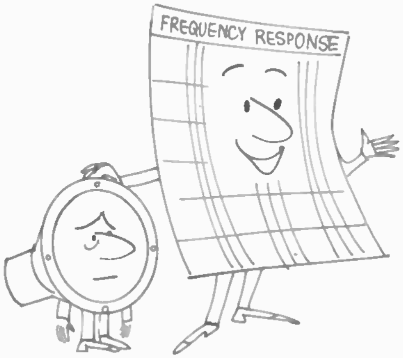
The features about a woofer enclosure with a big unit necessary to produce good response down there also apply to a mid-range unit at the lower end of its range. And the features that apply to getting smooth response in a tweeter apply equally to the upper end of a mid-range unit's response.

Mid-range units can be the extended-range type (a simple cone unit, designed for this specific job). A horn unit can do the job, but it gets much bigger than the corresponding tweeter horn. Electric units are also quite convenient for mid-range work. Any of them *can* be designed to have the desired directional characteristics and a smoothness in response with low distortion.

What does frequency response tell you?

But the published frequency response (if one is published) does not always indicate how the unit will sound. Two things make the measurement and assessment of frequency response difficult at the higher frequencies – above, say, 4,000 cycles: resonances and standing waves.

Both can cause extremely sharp changes of response with very small changes in frequency, measured at one spot by a microphone. The difference is that the sharp spike due to a resonance will show wherever the microphone is put, because it's actually in the speaker output curve. But the sharp spike due to a standing wave will be missing, or moved to a slightly different frequency, if you move the microphone a few inches.

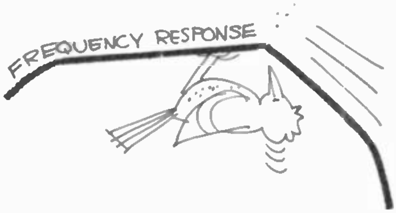


Because of this, one either has to average the ups and downs, assumed to be due to standing waves, or wobble the frequency a little to prevent standing waves from building up. Unfortunately, both of these things really do the same thing. Wobbling the frequency also prevents a resonance effect from building up and tends to average the response of the unit over those few cycles. In short, it's extremely difficult to tell whether a particular spike is due to standing waves or a resonance and, if you average it, you never know.

If it's due to standing waves, a transient sound will not start it and, at any given spot, the ups and downs average to what the speaker actually gives in that range.

But a resonant peak from the speaker unit does not necessarily have a corresponding dip. Even if it does, the effect will still show up on transients. And if the peak does not have a corresponding dip, the averaging process can lead to the impression that the response is holding up when really it isn't.

Keeping the end up



None of the units is free from the possibility that the apparent high-frequency response is maintained by a resonance or resonances. The direct-radiator cone unit, both mid-range and tweeter, may maintain the top end of its response (according to size) by breakup modes of vibration. The cone ceases to move back and forth in one piece (as a piston) and starts vibrating in a complicated manner.

Almost any cone with straight sides will do this, without any effort to try and make it do so. The usual approach has been to select a material that sounds good or produces the best-looking frequency response under measurement. This merely means that the breakup resonances happen to make the response look better instead of worse.

Or holding it down

The way to stop these breakup resonances is to use some shape other than a straight-sided cone to reinforce the sides in some way, or to break up possible modes of vibration. The first of these methods uses a curved contour for the "cone" (strictly it is no

longer conical). Advantages and disadvantages of this were discussed in the previous chapter.

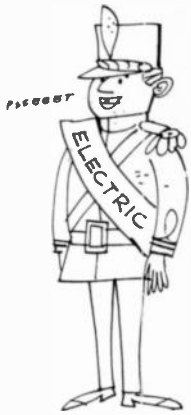
The second method uses foam plastic or other material, in the form of radial reinforcing ribs, cemented to the cone. The third method adopts the opposite approach of impressing slight circular corrugations in the surface of the cone to destroy its rigidity, so the motion is controlled all the way up the frequency range. Instead of breaking up in an erratic way that produces resonances, this sort of cone gradually uses less and less of its total radius in radiating the higher frequencies.

Throat trouble

Horn speakers have their troubles in the throat. The horn is usually designed on mathematical calculations and, provided the material does not resonate (as thin tin is apt to do), it cannot cause any serious deviation in performance. To do a good job, the mechanical clearances between the diaphragm and surfaces leading to the throat need to be almost infinitesimal. This is where the peaks and dips in horn response usually get made. The inexpensive units are usually made sloppy in the acoustic sense, with the object of having them rugged in the mechanical sense. It pays to buy a good, precision-built horn.

Whistling complex

Electric speakers are not so free from possible resonances as their partisans claim. The diaphragm must be enclosed to get the electric drive to make it work. Holes have to be provided to "let the sound wave out." So the most common problem with electric speakers is acoustic resonances in the hole structure — a tendency to want to whistle!



The diaphragm material can also have characteristic "rustle" effects, due to its material. This the horn speaker is not exempt from either. The cure for this, if a metallic diaphragm is used, is to make sure that the metal is not prestressed in its manufacture, so as to want to ring. It should be completely relaxed, or annealed in its working shape.

If the diaphragm is normally under tension (as it is in some electric units), then that tension should be the simple one intended and not complicated by other tensions left in during manufacture.

Please remember the transients

These are some of the tricks to making a good speaker unit. There are many more — enough to fill several books. But in speakers, at least as much as elsewhere, the performance on transients is

a very important factor. A resonance may be well concealed in the frequency response. Sometimes two resonances buck each other so the result cancels and looks good on the frequency response. But the transient performance will find it. The unit will sound quite definitely "colored" at the frequency of the resonance.

Frequency response and transient handling, along with freedom from distortion and dynamic range, are not everything with speakers. Directionality can be important too. In the complete unit, a two- or three-way system, this means integration. The parts of a composite sound, made up of many frequencies, should all seem to come from the same spot when you sit in any normal listening position.

This requires careful attention to positioning of units, so the sounds merge together well, particularly in the vicinity of cross-over – frequencies where more than one unit contributes to the total radiation. At these points, the two units should be working together, not opposing one another.

As well as attention to placing, this needs care in connection so the diaphragms work together. The crossover unit needs to be chosen to suit the electrical and physical characteristics of the units and correctly connected. Reversal of leads to one unit can completely spoil the integration of a perfectly good system.

chapter 12

stereo

THE fact that human hearing is normally accomplished with two ears, and that two ears are obviously better than one, has intrigued researchers for quite some time. In fact, the first demonstration of binaural listening was conducted at the Paris Exposition in 1881, in conjunction with a performance at the Opera House. So the concept of stereo is not new.

But its exploitation is quite recent. After those early experiments in Paris, and some more recent ones in this country conducted by Bell Telephone Laboratories around 1930, the first experiments in which the public could more generally participate began in the 1940's using the transmission of two channels by radio, usually one AM and one FM, although some have used two FM stations.

In the early 50's recording of stereo got started on tape. By this time, the half-track tape had become well established for a great many home recorders, so it was a relatively simple step to use both tracks going the same way, for stereo.

But 1958 will be remembered as the big year for stereo. This was when stereo on disc first saw the daylight of publicity. And now stereo multiplex on FM is gradually feeling its way onto the air.

How to use it

But there's much more to stereo than finding somewhere to put two channels, on radio or disc. The real questions this chapter

has to deal with are what goes on the two channels, and what you do with it to get the right – or best – effect on playback.

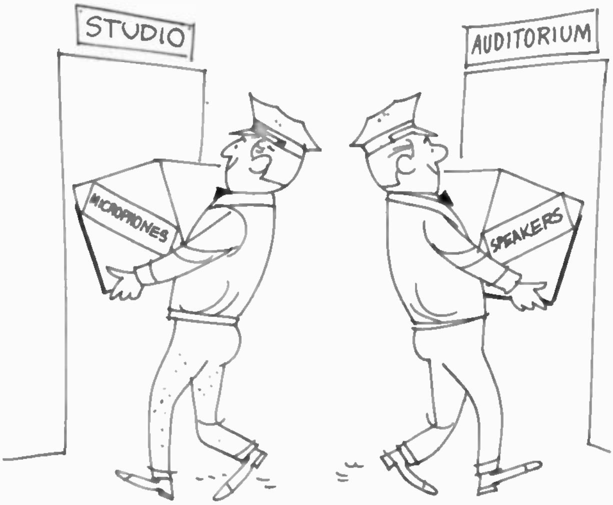
The first idea was that the two channels should carry the same "information" as would normally be received by each ear of the listener in a typical listening location. In some of the early experiments, using headphones for playback, with one channel connected to each earpiece, this was quite successful within the quality limitations of the system. It still works, and some prefer this *binaural* reproduction to stereophonic.

Basic concept

The first idea of stereophonic reproduction was based on a concept of many channels rather than just two. It visualized one wall of the studio covered with microphones, each with its own channel. At the "receiving" end, the channels would be connected to correspondingly placed speakers on one wall of the listening room or auditorium. It seemed this would provide, by electronic means, an acoustic "transparency" between the two walls.

The idea worked. So the next step was to see how it could be used practically – how few channels would do. It was found that taking away channels made little difference, until you got down to three. Reducing from three to two made quite a difference, but even two was a remarkable improvement over one.

Naturally, though, these experiments used somewhat ideal arrangements that cannot always be duplicated for all performances. How would you apply this technique to obtaining a stereo recording of a performance in Carnegie Hall, for example? If you



put the microphones along the back wall, the sound picked up along this big wall has to be compressed to go into the size of your living-room wall on playback. And you're sitting farther back, in effect, than the back wall of the auditorium.

Adaptation

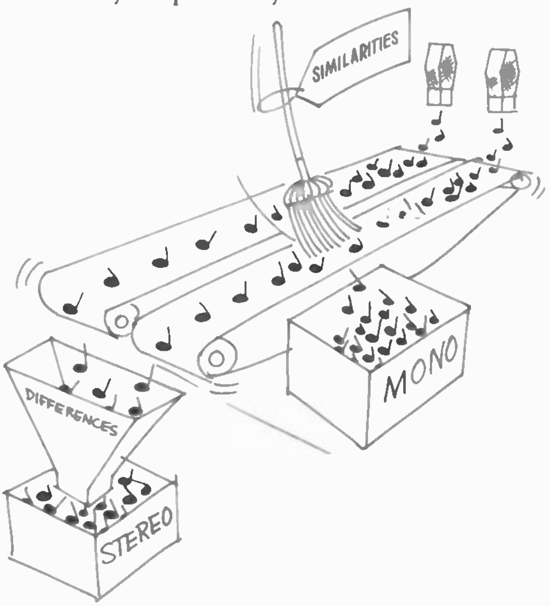
So the mikes should be along the footlight line? Then the reverberation picked up comes from behind the mikes, but in

your living room it still comes from the speakers. The impression you would get is more that of listening through the stage back wall. It begins to become apparent that some "studios" just don't fit into this ideal method of getting stereo.

The Europeans have been working along a different line of attack. They asked themselves the question, "What are the differences in sound picked up by the two ears, that enable us to get the sense of perspective?" What we have to preserve in the microphone technique is the means to reconstruct a sound field that will give a satisfactorily similar illusion.

Similarities and differences

On this basis, they came to the conclusion that there are many similarities between the programs heard by the two ears. Heard separately, in fact, you probably could not tell which is which.



Then there are differences which enable your hearing faculty to get the "depth" or stereophonic illusion.

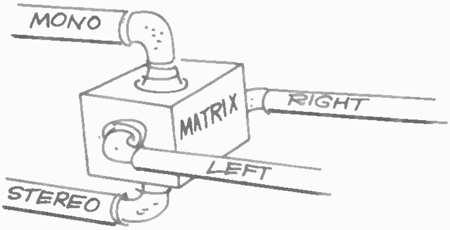
Instead of spacing mikes out, their approach was to use a single microphone location. But, unlike the binaural reproduction method, which uses two mikes mounted in the ear positions of a dummy human head, they use one microphone to pick up the similarities of the sound, which we could call "monophonic," and another that detects the difference components that would exist, between the two ears, which could be called the "stereophonic" component.

This is the MS system. M and S do not really stand for monophonic and stereophonic, but for German words describing the spatial function of the microphones, but the English words mono and stereo fit very well.

The stereo mike is a ribbon or other bidirectional type, placed sideways on. Sounds coming from the center hit the stereo mike right on its dead spot and don't get picked up at all. Sounds from the left get picked up one way, so as to add to the mono pickup, while sounds from the right get picked up the opposite way, so as to subtract.

Matrixing

To reconstruct "left" and "right," we need only electrical adding and subtracting circuits. Adding cancels the "right" and doubles on the "left." Subtracting does vice versa. Thus from pickup at one spot, a composite double program is produced that can be used to make stereo sound that suits the scale of any listening room. At least that's the idea.

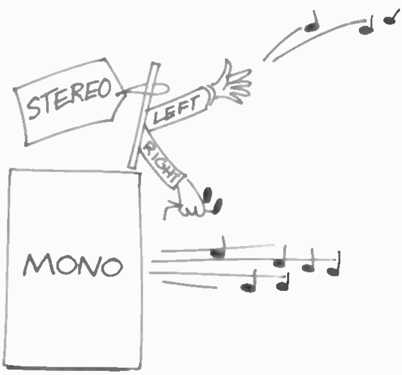


The adding and subtracting can be done either before the program is recorded or after it is played back. This mutual adding and subtracting to get left and right from mono and stereo is called *matrixing*. It is not necessary to do it electronically. It can be done acoustically or mechanically.

Acoustic matrixing

To do it acoustically, a speaker setup quite similar to the microphone arrangement is used. The main speaker faces front and can be any conventional wide-range system. This plays the mono channel. Then a speaker mounted on a small baffle board, edgeways on, plays the stereo channel.

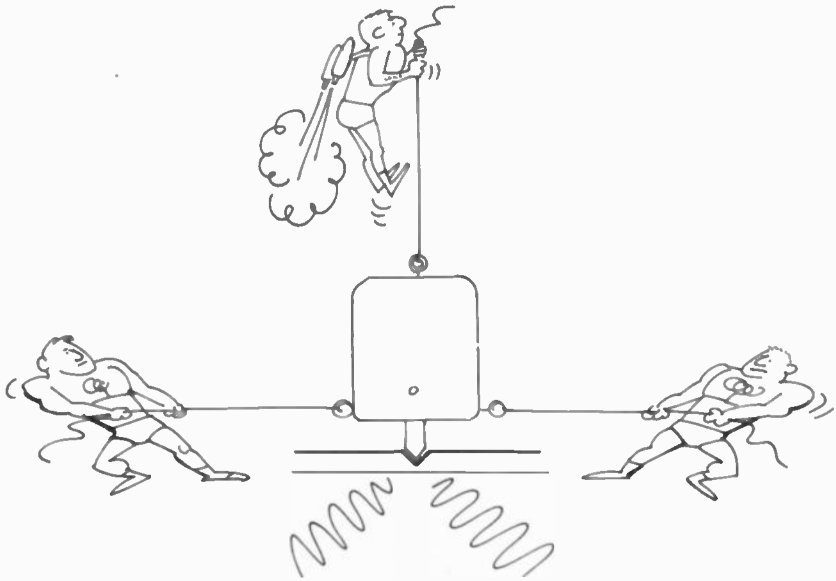
Center program has nothing



on the stereo channel, so only the main speaker sounds. Left and right program have something on both. The stereo speaker diaphragm moves in opposite ways, relative to the main speaker, according to whether the sound is to be left or right. The way the sound waves combine in space determines which way the apparent source of sound is pushed by the stereo speaker. This too works – quite well in certain places.

Mechanical matrixing

The mechanical matrixing uses the difference between vertical/lateral and 45/45 recording. Recordings on the MS system



use the vertical/lateral. The lateral movement of the stylus represents the mono channel, the vertical the stereo. Now, if you combine vertical and lateral movement at the same time, you get movement at 45°. Adding them will make the 45° angle one way, while subtracting makes it the other way. This means that a recording made from MS miked program as vertical/lateral can be played back with a 45/45 pickup and automatically we have correct left and right channels.

If this conversion works one way, it will work the other way too. This does not mean the two systems are completely interchangeable, but they are compatible.

Intensity or time

When recordings are made by the MS system, with two mikes at the same location, whether the two channels used are mono

and stereo or left and right, the differences are entirely a matter of intensity and polarity. There are no time differences because the sound arrives at both microphones always at the same instant.

In terms of stylus motion, this means the stylus always moves back and forth along a straight line whose angle represents the apparent location of the sound.

But where the mikes are spaced apart, as they are in most American stereo recording, the only time that both channels get the program at the same instant is when the sound source is at equal distance from both mikes — central. When the sound comes from one side, the sound not only reaches the microphone on its own side a little stronger but also a little earlier. So there is a time difference as well as, and somewhat more than, the intensity difference on which the MS pickup technique depends.

In terms of stylus motion, since only rarely will sound appear in both channels at precisely the same time, the movement can be in circles, ellipses, as well as up, down and sideways.

Multiple-mike technique

The MS system is hardly used, if at all, in America. But another method commonly used produces very similar recordings. This can be described as "electrical synthesis." Many microphones are used, enough so that any one instrument is picked up at appreciable level by only one of them. One or more extra microphones are used to pick up reverberation. Then the output from these many channels are mixed in appropriate proportions into the final left and right channels.

A sound in the center will be given equally to left and right. One toward the right, or that the director wishes to appear to be

toward the right, is mixed more strongly in the right channel than the left. If he wished he will put it *only* in the right to get a maximum separation.

This method of recording produces two-channel program whose characteristics are essentially similar to the MS recordings. The difference between left and right is almost entirely a matter of intensity and polarity, with practically no time differences. Beyond this fact, of course, the recordings could be very different.

A complete system

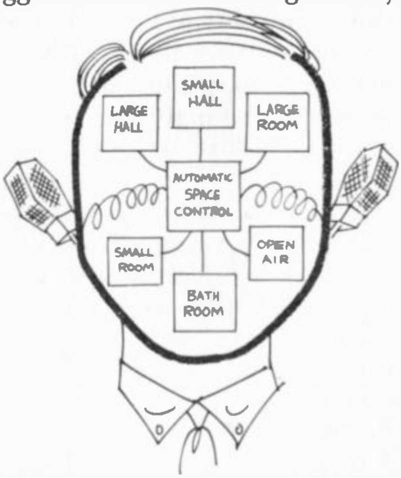
The important thing about stereo is to realize that the complete system is really from the input of sound to the microphones to the output sound from your speakers. This is why what you can learn from Chapter 8 is valuable. Each original concept of stereo or binaural is based on such a complete system. So the overall effect must take into account how the channels are picked up at the studio as well as how you deliver them in your listening room.

Failure to realize this important fact explains many of the prevalent differing opinions. Two reviewers may listen to the same piece of program over different systems without regard to how the program was miked. They may or may not agree about the musical quality. After all, this is to some extent a matter of taste. But they may also disagree about the recording

quality — or the effectiveness of the stereo. This can be because they used different listening arrangements, one suited and one not.

Automatic space control

Further study of this subject suggests that our hearing faculty adjusts its means for getting "perspective" according to surroundings. In a big room, with long reverberation times, the hearing faculty gets accustomed to relatively large time differences. Small ones get lost and relatively meaningless. In a smaller room, the time differences are smaller and the hearing faculty pays much sharper attention to them.



Recently the author conducted a whole series of carefully controlled experiments. The following are the results obtained, with suggestions about the reasons, based on the above deductions.

Basically, there are four possible combinations, from two variations at each end. Microphones can be used to get either intensity differences or differences in space or time. Likewise with speakers: they can rely on differences in intensity, with directional radiation or on space and time differences.

Knowing the combinations

Where both studio and listening room are small, the best results are obtained by using intensity differences at both ends. Directional microphones are used at a single location, or close together, to get pickup in which location is identified by intensity rather than time difference. Directional speakers, or the acoustic-matrixing arrangement, radiate from a spot close together to distribute sound that fills the room according to the intensity differences.

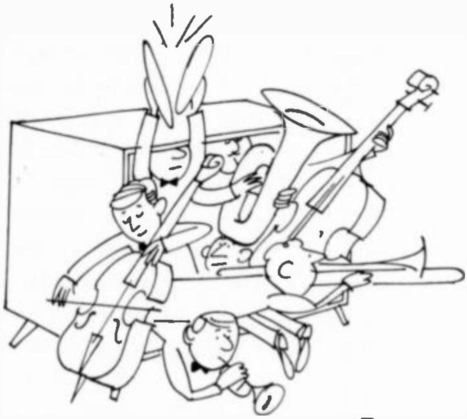
Because both rooms are small, the hearing faculty is extra critical of small time differences. Actually the acoustic-matrixing effect does produce time differences at the listener's ears, just the same as with original sound in a room of equivalent size. But separating either microphones or speakers too far apart introduces time differences that become unnaturally large for the room size.

Where both studio and listening room are large – auditorium size – wide spacing with the consequent time differences are required for best results. This is the basis for making stereo for motion-picture presentation. Any attempt to rely on only intensity differences here gives an overall feeling of inadequacy. The intensities get mixed in the air, and the time difference is not there to help.

Where both studio and listening room are of medium size (dimensions ranging from 15 to 25 feet), we are betwixt and between. A complementary combination is best. If the studio uses directional mikes close together, widely spaced speakers are best. If the studio uses widely spaced mikes, the speakers should be close together and directional, angled apart.

A matter of size

Tying together these interesting facts somewhat was this finding! Almost any recording made in a large hall, such as Carnegie, sounded good played back in the same hall. It was the right kind of sound for that hall. But to get realism in a small room, the technique is much more critical. Our hearing faculty never loses consciousness of the fact we are in a small room, so more critical judgment is possible. But we are listening to *reproduced* sound representing a large hall, so we expect it to have the proper character.



THIS MAY BE ALL RIGHT FOR COLLEGE CRAZES, BUT...

The big problem, of course, is how to mike each particular program so it will sound right in any playback system, or how to set up a playback system in a certain room so any program sounds right played back on it. The first is the recording man's problem, the second is the listener's and, incidentally, the component manufacturers'.

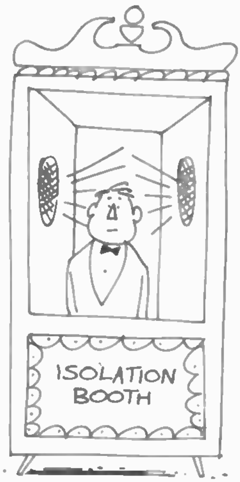
Small room

In a small listening room, the integral system, with directional speakers close together, facing apart, gives best overall coverage. Use of the more conventional widely spaced speakers misses out for different reasons on different programs.

Where the program has been miked for this kind of room (with dominantly intensity, rather than time or space, differences), the sound seems to come from whichever speaker you happen to sit nearest to. Only reasonably good stereo can be heard from a relatively small spot in the middle.

Where the program has not been miked this way and already has some time differences in it, using spaced speakers exaggerates some of these times and minimizes others, creating a "confused" effect.

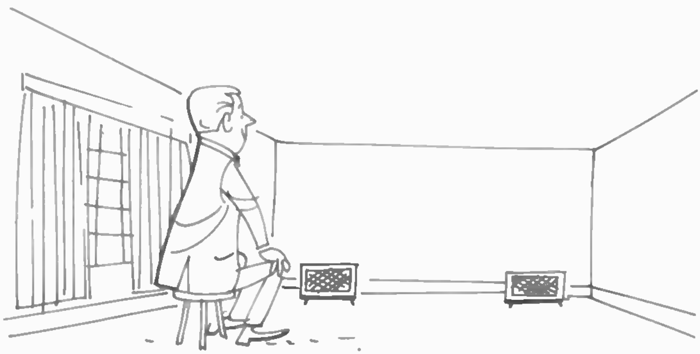
Large rooms



In a larger listening room — auditorium size — the conventional placement is indisputably best. A close-together arrangement, even with directional speakers, can never push the sound out far enough to get a satisfactory perspective effect.

Some systems have been devised to get better coverage of the whole listening area. In a low-ceilinged restaurant, alternating ceiling speakers are more effective (if academically incorrect in places!) than any other way. Everywhere an adequate differential is obtained to give a stereo illusion.

For home listening, use of "outrigger" speakers in various ways, principally to build up the stereo wave pattern, are quite effective. The isophonic system uses an adaptation of the acoustic matrixing we referred to



earlier, with two open backed speakers. This extends effective room coverage much further. In some of these systems, the bass is arranged to come all from one woofer.

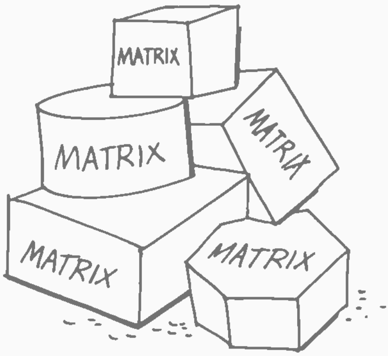
Common woofer stereo does not lose any of its illusion, provided the crossover frequency is low enough to suit the room size. In most rooms, 250 cycles is quite low enough. In smaller rooms, a somewhat higher frequency may be equally effective.

The matrix question

Another area in which considerable controversy has raged, and for some purposes still storms, concerns the proper use of matrixing. As we said before, this can be done two ways, either combining left and right to make mono and stereo, or vice versa. When the FM multiplex idea began to be investigated, two ways of matrixing suddenly became a gross understatement.

The obvious thing, one would imagine, would be to put the mono as the main-channel audio, and the stereo as the multiplexed subcarrier. But a number of people found different reasons to object to this, some valid, some not.

The only valid reason seems to simmer down to the fact that some stations are already transmitting stereo, with one channel on AM and the other on FM. If they multiplex both of them on FM, they don't want to deprive their AM/FM audience of stereo. That's part of it. The other part is that some listeners now listen to either FM and AM only, and don't have stereo at all. They don't want these to have their program degraded.



Plus and minus!

In talking about the channels, they have not used the words we have, mono and stereo, but the mono they have called *sum,* and stereo *difference.* This has led some to believe that the "difference" channel will only have some kind of almost unintelligible hogwash on it, and that any central part of the program will be entirely missing! They should try listening to some before they shout so loud, because it just tain't so.

For there to be this theoretical null zone, the source of sound would have to be precisely at equal distances from both microphones or on the dead spot of the stereo mike. There would also have to be no sound-reflecting surfaces whatever in the studio. In point of fact, there isn't very much audible difference between mono and stereo channels, listened to separately, any more than there is between left and right. One difference there is: the difference or stereo channel is almost invariably deficient in bass.

The transition

Since AM receivers do not have notoriously high quality anyway, they could easily put the stereo channel on that for the time being, with the mono on FM. Also the stereo, of course, on the FM multiplex. People who have AM/FM stereo, it's true, use

better-than-average AM receivers. But they won't lose bass because of this. They'll have that on the FM channel. And for a small extra charge, when they're ready, they can buy a multiplex adapter and have full-fidelity stereo, which they could never have had on AM/FM anyway because of AM's built-in limitations.

But to circumvent this problem, individuals have sought all kinds of "matrixing" so both transmitted channels get a bit of both left and right. Undoubtedly, in due course, this will all settle down, with the blessing of the FCC, to a reasonably simple system.

Another application of matrixing for stereo also uses multiplex. This is the Minter disc system. This puts the mono channel on in the normal way but includes a 25,000-cycle subcarrier which is frequency-modulated with the stereo channel. This is another system that definitely works. It doesn't take more equipment than existing systems, merely different.

Where the 45/45 disc system needs two preamplifiers from both pickup outputs, the multiplex disc only needs one. But then it needs an FM demodulator and matrixing circuit to recover left and right channels to feed to the basic or power amplifiers. All in all, both systems require about the same *amount* of equipment. But the 45/45 had a head start and so has become the accepted system. Additionally, it is compatible with the European MS system, which the multiplex is not.

What are stereo frequencies?

One controversial question in stereo is, "What frequencies are necessary to the stereo effect?" This you will also get different

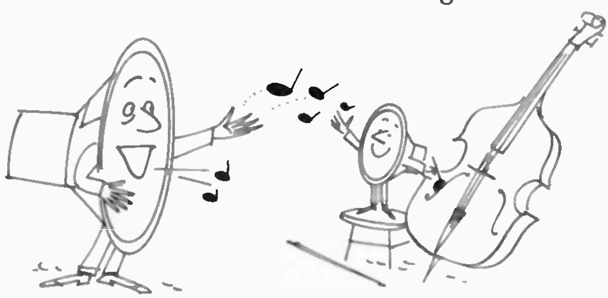
answers to. And largely because they are based on different reasons! Taking the low-frequency end first:

We have already stated that 250 cycles and below can be combined in one speaker without loss of stereo effect. This is true – and if you don't believe us, try it. People argue here because they feel they will lose their sense of location for such instruments as the string bass or tuba.

The fact is these instruments all produce a lot of frequencies *above* 250 cycles, as well as the fundamental frequencies which are usually below 250 cycles. All the transient effects, so important for giving the character of the instrument as well as in locating it, occupy frequencies above 250 cycles.

One-box bass

Until you've experienced listening to one of these common-bass systems that's correctly designed for its job, it may be unbelievable. But you would swear the sound of those big instruments actually



comes from those little outrigger speakers, until you put your ear right into the one it appears to come from and find it must be coming from somewhere else. The illusion is amazingly complete.

Now about the high end. The commonest relevant observation here is that stereo seems to extend the range of a speaker system at the highest frequencies. You may listen to a system that falls off above about 6,000 cycles on mono program. Now play good stereo on a couple of such units and you're ready to swear the response goes on out to 15,000 cycles, or whatever the top limit of your hearing is.

Stereo vs high frequencies

Obviously, just connecting to a stereo program two units that respond to only 6,000 cycles cannot miraculously extend their response to 15,000 cycles. So why does it sound that way? Remember what we said about what tweeters add to the program? They

make it seem clearer, more distinct, *without adding any new sounds*. Which is exactly what good stereo does too, by a different method. It enables you to identify individual sound sources as separate entities, which you could not with mono sound going only to 6,000 cycles.

It takes skilled listening to tell the difference between the clarity due to smooth response above 6,000 cycles on mono, and that due to accurate stereo reproduction without that frequency extension.

Because of this, some have argued that stereo separation of frequencies above a certain point – some have gone as low as 1,500 cycles – does nothing to aid the stereo effect. What this shows is that some people still need their auditory function educated a little more. If you want to say that good stereo with a frequency range up to 6,000 cycles is a good substitute for mono with a range to 15,000 cycles, we'll agree. It does enable the system with more restricted frequency range to sound *better.*

But this does not mean that extending the range of stereo to 15,000 cycles could not improve that too. It can. And it is important to exercise care in how the higher frequencies get distributed by the stereo system. In almost any room, there will be spots where the stereo effect is not so good as at others. There are places in a concert hall where auditory perspective is lacking, *even in the real thing.*

The upper frequency range in a stereo system can help fill in the weak spots in the listening area to an almost unbelievable degree. So the tweeter, or super tweeter, should be beamed to do this deliberately rather than adding confusion where stereo has already achieved clarity. This requires individual case consideration, beyond the scope of this book.

What next?

But we have said enough to show that stereo is young yet — like LP's were about 10 years previously. We can expect some really

significant developments in the next decade. The criticism leveled at stereo, that it is retrograde step from high-quality mono will truly be lived down, and a new advance in high fidelity realized.

Some people ask what will be next after stereo? Most who ask this seem to think that stereo *superseded* hi-fi and they want to know what will outdate stereo. Actually, of course, stereo is just another step *forward* to high fidelity. Explaining this leads to the question whether further channels may be added.

It is difficult to prognosticate, in view of the amount of work it takes to get the best out of two-channel stereo as yet. Some already use three channels, by taking a mixture of left and right to get center. This can be done in a variety of ways. Some pre-amplifiers provide this as an additional output, so you use three basic amplifiers and speaker systems. Another way is less costly: connect one speaker so it gets a combined output from the left and right amplifiers, without mixing them for the left and right speakers.

At this point, it seems unlikely that the complete system will be increased to three channels. Our hearing faculty has always done a good job with only two ears. If two channels are used properly, they should be able to convey enough detail about the program to produce the best realism our ears can appreciate, at least from that viewpoint.

Achieving good stereo is dependent on having a really good, smooth frequency response—if anything, more so than for monophonic reproduction. This means we should expect to see new developments all around that will improve quality so stereo can realize the potential of which it is capable. Many new components have already appeared, but this is only the beginning. But don't wait until it's perfected before you buy, because you'll wait forever.

index

E

F

G

H

I

J

K

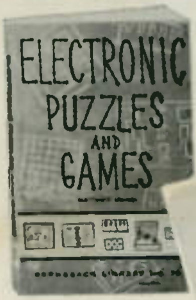
L

M

PRINTED IN THE UNITED STATES OF AMERICA

HI-FI MADE EASY . . . $2.90

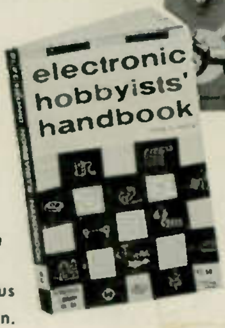
BE SURE TO READ THESE OTHER GERNSBACK LIBRARY BOOKS



ELECTRONIC PUZZLES and GAMES *by Matthew Mandl.* Build your own projects with simple, inexpensive materials. Easy-to-understand and easy-to-construct! No special electronic knowledge needed. *G/L No. 70. 128 pages.* ***$1.95***

HINTS AND KINKS FOR TV, RADIO AND AUDIO *edited by Martin Clifford.* Hundreds of ideas and suggestions to make servicing easier, faster. Time-saving and money-saving gadgets you can make to help you in servicing. From *Radio-Electronics Magazine.* *G/L No. 103. 128 pages.* ***$2.35***

MODEL RADIO-CONTROL *by Edward L. Safford, Jr.* Want to know more about this fascinating hobby? Learn how to put together a complete radio-control system of your own. Control planes, boats, toy cars and trucks at a distance. *G/L No. 74. 192 pages.* ***$2.65***



ELECTRONIC HOBBYIST'S HANDBOOK *by Rufus P. Turner.* Scores of circuits for planning, designing and building amplifiers, oscillators, intercoms, door openers and numerous other hobbyist's projects. Also covers safety and shop practice. *G/L No. 69. 160 pages.* ***$2.50***

FUN WITH ELECTRICITY *by Tom Kennedy, Jr.* Build and learn! A unique introduction to electricity plus shop projects that make learning fun. Ideal for the practical hobbyist or student. *G/L No. 83. 128 pages.* ***$2.65***

TRANSISTORS—THEORY and PRACTICE *(Second Edition) by Rufus P. Turner.* Learn how transistors work, how to use them in circuits, how to get the most out of them. Theory written so you can understand it—circuits explained so that you can follow them. *G/L No. 75. 160 pages.* ***$2.95***

GERNSBACK LIBRARY INC.
154 West 14th Street
New York 11, N. Y.

Use the Gernsback Library book way of getting more knowledge and more enjoyment out of electronics. You can increase your understanding and enjoyment of high fidelity, build your own projects, use the latest circuits. Learn every phase of electronics. All of this—and more, in Gernsback Library books.