Some Principles of Radiotelephony

Part I — *Plain Talk About A.M. Fundamentals*

BY BYRON GOODMAN,* W1DX

Actually the basic principles of radiotelephony are not complicated. At least they are no more so than most other phases of radio that we amateurs are interested in. Why so many misconceptions about radiotelephony should have developed through the years is only a matter for conjecture, but surely much of it can be attributed to a lack of knowledge of fundamentals.

Some (or all) of the information in this article may not be new to you. It is being included so that a complete picture can be presented, and the only assumption we'll make is that you have passed an exam or have hopes along that line.

Sound

Everyone knows that he "hears" with his ears. If questioned further, he might add that what he hears with his ears is "sound." If he remembers some of his high-school physics, he can tell you that "sounds" are vibrations in the air caused by any of a myriad of sources. The only requirement of a sound-generating source is that it be able to transfer a vibration to the air (or other medium, since sound can travel through gases and most liquids and solids). Vibrations of less than about 20 cycles per second are "felt" rather than heard — examples are the rumblings of passing trucks or trains sometimes felt through the floor or earth, or the blast from an explosion. The average person can "hear" as high as 12,000 or 15,000 cycles per second, with a few capable of hearing up to 20,000 or above. Many animals can hear much higher than this. The important thing to remember is that sound is a mechanical vibration transmitted through gases, solids and liquids — our inability to hear it doesn't make it something else, although sounds outside the normal hearing range are usually called "sub sonic" or "super sonic."

Various sounds have two basic properties that enable us, through our hearing, to distinguish them from others. These basic properties are the *pitch* (or frequency) and the *amplitude.* For example, our ears can tell us when different keys are hit on a piano (different pitches or frequencies), or when the same key is hit harder or softer (different amplitude or loudness).

It is relatively rare to encounter a sound that is a single frequency — most of them will contain several frequencies. The piano notes mentioned above, for example, do not consist of a single frequency, and that is true of most musical instruments. Instead, a single musical note will be made up of a predominant frequency plus several others of lesser amplitude. The same note (predominant frequency) played on another instrument will sound different because the amplitudes of the lesser-amplitude frequencies will be different than for the piano. Sounds that are made up of more than one frequency are called "complex," in contrast to a single, or "pure," frequency. The "voice" sounds are quite complex, and are always made up of at least two or three frequencies at any instant. The closest we can come to a pure frequency is in whistling, and even then it isn't too close to pure.

In a wonderful manner that has never been duplicated mechanically, we generate voice sounds through complex control of muscles and air in the throat, and transfer the vibrations (variations in air pressure) to the air just outside of the mouth.

Telephony

The DX limitations of the human voice were recognized some time ago, and several methods have been devised for increasing the range of communication by sound. One method is to "can" it on a solid of some kind, carry it bodily to the distant point, and by some means get it out of the "can." Examples of this are the phonograph and tape and wire recordings — they work, but a significant time delay is involved.

A much faster approach is to use a direct link of some kind between the speaker and the listener. The simplest answer to this is the familiar "tin-can telephone" that children make with two tin cans and a long piece of string. Here the vibrations of the speaker's voice strike the tin can and make it vibrate the same way. This vibration is transmitted directly to the string, which in turn vibrates the tin can at the listener's ear. That tin can then vibrates the air around it, and the sound travels a short distance through the air to the listener's ear. The tin-can telephone has distinct DX limitations, and is not offered as a substitute for the electric telephone.

The telephone we use every day starts out like the tin-can telephone, in that the speaker's voice vibrates a thin plate called a "diaphragm"

• Here's some "down-to-earth" talk about the whys and why-nots of a.m. radiotelephony. Reading it won't make you an expert on single-sideband (sidebands aren't even mentioned) or even the Einstein Unified-Field theory, but it will give you a clearer picture of some aspects of a.m. and basic electricity, if you are a little hazy in that department.

* Assistant Technical Editor, *QST*.

that is mounted in the "microphone" you speak into. These vibrations are transformed into similar variations in an electric current. In other words, if at some instant your speech consists of a 500-cycle and a 1500-cycle tone, the current through the microphone varies at the same rate and in the same proportionate amplitudes. This electric current passes along wires to the listener's "headphone" or "receiver," where it is used to vibrate another diaphragm.

At first glance it would seem that there isn't much basic difference between the tin-can telephone and the electrical telephone (there are obvious physical differences). But there is, and it is a big difference. In the tin-can telephone we always worked with sound, although it went from sound in air to sound in a solid and back to sound in air again. In the electric telephone we went from sound in air to *electricity in wires* and then back to sound in air. In other words, to be able to duplicate the speaker's sounds at a far

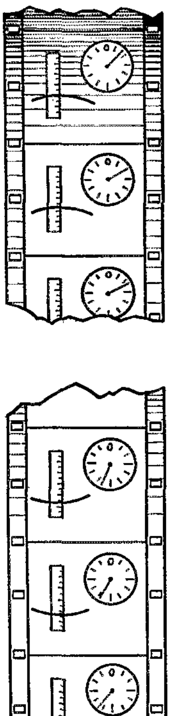


Fig. 1 — Sections of a high-speed movie taken of microphone-diaphragm positions, with a high-speed clock in the background for timing. (The time could also be established by the film frame.) Adjacent film frames show very little change, because the pictures are taken at high speed. Other sections, taken when the diaphragm is moving fastest, would show more change in the diaphragm from frame to frame.

point we had to make an electrical current vary in the same way that the air pressure was varying at the speaker's mouth, and then re-create these air-pressure variations near the listener's ear. It is possible to substitute a beam of light or heat (or radio) for the wires, and thus "transmit" sound over these links. But always bear in mind that the sound itself is not transmitted — it is translated into similar variations in the link medium and then translated back into sound at the far end. You can't "hear" 60-cycle house-wiring electricity — you can hear a 60-cycle "hum" if the electrical variations can be translated into mechanical vibrations.

Radiotelephony

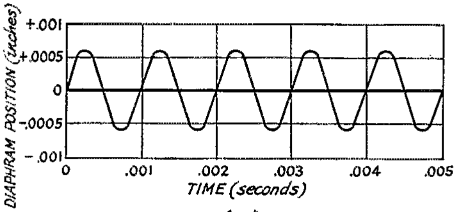
Before we settle down to a close look at radiotelephony, let's draw a picture or two. Suppose we could take some slow-motion movies of a microphone diaphragm. The movies would have to be taken through a microscope, because when the diaphragm is made to vibrate by sound hitting it, the diaphragm doesn't move very far. But to enable us to see how far the diaphragm does move, let's assume our movie is taken against a small scale graduated in ten-thousandths of an inch. And just to make everything still more scientific, let's put a clock in the picture, a special clock that shows hundred-thousandths of a second. Two random sections from the film might look as in Fig. 1 — only the clock hand seems to move in *adjacent* frames. However, when any such slow-motion movies are projected, we will see the diaphragm lazily moving back and forth as sound hits it.

Finally, let's take two movies — one with sound that is a pure (single) frequency of 1000 cycles per second, and another with a complex sound like a voice syllable. When we project these movies, the one using the 1000-cycle pure sound will show us a diaphragm moving back and forth at a regular rate. It will look just like a rope being twirled by two children for rope-jumping — viewing the rope from the side (and without 3-D) it looks like the rope is moving up and down. It moves most at the center and least at the edges.

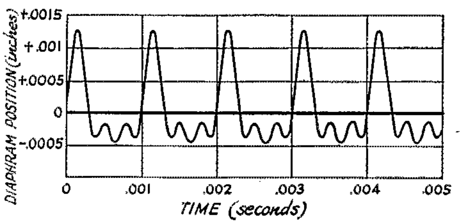
The other movie, the one with the complex sound, will show the diaphragm moving in a crazy sort of way. It will start down, back up a little, go down more, back up, and so on. However, if we watch it long enough we might detect a regular repeating pattern.

Now suppose that, having nothing better to do, we take the movies frame by frame and make a tabulation of diaphragm position (in ten-thousandths of an inch above or below the resting point) and the corresponding time as indicated by the clock in that particular frame. Some weeks later, when we've finished the job, we can then transfer these values to a chart. The charts will look like Fig. 2, if we've done our work accurately. But, you say, those curves don't show the individual points — they're drawn as lines. No, the points are there, but they're so close together the ink ran a little and made lines.

If your high-school physics is still coming back, you will recall that the shape of the curve in Fig. 2A is called a "sine wave." This term "sine wave" is one you will hear kicked around in radio and electricity a lot, so just remember how you arrived at it, by plotting the excursions of a diaphragm driven by a single frequency (or the excursions of a rotated rope as viewed from the side). One cycle of this particular sine wave (a cycle is the complete action that keeps repeating) takes 0.001 second — this time is called the "period." Dividing the period into 1 gives the "frequency" ($1 \div 0.001 = 1000$ cycles per second) and, conversely, dividing the frequency into 1 gives the period ($1 \div 1000 = 0.001$ second). You'll be talking "frequency" in radio,



(A)



(B)

Fig. 2 — **Hypothetical plots of microphone-diaphragm positions *vs.* time, as taken from a slow-motion movie.**
(A) A single audio frequency (rare in nature).
(B) A more likely "complex" signal.

so you don't have to remember that stuff about "period."

About the only thing you can say for the complex wave in Fig. 2B is that it's a peculiar shape. But wait a minute — there are a couple of significant points. For example, it repeats itself every 0.001 second, so its predominant frequency (actually its lowest, or "fundamental," frequency) must be 1000 cycles. And the diaphragm moves farther in the "+" direction than it does in the "—" — how can that be? Any complex wave can do that, but if you examine it carefully you will find that the *area* above the "0" line is the same as the *area* below. This has to be true of any complex wave like this — the *average* excursions either side of the mean must be equal, although the *peak* excursions may be unequal. Don't ever forget this point about the average of any waveform — if it isn't the same above and below the mean, something has been added. (We'll run into that something later on.)

By now it has probably occurred to you that if we were to plot the current through the microphone that we used in the movies, a plot of the current would look just like Fig. 2A, except that the vertical scale would read "Current (amperes)" instead of "Diaphragm Position." Of course the scale (numbers) might be wrong, but that's all relative and doesn't change the shape of the curves at all. Notice that the current would be increasing or decreasing about some value (marked "0" in the sketches) — this value is called the "d.c. (direct current) component," since it is the steady value about which the changing value, or "alternating current" (a.c.), swings. It is the steady value that would be read by a d.c. milliammeter, since the needle wouldn't be able to follow the rapid current changes. If this "a.c. + d.c. signal" is fed to a transformer- or capacitor-coupled circuit, only the variations can get through, and we will have removed the d.c. component. Carbon microphones give an output that has a d.c. component, while crystal or dynamic microphones give an a.c. output with no d.c. component. It makes no difference, of course — you're only interested in the variations (a.c.). Fig. 3 illustrates this point.

Thus far we have an electrical current, or signal, that follows the diaphragm variations which, in turn, are following the air-pressure variations that we call "sound." Our next problem is to get it to some distant point by radio, and this you can already guess about. We will take our radio transmitter and turn it on, so that we have a steady flow of power from the antenna and (we hope!) through space to the receiving antenna. To the receiving antenna we connect a radio receiver, a device of

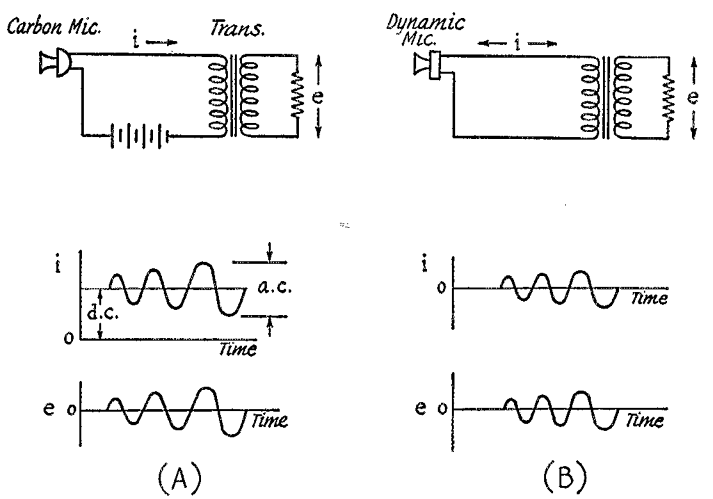


Fig. 3 — **Illustrating the presence of a d.c. component in the current through a carbon microphone (A) and the lack of it through a dynamic microphone (B). In either case, the resultant output after transformer coupling is the same.**

The carbon microphone controls the flow of battery current in (A), while the dynamic microphone generates its own alternating current (B). The permanent-magnet field of the dynamic microphone might be considered to be the d.c. component in the latter case, although it isn't obvious. If the dynamic microphone had no permanent magnetic field, it could generate no output.

some kind that responds to radio-frequency energy. For convenience right now, let's think of the receiver as something that gives, say, one volt of d.c. output for a given radio signal. This receiver has an electrical output proportional to the amplitude of the radio signal, so if we double the amplitude of the signal that gave one volt output, we will have two volts output. And, of course, no signal gives no output. Thus the receiver output is proportional to the signal coming from the transmitter, if all other things remain constant (no fading over the radio path).

Now we take a "modulator" and connect it to the transmitter. This modulator has the ability to control the transmitter output in accordance with an a.c. signal fed to the modulator. If we take the signal of Fig. 2A and feed it

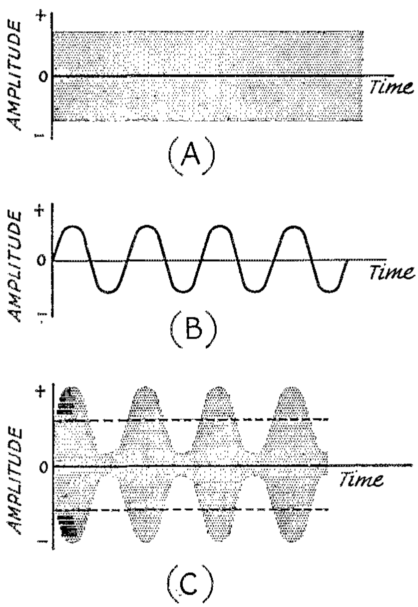


Fig. 4 — (A) Plot of a constant-amplitude radio-frequency signal *vs.* time. Because the time interval is long, individual cycles are hard to see. If the sketch could be expanded along the "Time" axis, the cycles would have the same shape as in Fig. 2A.

(B) A single audio frequency plotted to the same "Time" scale as Fig. 4A.

(C) When the audio freqnency of Fig. 4B is used to control the amplitude of the r.f. in Fig. 4A, the resultant would look like this.

to the modulator, the transmitter output will increase and decrease about its steady value in just the way that the signal (and also the microphone-diaphragm position) did. If we use the signal of Fig. 2B, the same thing will happen, and the transmitter output will vary in this more complex manner.

Since the output from the receiver is proportional to the transmitter output, it's easy to see that we now have a signal at the receiver that varies like the transmitter (and the microphone diaphragm). All we have to do is take the electrical signal at the receiver, feed it into a headphone (or loudspeaker), and the diaphragm there will re-create the sound that was striking the microphone diaphragm. In the earlier paragraph the receiver output was called "d.c.," but it should be apparent that if this d.c. changes, as a result of the changing signal, it becomes the "a.c. + d.c." signal mentioned earlier. If the headphones are connected to the receiver output through a transformer, the diaphragm responds only to the a.c., which is all we're interested in anyway.

This process of radiotelephony is called "amplitude modulation," and its derivation is obvious: we are changing, or "modulating," the *amplitude* of the radio signal. There are other methods of using a radio link for reproducing sound (f.m. or "frequency" modulation, and p.m. or "phase" modulation), but amplitude modulation ("a.m.") is the most common.

Incidentally, that steady value of transmitter output we have when no modulation is present is, unfortunately, called the "carrier." This leads some people to think that it "carries" the sound, which of course it doesn't. It is there to give an operating point about which the amplitude can swing. Many amateurs wonder why the carrier is required at all, and some of them "invent" ingenious ways for doing without it, but the sad fact is that it *must* be there when the modulation takes place, and it must be there again at the receiver when "detection" takes place. In between we can sometimes do without it, but that's another story.

Perhaps you have been wondering why we don't just connect the transmitting antenna to the microphone output, and use the voice-controlled electrical current to jump through space to the receiver. This would be a very convenient way to do it, but the trouble there is that electrical currents at audio frequencies don't travel far through space, and higher-frequency ("*radio*-frequency") currents are required.

To review this business a little, let's draw a picture of the r.f. we've been talking about. Fig. 4 illustrates a few of the points. For simplicity, only the single-frequency modulation case is used, but it should be remembered that in practice a voice signal is always complex. However, the single-frequency case is easier to talk about, and everything that is said about it applies in the same way to the complex case.

The unmodulated r.f. is shown in Fig. 4A. Although it is a sine wave like Fig. 2A, the frequency is so high that the cycles are crowded close together and you can't distinguish them when they're drawn to this scale. Fig. 4B is a duplicate of Fig. 2A — it represents the single-frequency modulating signal. When it is applied to the modulator, the modulator varies the output of the transmitter accordingly, and the resultant r.f. output is as in Fig. 4C. The dashed lines represent the no-modulation, or carrier, level, about which the output varies. You have heard, or certainly will hear, about "percentage of modulation." This is a percentage of the available carrier that is modulated or changed. If, for example, the sketch of Fig. 4C showed the amplitude at times being reduced to zero instead of not quite reaching zero as shown, it would be said to represent "100 per cent modulation." Since the headphones at the receiver are responding to the *changes* in transmitter output, it is apparent that a high percentage of modulation, for any given carrier, will give a receiver headphone (or loudspeaker) output of greater amplitude. This, of course, is why you hear so much concern expressed over one's "percentage of modulation" — to make maximum use of the available carrier amplitude, the modulation percentage must be kept high.

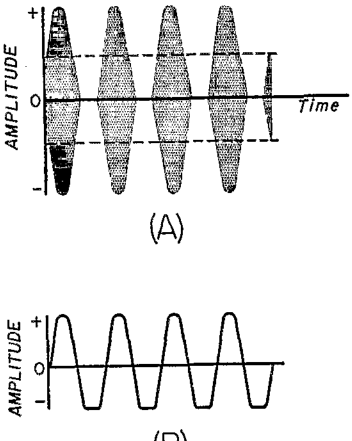


Fig. 5 — If the audio signal of Fig. 4B is used to control the r.f. signal of Fig. 4A beyond its capabilities, the result is as (A). This is called "overmodulation."
(B) The audio output of a receiver receiving this signal would be "distorted" in this manner.

But "percentage of modulation" is somewhat meaningless except for the maximum-amplitude components of speech, because obviously with one's voice varying in amplitude from syllable to syllable, only the highest-amplitude components should modulate the rig 100 per cent.

"But," you ask, "if I talk still louder, won't the output of the transmitter increase even more?" Good question, and the answer is "Yes." But let's take a closer look at what happens, and get the whole story. To simplify things, let's say you're talking with a sine wave, although we know you won't be. If we apply enough sine-wave signal to the modulator, the transmitter output might look like Fig. 5A. Assuming a good transmitter capable of this, the output does increase on half of the audio cycle, and the increased-output swings do go right on up. But since we can't go below zero output, the decreased-swings flatten off, as shown. The receiver output would look like Fig. 5B, and you can see that the resultant audio signal is not the same as the original (Fig. 4B) — the bottoms are flat, and the signal is said to be "distorted." In actual practice, a little distortion like this isn't too bad, and only a trained ear would notice much difference when listening to something like Fig. 4B or Fig. 5B. But for reasons that won't be taken up now, the transmitter signal of Fig. 5A requires a wider band of radio frequencies for its transmission than does that of Fig. 4C. It is said to "splatter," and these splattering signals interfere in adjacent radio channels that would otherwise be clear. It is easy to see that "overmodulated" signals like this would be undesirable in a crowded amateur band. And even though the receiving operator might not notice the distortion, overmodulated signals are illegal by FCC rule because they interfere unnecessarily with adjacent-channel signals. It is important to remember that one cannot be certain of overmodulation when listening "on" the signal (any distortion present might be a result of other causes, or the operator might not be able to hear the distortion). The only check for overmodulation (lacking an oscilloscope properly connected in the receiver) is to listen "off" the signal for "splatter," with a selective receiver that isn't overloading.

An interesting point is the fact that the percentage of modulation is not always the same for positive peaks (increased output) and negative peaks (decreased output). Referring back to the voice signal of Fig. 2B, you will recall that it swings upward more than downward (for reasons given when it was described). If this signal is fed to the modulator, the transmitter output could instantaneously increase more than it decreased, or the converse could be true, depending upon how we connected the modulator. Thus in practice we can exceed 100 per cent modulation on positive peaks without distortion, but we can never exceed 100 per cent modulation on negative peaks without running into distortion and splatter. This is illustrated in Fig. 6.

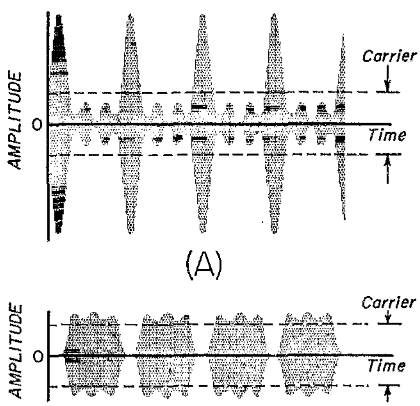


Fig. 6 — Showing how positive peaks can exceed 100 per cent modulation. In each case above, the modulating signal is similar to that of Fig. 2B. In (A) the positive peaks rise well above the 100 per cent modulation. In (B) the polarity of the modulating signal is reversed, and it is obvious that the positive peaks cannot now go to 100 per cent modulation without overmodulation on the negative peaks.

[*Part II of this article will appear in a subsequent issue.* – ED.]