Chapter 9
The Superheterodyne

Now that we understand the fundamentals of AM and FM, it is time to look at the typical radio receiver.

Fig. 9-1 is the same crystal radio we introduced in Chapter 7. We pointed out that it consists of just a few parts:

- An outside wire antenna, which captures the various radio signals coming our way
- An antenna coil and variable capacitor, which do two jobs. The antenna coil is wired as a transformer, coupling the antenna signal to the radio. The secondary of the coil, along with the variable capacitor also form a tuned circuit which selects the station we want, while rejecting stations we do not want.
- A diode which rectifies the AM signal.
- A 0.001 µF capacitor which filters out the high-frequency carrier and sidebands, and keeps only the envelope — the audio signal.
- A pair of headphones which convert the audio signal to sound.

Variations of this radio circuit date back to the early days of radio. The big advantage, of course, is that it is simple, and requires no batteries. But that simplicity carries a price — the radio doesn't work too well. It has poor sensitivity, and poor selectivity.

Sensitivity

Sensitivity describes the ability of a radio to pick up weak signals. Our crystal radio has low sensitivity, because it can only pick up really strong stations.

Sensitivity has to be judged in relation to noise. Just picking up a station is not enough, if the station is so noisy that it is not pleasant to listen to. Spec sheets and advertising literature usually specify receiver sensitivity by measuring how much voltage from the antenna (usually measured in microvolts) is required to make the desired signal (usually the sound out of the speaker) 10 times or 100 times stronger than the noise. This ratio of signal to noise is then called the *signal-to-noise ratio*; a decent radio might provide a 10-to-1 or 100-to-1 signal-to-noise ratio with an antenna signal of under 1 microvolt.

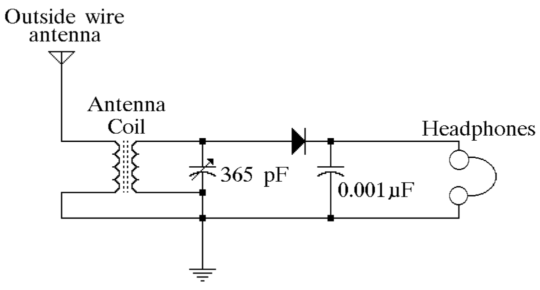

Fig. 9-1. A simple AM crystal radio

This definition of sensitivity is useful for most radio receivers, but not for a crystal radio. Typical receivers have amplifiers which produce noise when tuned to a weak or no station, so measuring the signal-to-noise ratio is possible. With a crystal set, however, there is really no noise to be heard from the headphones, so measuring the ratio is tough. Still, you need several hundred thousand microvolts of antenna signal to hear anything at all, so sensitivity is clearly bad.

The signal-to-noise ratio is usually measured in decibels as follows.

Suppose you turn on a receiver, disconnect the antenna (so there is no input signal), and adjust the volume control to get some measurable amount of noise from the speaker. Measure the speaker voltage and call that the "before" voltage. (For our example, suppose it is 0.1 volts.)

Now connect the receiver's antenna leads to a signal generator, properly adjust its frequency so the radio can receive it, and set the generator output until the receiver's speaker voltage is 1 volt (10 times more than before.) Call this the "after" voltage; it is the voltage needed to give a signal 10 times stronger than the noise.

Now insert the "before" and "after" values into the formula for decibels:

$$\text{db} = 20 \log \frac{\text{after}}{\text{before}} = 20 \log \frac{1}{0.1} = 20 \log 10 = 20 \times 1$$

(since 10 to the 1st power is 10, log 10 is 1). And so the SNR or signal-to-noise ratio is 20 db.

Now go back to measure the amount of signal coming from the generator into the receiver antenna connection; this value (in microvolts) is the sensitivity to produce a 20 db signal-to-noise ratio in the receiver output.

Strictly speaking, this is not entirely correct, because the "after" measurement is not just the signal, but also includes a bit of noise. Hence many people will call our 20 db value the *signal-plus-noise-to-noise* ratio, rather than just signal-to-noise ratio. They may also write it as (S+N)/N, meaning that the signal-plus-noise is divided by the noise output.

Selectivity

Selectivity describes the ability of a receiver to select the station you want, and keep out other stations that you don't want. Our crystal radio has poor selectivity, because it has trouble separating nearby stations.

As we've discussed in previous chapters, a radio signal consists of a carrier and sidebands, and has a certain bandwidth which depends on the type and amount of modulation. Resonant circuits in the receiver tune in the station you want; ideally, they should pass all the frequencies in the carrier and sidebands equally well, and completely reject all other frequencies below and above that. Practically, however, that is not possible.

Fig. 9-2 shows the actual frequency response of a tuned circuit at the left, and the ideal response we would like to have for a radio at the right. The ideal response would be a rectangle, where *all* the signals within the bandwidth of the radio signal (carrier and sidebands) get through equally well (the response is 1 or 100%), while *nothing* gets through above or below that range (the response is 0 or 0%.)

You can see there is a big difference between what we want and what we get. The ideal rectangular response at the right has

1. A flat top; this lets the carrier and all sidebands get through the tuned circuit equally well.
2. Steep *skirts*; the skirt is the vertical part at the left and right. Steep skirts make sure that the response drops very fast, so that no adjacent stations get through.
3. A definite bandwidth; ideally this should be just as wide as the bandwidth of the signal we are trying to receive — no more, and no less.

The actual tuned circuit response on the left of Fig. 9-2 has none of these. The top isn't flat, so the carrier can get through, but the farther out a sideband is, the less of it gets through. The sides aren't steep enough to keep out adjacent stations, since even pretty far away from the peak, the curve still has fairly high response. And finally, there is no definite bandwidth to the circuit.

We can flatten out the top a bit by widening the whole curve. The bandwidth of a tuned circuit determines the relative width of the curve (see Appendix A.) The bandwidth in turn is determined by the Q or Quality Factor of the circuit. The higher the Q, the narrower the response is; the lower the Q, the wider it is. Unfortunately, there is a conflict here — a lower Q would flatten out the top and thus provide more even transmission of the desired signal, but it also widens the bandpass and makes the skirts even less steep.

It now becomes obvious that a single tuned circuit simply cannot provide the right selectivity for a radio, even under ideal conditions. Even worse, the circuit of Fig. 9-1 is very far from ideal. The antenna and the primary of the coil feed the signal into the tuned circuit, so their resistance affects the Q. The headphones are the load on the circuit, so their resistance also affects the Q. And because all of these have fairly low resistance, the Q is terrible! In a typical crystal radio, the response of the tuned circuit is so wide, and the skirts so broad, that it is almost impossible to separate stations from each other. Unless you are lucky to live very close to one radio station and far from all others, don't expect to get very good performance from such a simple crystal radio.

How do we improve sensitivity and selectivity?

Improving sensitivity seems fairly simple — just add some amplifiers. Well ... it's not quite that simple, because you have to do it just right, as we shall see in a moment, but it can be done.

Improving selectivity, on the other hand, is somewhat more complicated. There are some modern components, such as crystal or ceramic filters, which can provide a fairly sharp bandpass. The more traditional method, however, is to just add more tuned circuits. For example, the dark curve in Fig. 9-3 shows how using three resonant circuits, each tuned to a slightly different frequency, and having different Q's, can improve the overall response. There are several ways of getting this

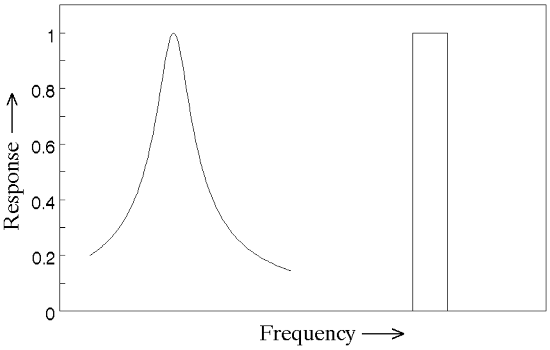

Fig. 9-2. Actual and ideal resonant response

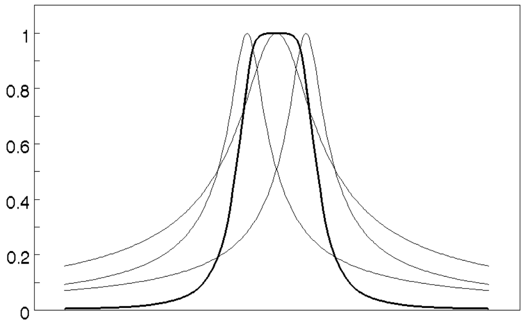

Fig. 9-3. Improving the bandpass by combining tuned circuits

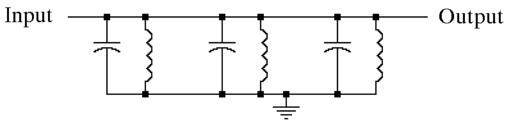

Fig. 9-4. This is one, not three tuned circuits

same result, and obviously the circuit must be carefully designed and set up, or else the resulting bandpass may be lopsided or have lumps in the top.

The problem is that you can't just parallel a bunch of tuned circuits together as in Fig. 9-4. Even though this looks like three separate tuned circuits, if you parallel the three inductors into one, and the three capacitors into one, you see that there is really only one tuned circuit here. To use more than one tuned circuit in the radio, you must somehow separate them so they are not all in parallel with each other. The secret is to separate them with amplifiers.

The TRF or Tuned Radio Frequency Receiver

The TRF, or Tuned Radio Frequency receiver, became popular as soon as the electronics industry got to the point where it was possible to build amplifiers cheaply enough. Fig. 9-5 shows the block diagram.

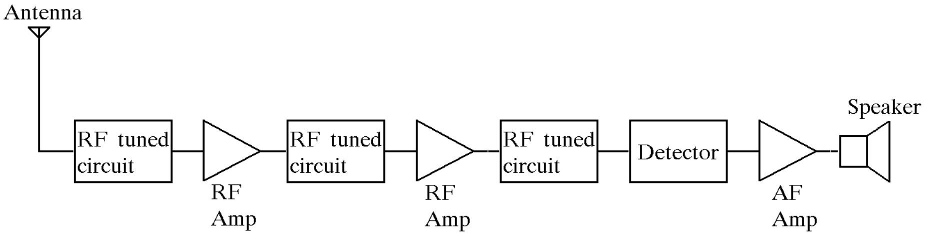

Fig. 9-5. The Tuned Radio Frequency (TRF) receiver

The circuit started with an antenna, usually a long wire strung outdoors. Then came two or more RF tuned circuits, separated by RF amplifiers. These were called RF because they all amplified the actual radio frequency (RF) signal. Eventually came a detector, which was simply a rectifier diode and capacitor, that worked just like those in the crystal radio of Fig. 9-1. This was followed by an AF amplifier, called AF because it now amplified the audio frequency signal, not the radio frequency signal. The audio signal then went to a speaker. You'll note how RF amplifiers separated the tuned circuits, so they would act separately instead of becoming one single tuned circuit, as in Fig. 9-4.

The TRF receiver worked quite well for its time, but it had some major flaws. One difficulty was that, each time you wanted to change stations, you had to retune all the tuned circuits. Although Fig. 9-5 shows only three, some more expensive radios had four or even more. But as Fig. 9-3 shows, even three tuned circuits have to be carefully adjusted if you want to get an overall response with a fairly flat top and steep skirts. It was almost impossible for the average owner to get it right.

A second problem had to do with the actual physical construction of the radio. If two tuned circuits were too close to each other, the two inductors would act as a transformer. Some of the amplified signal from one of the later stages would get back into an earlier stage, only to be amplified again and again — this positive feedback made the radio into a perfect oscillator! And the more tuned circuits there were, or the more gain the amplifiers had, the worse the problem became. It was very difficult to build a receiver that had both high sensitivity and high selectivity.

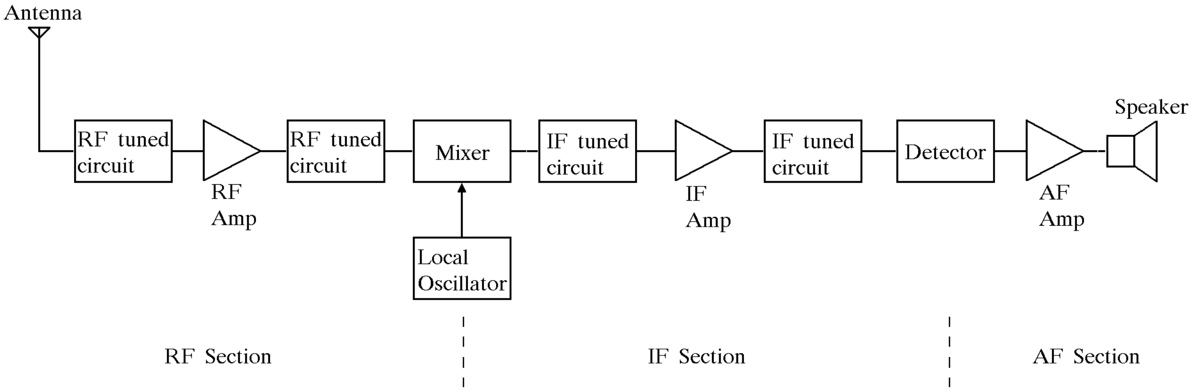

Fig. 9-6. The superheterodyne receiver block diagraam

Superhet to the Rescue!

The solution was the superheterodyne (superhet for short) receiver, invented around 1930. This idea revolutionized radio; all modern receivers use it.

Fig. 9-6 shows the superhet's block diagram. The amplification in the circuit is provided in three separate sections: the RF section, which extends from the antenna to the mixer, the IF section which goes from the mixer to the detector (and is the main feature of the superhet), and the AF section, which extends from the detector to the speaker.

As before, the RF section contains some RF tuned circuits and amplifiers, which amplify the radio frequency signal. Similarly, the AF section contains some audio amplifiers, and amplifies the audio signal. But between them is the IF or *intermediate frequency* section, which further amplifies the signal, but also provides all of the selectivity for the entire radio.

The superhet solves both problems of the TRF design: First, the IF section operates at a different frequency from the RF section; moreover, the IF section stays tuned to the same frequency regardless of which station we listen to. So the tuned circuits in the IF section can be properly aligned in the factory to give the best bandpass curve, and they don't get retuned by the user.

Further, because the overall radio gain is split into three sections, each section's gain is smaller. With less gain, feedback from the output back to the input is less of a problem. And because each of the three sections — the RF, IF, and AF — operates at a different frequency, it doesn't matter if a signal from one section sneaks into another, since it just gets rejected.

So the superhet's significant feature is that the signal in the IF portion of the radio stays at a constant frequency regardless of what station you tune to. This is done by *heterodyning* or *beating* two signals.

DETOUR

Heterodyning is so important to radio that we have to look at it some more.

Consider the circuit of Fig. 9-7. We have a box containing some circuitry, and two inputs into the box; one is a 100 Hz sine wave, the other a 1000 Hz sine wave. What comes out?

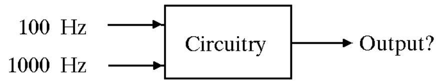

Fig. 9-7. Mixing two signals in some circuit

Assuming there is *something* inside the box (not just empty air!), the two input signals will usually somehow combine into the output. Electrical engineers will now explain that there are two main possibilities:

- If the circuitry in the box contains only resistors, inductors, and capacitors, it is called a *linear* circuit. In linear circuits, the output is proportional to the inputs; there is nothing in the output which didn't come from the input. This is just a fancy way of saying that, if 100 Hz and 1000 Hz go in, then only 100 Hz and 1000 Hz can come out.
- But if the circuitry in the box also contains some diodes, transistors, tubes, or other *non-linear* components, then this becomes a whole new ballgame — things can come out that didn't go in.

For one thing, non-linear circuits can distort; they can change the waveshape of the sine waves going in. As we explained way back at the beginning, this introduces harmonics. So the 100-Hz signal could now produce harmonics of 200, 300, 400, or more Hz, while the 1000-Hz signal could now have harmonics at 2000, 3000, Hz. etc.

Much more important for us, though, is that the two input signals can interact with each other. This process is called *heterodyning* or *beating*. When two signals interact like this, they produce new signals whose frequencies are the *sum* and *difference* of the original two signals. In our case, the sum would be 1100 Hz (1000 plus 100), and the difference would be 900 Hz (1000 minus 100). These new frequencies would be called *heterodynes*.

As usual, things are just a bit more complicated. The distortion harmonics also produce sums and differences. For example, the 200 Hz harmonic of the 100-Hz signal could heterodyne with the 3000 Hz harmonic of the 1000-Hz signal to produce 2800 and 3200 Hz, and so on. Fortunately, the harmonics are usually smaller than the fundamentals, and so these heterodynes are also smaller than the main ones at 1100 and 900 Hz.

At a first glance, you may think this heterodyning is a terrible complication. But remember that, without heterodyning, the superheterodyne receiver would be impossible, and radio and TV reception would be a lot worse today.

END OF DETOUR

Let's now see how heterodyning is used in the superhet. As an example, Fig. 9-8 shows a superhet AM radio tuned to a radio station at 880 kHz. Coming in the antenna is not just this station's signal, but also signals from all sorts of other stations — radio, TV, radar, etc. The tuned circuits in the RF section remove most of the

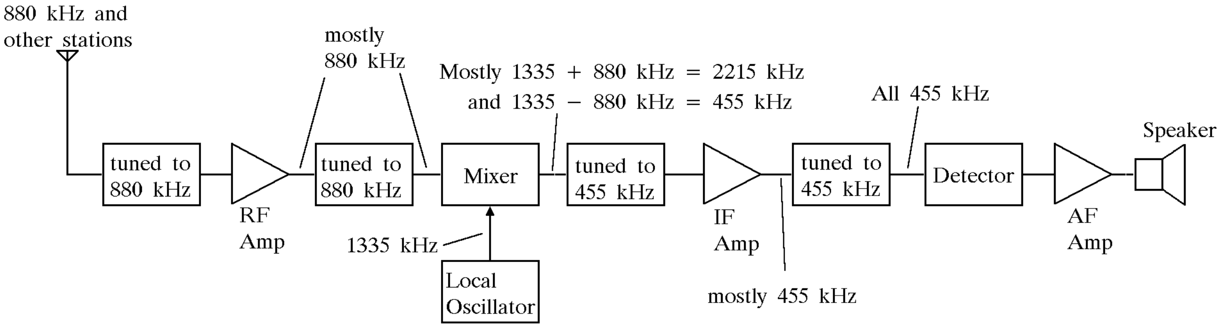

Fig. 9-8. Superhet with a 455 kHz IF, tuned to 880 kHz

undesired signals, but not all, so that the signal coming into the mixer is mostly 880 kHz, but still has many other signals at nearby frequencies.

The mixer is a nonlinear circuit; it receives this combined signal, but it also gets a 1335 kHz signal from the oscillator below it. Since it is nonlinear, it heterodynes these signals. There are a lot of different signals going in so it produces a lot of heterodynes, but the most important ones are the sum and difference of the desired station at 880 kHz, and the oscillator signal at 1335 kHz. This gives us 2215 kHz, the sum, and 455 kHz, the difference.

But note that the tuned circuits in the IF section are all tuned to 455 kHz, so they keep the 455 kHz signal and reject the others. By the time signal gets to the detector, the filtering has been pretty much completed, and the signal is almost pure 455 kHz (plus any nearby sidebands .)

Now, suppose we retune the radio to a different station, say one at 770 kHz. We retune the RF tuned circuits, but these only do a rough job of removing faraway signals; they aren't the main tuned circuits in the radio, so it isn't important to get them just right. But — and here is the important thing to note — we also retune the oscillator to 1225 kHz. The difference between 1225 kHz and 770 kHz is again 455 kHz! And so the IF section again amplifies the resulting signal, without having to be itself retuned.

So the trick when changing stations is to retune the RF circuits (but a slight error here isn't critical), and also retune the oscillator (and this is important!) so the difference frequency between the station you want, and the oscillator, stays at 455 kHz. Since the RF tuning adjustment isn't that critical, it is possible to use a single knob to adjust all the tuned circuits at the same time, without having to worry about whether all of them are right on target.

In case you wonder why we chose 455 kHz for the IF frequency ... well, other values are possible, but this just happens to be a popular one in AM broadcast receivers. FM broadcast receivers usually use 10.7 MHz IF, and other IF frequencies are also used in other kinds of receivers.

If we let $f_{station}$ be the frequency of the station we want, and F_{IF} be the IF frequency, then the oscillator frequency f_{osc} should be

$$f_{osc} = f_{station} + f_{IF}$$

But it's also possible to let

$$f_{osc} = f_{station} - f_{IF}$$

Either way, the difference between $f_{station}$ and f_{osc} is equal to the IF frequency f_{IF}, so either will work.

The Converter

Many radios combine the mixer and the oscillator into one circuit called the *converter*. This is a popular technique for lowering the radio's cost, because several components in the circuit do multiple jobs at the same time. Fig. 9-9 shows the converter used in many popular

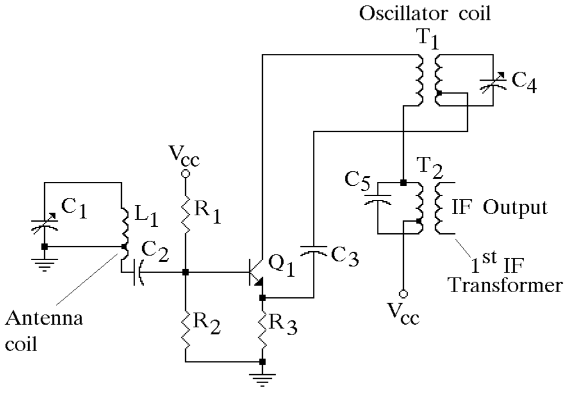

Fig. 9-9. Typical converter in an AM receiver

AM broadcast radios; there are several useful techniques that are worth mentioning.

L_1 and C_1 are the RF tuned circuit, with C_1 being the tuning capacitor. But L_1 does several different jobs. The top part of the winding (above the ground connection) is the part that actually resonates with the capacitor; the bottom part (connecting to C_2) acts as the secondary of a transformer, to bring the signal from L_1 to the transistor without loading down the tuned circuit (which would reduce the Q.)

At the same time, L_1 is also the antenna. As we have seen, coils or loops of wire can act as antennas; in this case, L_1 is wound on a ferrite core (a ceramic core which contains ferrous metal particles); the core helps to pick up the energy from the radio signal, and concentrate it in the coil.

The transistor also does two jobs. First, it oscillates at a frequency 455 kHz above the signal you want to pick up. To do this, we need an amplifier with positive feedback. The transistor is the amplifier, with its output coming out of the collector, going through oscillator coil T_1, and back through C_3 into the emitter of the transistor. Capacitor C_4 resonates with the secondary of this coil to control the oscillator frequency.

At the same time, however, the transistor amplifies the RF signal coming from the antenna coil, and mixes it with the oscillator signal. Because the transistor is non-linear, it also produces the sum and difference heterodyne frequencies. The primary of IF transformer T_2 and capacitor C_5 resonate at 455 kHz, and send the 455 kHz difference frequency on to the IF amplifier.

Note how T_1 and T_2 both use taps on one winding (the tap is a third connection part way into the winding.) This reduces the loading on the resonant circuit, and keeps the Q from being lowered.

Variations on a Theme

Although Fig. 9-6 showed one RF amplifier (with two RF tuned circuits), and one IF amplifier (also with two IF tuned circuits), there is nothing sacred about these numbers. Many cheap radios use only one RF tuned circuit, and no RF amplifier; the converter in Fig. 9-9 is a good example.

On the other hand, more expensive radios might have more RF amplifiers and/or more IF amplifiers. In fact, quite a few radios use *double-conversion*; in this scheme, there are two mixers and oscillators (or two converters), and two different IF amplifier sections. We'll see the reason for this in a moment.

Superheterodyne Sensitivity and Selectivity

By splitting the amplification into separate sections, a superhet can provide more total gain without the danger of signals feeding back and causing oscillation. Further, because the IF amplifier does not need to be retuned each time you change stations, it can be optimized, and carefully adjusted at the factory, to provide the best possible bandpass characteristics — steep skirts and a flat top. But there is more to it than that.

Recall our definition of the Quality factor Q of a resonant circuit:

$$Q = \frac{\text{resonant frequency}}{\text{3–db bandwidth BW}}$$

The 3-db bandwidth doesn't really specify how well the circuit will reject adjacent stations; in order to reject such interference, the response of the tuned circuit has to be 30, 40, or even more db down from the top of the curve at the frequencies of any adjacent stations. Still, it provides a useful yardstick for comparison.

We can rewrite the above equation as

$$\text{3–db bandwidth BW} = \frac{\text{resonant frequency}}{Q}$$

To get a small bandwidth, you have to either make the resonant frequency small, or make the Q big. But in most resonant circuits, there is a limit on how big Q can get; it is affected by the resistance of the rest of the circuit, and is seldom more than 20 or 30. So making Q big is not a feasible approach to making the bandwidth small. So, to get a small bandwidth, it would help if you could make the resonant frequency small.

But in a TRF receiver, you must tune the resonant circuits to the frequency of the station you want to receive, so you really can't make the resonant frequency small to get a good bandwidth. And the bandwidth will change as you tune to different stations, further complicating the design.

In a superhet, on the other hand, all the selectivity is obtained in the IF stages, and their frequency stays the same for all stations. Moreover, the IF frequency is lower than *any* of the stations you want to receive, so you can get the same narrow bandwidth for every station you listen to. In theory, at least, you could get the bandwidth as narrow as you want, simply be going to a lower IF frequency. (Note how AM broadcast radios, which need a lower bandwidth than FM broadcast radios, also have a lower IF frequency of 455 kHz instead of 10.7 MHz.) But there is a fly in the ointment, as they say — the image.

The Image

Let's return to the radio in Fig. 9-8. It is tuned to 880 kHz, has a 455 kHz IF, and an oscillator frequency of 1335 kHz. Here we see that

1335 kHz – 880 kHz = 455 kHz.

So far, so good. But suppose there was a station at 1790 kHz. Look at the following calculation:

1790 kHz – 1335 kHz = 455 kHz

In other words, the difference between the new station at 1790 kHz and the 1335 kHz oscillator frequency is *also* 455 kHz. This new radio station could also now be heard, though not as well as the one at 880 kHz because the RF tuned circuits largely remove it. But if it were strong enough, it would come through anyway. The 1790 kHz frequency is called the *image frequency.*

Note how the image frequency is calculated:

	Desired station	880 kHz
+	IF frequency	+455 kHz
	Oscillator frequency	1335 kHz
+	IF frequency	+455 kHz
	Image frequency	1790 kHz

That is, the image frequency f_{image} is

$$f_{image} = f_{desired\ station} \pm 2\, f_{IF}$$

(We used the ± sign in the equation because in some radios the oscillator could also be below the desired station frequency; in that case, the image frequency would be below the oscillator frequency, and we would need the minus sign.)

This brings us to a problem — just a few paragraphs ago, we said “In theory, at least, you could get the bandwidth as narrow as you want, simply be going to a lower IF frequency.” But if you do that, then the image frequency gets closer to the desired frequency, and then the RF tuned circuits may not be able to get rid of it. So you have two conflicting requirements:

- To get better selectivity — lower bandwidth — you want to *lower* the IF frequency.
- To get better rejection of the image frequency, you want to *raise* the IF frequency.

This is particularly a problem with high-frequency receivers intended to receive narrow-band signals. For example, consider an amateur FM receiver for 146.94 MHz. Since the bandwidth of FM signals on this frequency is typically only 10 or 15 kHz, a low IF frequency (such as 455 kHz or even less) would be ideal. But then the image would be at

$$146.94\ \text{MHz} + (2 \times 455\ \text{kHz}) = 147.85\ \text{MHz}$$

which is not even 1% away from the desired frequency. There is no way that a typical RF tuned circuit could keep the image out — you’d need a tremendous Q to do it.

Typical receivers solve the problem one of two ways. A few use a much higher IF frequency (around 10 MHz), but with special crystal or ceramic filters which can achieve the narrow bandwidth even at this higher IF frequency.

But a much more common alternative is to use two separate IF sections and double conversion. Fig. 9-10 shows the block diagram of a double-conversion superhet to receive 146.94 MHz. Since 10.7 MHz and 455 kHz IF transformers are fairly inexpensive (they are manufactured by the zillions for use in standard AM and FM broadcast receivers), many communications radios use them as well, and we show them here.

To receive 146.94 MHz, the first oscillator runs at 146.94 minus 10.7 MHz, or 136.24 MHz (the oscillator could be either 10.7 MHz *above* the desired signal, or 10.7 MHz *below*; in this case, we chose to use the lower frequency.) The second oscillator and mixer converts the 10.7 MHz first IF signal to 455 kHz by using an oscillator at 10.7 MHz + 0.455 MHz, or 11.155 MHz.

By using two IF frequencies, the double-conversion receiver solves our two problems. The high first IF frequency does not provide much selectivity, but it helps to eliminate the image. Since the image frequency is at

$$f_{image} = f_{desired\ station} - 2\, f_{IF}$$

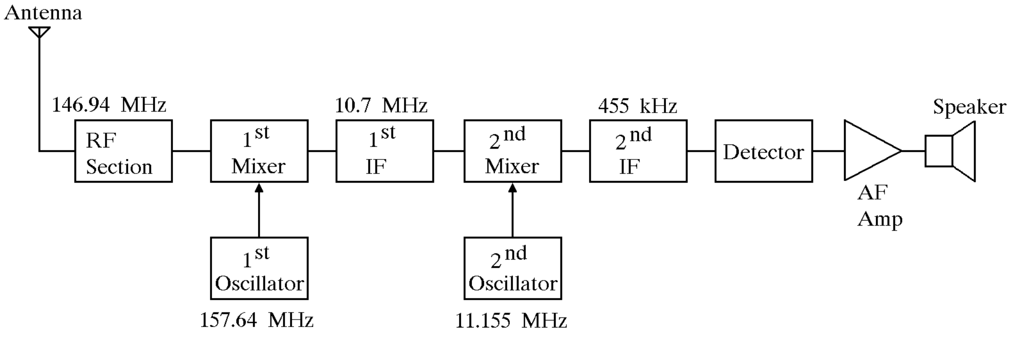

Fig. 9-10. A double-conversion superheterodyne

(note that we use a minus sign since the oscillator is below the desired signal, so the image must be even farther below that), the image frequency is now

$$146.94\text{ MHz} - (2 \times 10.7\text{ MHz}) = 125.54\text{ MHz}$$

which is far enough away from 146.94 that the RF tuned circuits can remove it (or at least significantly reduce it).

The second IF frequency of 455 kHz, on the other hand, is low enough that even transformers with reasonable Q can provide a narrow bandwidth.

Incidentally, suppose we wanted to use a similar circuit to receive 145.015 MHz instead of 146.94 MHz. This circuit would not do, and for an interesting reason: The 11.155 MHz signal from the second oscillator goes into the second mixer, and the mixer is intentionally non-linear (to produce a heterodyne.) Hence it also generates harmonics of all the signals going in. It turns out that the 13th harmonic of 11.155 MHz is exactly 145.015 MHz. Although this harmonic is weak, a slight amount of it will still sneak back into the RF stage, and fool the receiver into thinking there is a weak, unmodulated signal at that frequency. Unless your desired signal is substantially stronger than this false signal (called a *birdie*,) it will not be heard.

The solution in this case is to change the second oscillator frequency from 10.7 *plus* 455 kHz, to 10.7 *minus* 455 kHz, or 10.245 MHz. This new oscillator frequency has harmonics at different places; while this removes the birdie at 145.015 MHz, it introduces birdies elsewhere, such as 143.43 (which is the 14th harmonic of 10.245 MHz.) Designing wide-band receivers (receivers designed to receive a wide range of frequencies) is thus always a problem; there are always some birdies somewhere, and the designer has to carefully choose his oscillator and IF frequencies to try to place the birdies at places where they will not interfere with normal operation.

Summary

If the superheterodyne receiver had never been invented, communications as we know it would probably not exist. The combination of features we have described allows radio receivers to have selectivity and sensitivity which allows millions of transmitters around the world to coexist with each other, yet allows us to select and listen to even extremely weak signals from far away.

We have touched on some of the important concepts, yet have had to skip many others. In the next chapter, we will try to cover some more concepts having to do with the transmitters and receivers which we run into daily.