17.1.3: Amplification

In the diagram on the left, below, we can imagine a small flow of water (labeled Signal) controlling a larger flow of water. This is the general idea of ***electronic amplification***; a large current flow is controlled by a small current or voltage signal. Originally this was done using a ***vacuum tube***. As shown in the second figure below, the tube has a positive end (***anode***) and a negative end (***cathode***) and all air has been removed. A large voltage is applied from anode to cathode so that electrons will stream from the cathode to the anode (just to be clear, recall that conventional current, I in Amperes that we talked about in the last chapter is labeled to flow in the direction opposite to the electron flow). A wire grill called the ***grid*** is placed in between the anode and cathode. If the grid is neutral, the electron flow from cathode to anode passes through the grid and a steady, direct current is produced. If a small signal is applied to the grid it affects the much larger current flow between anode and cathode. So for example, a low voltage sine wave applied to the grid becomes a large current sine wave flowing through the tube. The ratio of the output signal to the input signal is called the ***gain***. This can be measured as voltage gain (ratio of voltage out to voltage in) or current gain (ratio of current out to current in).

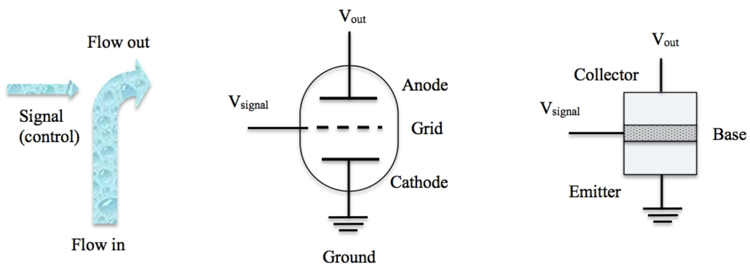

Figure 17.1.3.1

Although they take much more energy to run than modern solid state amplifying circuits, tube amplifiers are still used by the military for radar and high power radio transmission. Some audiophiles claim that tube amplifiers produce a better sound than solid state amplifiers. A disadvantage of tube amplifiers is the tubes take up a lot of space, are heavy, use much more energy than transistors and will stop working if the vacuum seal breaks.

Transistors were developed in the 1940s and gradually replaced most vacuum tubes in electronic applications. Transistors work basically the same way as a vacuum tube. Two crystals (made of silicon or germanium) with slightly different electrical properties are put together separated by a third crystal as shown in the figure on the right, above. The ***emitter*** end of the transistor is made so that it has electrons ready to flow and the ***collector*** end will accept these electrons. When a voltage is applied across the transistor a current flows through the transistor and electrons flow from emitter to collector. The function of the grid in a tube is performed by a crystal called the ***base*** which has slightly different electrical properties from the emitter or collector. When a small signal is applied across the base, the larger current flowing from emitter to collector is controlled to have the same variations as the signal. The terms emitter, collector and base are replaced by source, drain and gate for some other types of transistors using different crystals with different electrical properties.

Transistor amplifiers take much less energy to run than tube amplifiers and generally can take much more physical abuse (shock, temperature changes, etc.) and still work. In theory a transistor should last forever because it is a solid crystal although some transistors can be destroyed by electrostatic sparks. They can also be made incredibly small (today's computer chips contain billions (10^9) of transistors), making cell phones, mp3 players and other small electronic devices possible.

In addition to being used as amplifiers, transistor (and tubes) can also be used as ***electronic switches***. In this application, the signal to the base either allows current to flow from emitter to collector or turns the current completely off. Transistors that function this way can be used to represent ones (all the way on) or zeros (all the way off) in a digital computer.

All amplification suffers from ***distortion***. A perfect recording and playback process would exactly reproduce the original sound but this is never possible; there are always some unwanted modifications to the sound. Microphones, recording devices, speakers and amplification systems all produce some distortion of the signal. The difference between noise and distortion is that noise is extra, unwanted signals whereas distortion is the unwanted modification of the signal.

One form of distortion is when the voltage output is already at the maximum possible for the circuit and the signal amplitude increases. The result is the peaks of the output signal are cut off as shown in the diagram below. The original signal is the blue curve but the amplifier has a maximum output of $0.6 \mathrm{~V}$ so it clips off any part of the signal above $0.6 \mathrm{~V}$ (the red curve). The amplified signal will not sound right because the volume doesn't increase the way it was suppose to in the original recording. This is called ***amplitude distortion***.

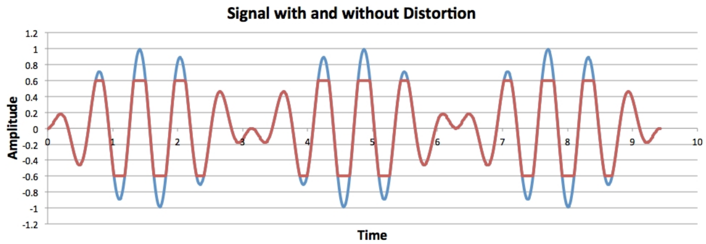

Figure $17.1.3.2$

When there is amplitude distortion the effect of flattening the top of the output wave when it reaches the speaker is to introduce unwanted overtones. We know from Fourier analysis (Chapter 9) that the sudden change from a smooth curve to a flat top requires many extra harmonics to describe the new wave. These new harmonics change the sound produced by the amplifier/speaker system. This type of distortion is called ***harmonic distortion***. Harmonic distortion also occurs if different frequency ranges are amplified by different amounts (in any part of the recording to playback sequence), even if the amplitude is not flattened.

Most amplifiers (and microphones and speakers!) do not treat all frequencies the same. For example, the amplifier may be able to amplify high frequencies better than low frequencies. Or it may not be able to react fast enough to amplify the high frequencies accurately. This is called ***frequency distortion***.

We have already mentioned that microphones do not produce a current with the same phase of the original sound. Amplifiers also can have this problem depending on how they are constructed. If the phase of the output wave is not the same as the input source there is ***phase distortion*** in the output.

Sometimes electronics and software are used to intentionally modify a signal to make it sound different, as mentioned in Chapter 16. One example is adding in reverberation to make the recording sound more natural, as if it were recorded in a concert hall instead of a studio. The following two videos explain in detail one way reverb can be added to a recorded signal to make it sound like it was recorded somewhere other than a studio: Altiverb web site and The 3-D Audio and Applied Acoustics (3D3A) Laboratory at Princeton University web site. Commercial software is also available, for example Room EQ Wizard. Similar techniques can be used to improve the sound of an auditorium if electronic amplification is being used; signals from the microphones can be delayed or modified to provide a better interaction with the room acoustics.
