A2420 SOME FACTS ABOUT DISTORTION AND SPEAKER DAMAGE

By Richard Clark

The subject of speaker damage is always a high interest topic in car audio circles. Because many of us tend to run our speakers close to the edge, the possibility of damage is greatly increased. A good understanding of why speakers actually fail can be of great benefit to designers, installers and operators of audio systems. It is unfortunate that there is so much misinformation about this subject. Because certain aspects of the subject are very complex, it is no wonder that there are so many explanations floating around.

Some of the aspects of speaker damage are rather easy to understand. Mechanical failure due to certain physical related factors is obvious. For instance, poking a hole through a cone with your foot doesn't take much of a study to evaluate. Wrecking a speaker because it moved so far that it literally ripped its suspension and cone apart is also a common mode of failure. Many woofers that are used in poorly designed boxes have met this fate. Tweeters without adequate crossovers have also seen a lot of failures due to over-excursion.

Thermal Failure

Although for this article we don't intend to cover such obvious modes of failure, there is one that we want to cover. This is known as thermal failure. As the word implies this failure is a result of the temperature of the speaker getting too hot. It would seem that the cause of this should also be obvious, however, after talking to many consumers, subscribers and manufacturers, it seems that this failure is not clearly understood.

After hearing some of the explanations for the actual causes, I am no longer surprised that it is so common. It seems that there are a lot of technically unfounded theories as to the cause. We felt it was time to clear up some of these myths!

When we say that a speaker "burned up" we are usually referring to the fact that something in the speaker actually got so hot that it burned up. Just like any other object that burns as a result of getting too hot, certain components in a speaker will actually burn up rendering the speaker useless. These parts are usually the voice coil and the former that it is wound on. Too much electric current flowing through the coil causes the heat. Just as with a hair dryer element or light bulb that glows when current flows through it, the voice coil of a speaker will also get excessively hot if the current exceeds a certain amount. This amount is determined by several factors. The size of the wire can matter. The larger the wire the more current it can handle. The type of insulation on the wire can have a big effect. Some types of insulation are more temperature resistant than others.

Even the shape of the wire can matter. Certain shapes such as flat ribbons can allow more exposure to cooling surfaces. The former that the coil is wound on can be made from anything ranging from paper to metal and can greatly influence thermal limits. The type of glue used to hold the parts together can have different resistance to temperature. Some of the more modern epoxies have the remarkable ability to withstand high temperature.

Regardless of how the speaker is made, there is always a limit to the power it can safely handle without damage. It is usually in defining this power limit that we have a lot of confusion. When we rate an electrical device we use the term watts. Some electrical devices are easy to rate. Take for example a light bulb. When we rate a light bulb at 60 watts, we know that 60 watts of power will flow continuously through that bulb as long as it is turned on. If we increase the power by a significant amount, the bulb will burn up because its filament will get too hot and melt. If, we reduce the voltage to the bulb then the filament will cool off and the light will become dim. As long as the voltage and current remain within the design parameters of the bulb, it will work as intended. With a constant source of power it is easy to rate the power capacity of a component.

Speaker Power Rating

If we were to rate a speaker the way we do a light bulb, it would be easy to arrive at a value but that number would be useless considering the way we use our speakers. Unlike AC line voltage, music signals are constantly changing. If we look at a typical music signal over a short period of time we will see that there are peaks that are several times as large as the average level. When we consider the power of such a complex signal, the actual values of the waveform must be integrated over a given time period. It is the total power under the curve that determines the power, therefore the heating, of the signal.

For an example of this examine the following charts. The measurements were taken from the speaker terminals of a power amplifier. Each chart is a display of power for a period of one minute so each vertical division represents 15 seconds.

Figure 1 shows a constant power level of a 1 kHz sine wave. For the entire one-minute measurement period the level shows no variation. This continuous signal is similar to the 60 Hz that a light bulb would see. Figure 2 shows the levels that are typical of music. The levels can be seen to vary over a range of about 20 dB. The interesting fact about Figures 1 and 2 is that when the music is averaged over the period of one minute, it represents exactly the same power level as the continuous sine wave. The highest power point in the music selection is near the start of the song and is nearly 10 dB louder than the average level. This high point is in contrast to the lowest point that occurs about 20 seconds into the song. The lowest point is about one tenth the average power (about 5 watts).

If the continuous sine wave of Figure 1 were used as a reference and played at a power level of 50 watts, it is easy to see the heating effect that would occur in the voice coil of a speaker. What is not always understood is that if the signal of Figure 2 were played into that same speaker, the heating effect would be the same even though there were momentary peaks of 500 watts. If our speaker were rated at a truly honest 50 watts continuous, neither of these signals would cause any thermal damage even though the music had peaks that reached 500 watts. It may cause excursion damage but that is another failure mode.

Thermal Mass

Because the duration of those 500-watt peaks was very short, the voice coil would not have time to actually heat up. To understand why such large power peaks do not damage the coil we have to think of the term thermal mass. Even though the power is high the coil can absorb a certain amount of heat before its temperature starts to rise to unsafe limits. This principal is easily demonstrated with a candle. It is easy and painless to pass a finger through the flame of a candle as long as it is done quickly. This doesn't mean the candle isn't hot. It just shows that the thermal mass of our finger is able to absorb a certain amount of heat before we start to get burned. As long as we give our finger time to cool between passes we can do this repeatedly.

Now for an extreme example of power variation. Figure 3 is a power vs. time plot of a snare drum beat reproduced by a very large amplifier. The average power of this signal was also the same as the signals in Figures 1 and 2. Some of the beats are 17 dB louder than the average level. Seventeen dB above 50 watts is over 2000 watts! If our 50-watt speaker were driven with this signal, it is very likely that it would not be damaged due to the cooling period between the peaks. Even though the momentary peaks are very high, the average power level, and therefore the average heating effect, is only 50 watts.

It is this very nature of music that makes rating the power of speakers so difficult. If our sample speaker could handle 50 continuous watts, we could rate it at that power level. If we were going to play only tones with our system then it would be simple to properly match it to a 50-watt amplifier. If we were going to play music with our system, a quick look at Figure 2 will show us that a 50-watt amplifier would probably not be adequate for our speaker.

The speaker was easily able to handle 500-watt music peaks. If the speaker were matched to a 50-watt amplifier and we tried to play the music of Figure 2 at an average power level of 50 watts, we would find that the amplifier would be greatly undersized for our 50-watt speaker. In such a case we would find that the amplifier would be very distorted during most of the music. Anytime the signals consisted of peaks that exceeded 50 watts, they would be clipped by the amplifier.

Distortion

This brings us to an interesting, but often misunderstood, subject. ---- Distortion. Let's consider the fact that all music is composed of sine waves. The combinations of frequencies, amplitudes, and phase relationships are virtually infinite. To keep the explanation simple, let's first consider a single sine wave. Most amplifiers have no trouble passing a sine wave with virtually no distortion. They can do this at all levels up to their maximum rating. The power supply rails usually determine the maximum rating. The rails of an amplifier are the outputs of the power supply of the amplifier and are a DC potential. In modern amplifiers there is usually a positive and negative rail. When the amp is outputting a sine wave, the energy is diverted from these rails by the output transistors that act as variable valves.

As the sine wave swings positive, the energy is diverted from the positive rail and when it goes negative, the energy comes from the negative rail. As long as the voltage required for the sine wave is less than the power supply rail, the amplifier will have no trouble reproducing the signal without distortion. When the signal required is higher than the rail voltage, then the transistor acting as a variable valve is turned on completely and the rail of the amplifier is literally connected to the speaker. This full "on" condition is known as saturation and it occurs for whatever duration of the wave that exceeds the rail of the amp. See Figure 4 for a simplified block diagram of an amplifier and the relative relationship of signals to rail voltages.

Amplifier Clipping

This saturation condition is more commonly referred to by the name that describes what it does to the signal. We call it "clipping." Perhaps no other facet of amplifier behavior is more misunderstood than this condition. Let's cover some of the myths. First it is commonly believed that clipping is bad for an amp. Nothing could be farther from the truth. The transistors in an amplifier operate in a linear mode. This means that for a given current, the power dissipated by the transistor is proportional to the voltage dropped across its junction. When the amp is clipping the transistors are fully saturated and have virtually no excessive drop across their junctions. This places them in their most efficient condition. True digital amplifiers operate in this condition all the time and this is why they are so efficient.

Pure DC?

It is also commonly believed that when an amplifier clips that it puts out DC. A quick look may make this appear to be true, but it would have to be a very quick look. Let's take the example of a clipped 100 Hz sine wave. If we examine it for more than 1/200'" of a second, we will see that the signal inverts polarity at the same periodic rate as an unclipped sine wave of the same frequency.

The waveform we are describing is known as a square wave. But wait a minute here; a square wave is not the same as DC. A pure square wave is a complex AC signal that is composed of a fundamental frequency and an infinite number of harmonics. Figure 5 shows an oscilloscope photo of a good square wave. If we examine the makeup of this waveform, we will find that it is composed of many individual sine waves. The first of these is the fundamental. We can observe the fundamental frequency by passing our square wave through a very steep low pass filter. This will remove all of the harmonics and leave only the fundamental. The fundamental is a pure sine wave of the same frequency as that of the square wave and can be seen in Figure 6.

If we move our filter up to three times the fundamental frequency, then we can observe the next harmonic along with the fundamental. This would be known as the third harmonic and can be seen in Figure 7. Moving the filter even higher will reveal more harmonics.

Figure 8 shows the fundamental with the third and fifth harmonics. By the time we add the fifth harmonic the signal starts to resemble the familiar square wave. If we continue to add harmonics, the wave will eventually become more and more filled until it becomes almost square. If you still think clipping is DC and all this seems a little too confusing, we have a little experiment you can try that will convince you that a clipping amplifier doesn't put out DC.

An Experiment into Distortion

All it should take to clear up the DC question is to put a large stiffening cap in series with a speaker. Then hook your speaker to your car battery. No DC will flow because DC doesn't flow through a cap. Now hook the speaker to your amp with the cap in series and turn up your amp until it starts to clip. The clipped signal will go right through the cap. We all know that caps don't pass DC, right? So now that we are on the right track let's get back to the facts. When an amplifier clips a signal, it causes the generation of harmonics. These harmonics are also known by another name. This is what we call distortion. The more an amplifier is over-driven, the more the signal is clipped and the more distortion is generated. No matter how much an amplifier is clipped, it will never produce a perfect square wave because the generation of a pure square wave requires infinite bandwidth. In fact perfect square waves only exist in theory because there is no circuit with the bandwidth necessary to pass one -however, for all practical purposes some of them come pretty close.

For the interests relating to loudspeaker power capacity, it is only necessary that we concern ourselves with the very basics of square waves. We have already seen the frequency content of a square wave. When a square wave is compared to a sine wave of the same peak amplitude, the square wave will have twice the energy of the sine wave. This is because there is more area under the curve but not all of that extra energy is from the harmonics. If we examine the actual power contained in the individual frequencies that make up the square wave we can see this more clearly.

Figure 9 shows the relative power content of a square wave. The power levels relative to the fundamental decrease with each ascending harmonic. If we add the total energy of the fundamental sine wave with the harmonics we will get a total of 100% of the energy. What is important when we are considering the heating effect on a speaker is that the fundamental frequency of a square wave, which is a sine wave, will contain 2 dB more energy than a sine wave with the same peak voltage as the square wave. Put another way, the fundamental of a square wave has a higher peak (27% more voltage, 62% more power) than the square wave itself.

More Power?

Analysis of this subject can get rather complex but what it really boils down to is that almost any amplifier can produce twice its undistorted output if it is allowed to clip excessively. This extra energy results in about 2 dB more energy at the fundamental and one dB in harmonic content.

It also is commonly believed that there is a difference between square waves that are recorded in program material (such as might be found in heavy metal music) and square waves that occur as a result of amplifier clipping. This is simply not true at all and an understanding of just what makes a square wave should make this clear. By the time we have reached the third overtone (also called the 7" harmonic) the power level is diminished to only a couple of percent of the fundamental. A whopping 81 % of the total power is still contained in the fundamental.

If we are concerned about power dissipation in speakers it is easy to see that we need not concern ourselves with harmonics much beyond this point. How the signal is generated is not important at all. The only thing that matters are how much energy the waveform contains. This is the only factor that determines the heating effect on the voice coil of a speaker. It has been suggested that clipping can destroy tweeters. This is seldom the case. If only 19% of the energy is contained in the upper harmonics, then this doesn't really leave us with much destructive energy.

Take the example of the case of a passive three-way system with crossover points at 100 Hz and 2000 Hz. Suppose we clipped a 50 Hz signal going to the woofer with a 100-watt amp. We know that the amp can produce 200 watts of square wave power. Looking at Figure 9 we can calculate that we would have 162 watts (undistorted) of 50 Hz, 18 watts of 150 Hz, 7 watts of 250 Hz, 3 watts of 350 Hz, and 2 watts of 450 Hz. By the time we reached the frequency range of the tweeter, the power would be fractions of a watt. If our system had a tweeter with a power rating of only 10 watts, it is unlikely that adding another fraction of a watt would make much difference.

These harmonic levels represent the absolute worst-case conditions for a single frequency and do not account for the fact that program material contains complex frequencies, all of which clip simultaneously if the amp is overloaded. But the amount of overdrive needed to clip an amplifier to the level demonstrated is well in excess of 30 dB. For an amplifier with an input sensitivity of 2 volts, this would require an input voltage to the amplifier of over 50 volts! This type of drive voltage is virtually unavailable in car audio products.

Of course the same effect could be achieved by feeding the signal to the amplifier in an already clipped state, something that happens frequently. At any rate these excessive conditions don't really apply here as the dynamic changes in music make it very difficult to continuously clip a signal to such an extreme state. Even when an amplifier is overdriven by as much as 20 dB (10 times the rated input voltage), clipping of the waveform will occur only for about 40% of the total time on music. At this point the music is so distorted that it is virtually unlistenable, at least for most listeners.

For a demonstration of the actual spectral difference between normal program material and terribly clipped material see Figures 10 and 11. Although the spectrum would be different for different types of music these two curves should be comparable as they were the same piece of music.

Figure 10 is the spectral content of a song averaged over a period of 40 seconds with 243 samples each. Figure 11 is the same song but it is massively clipped. The level was reduced so the curves would represent the same overall level. Notice that the overall spectral content is very similar to the original. These tracks are from our new amplifier level setting CD#104 and are Tracks 31 and 35 respectively. The total energy represented by both tracks is exactly the same and will produce virtually the same heating effect in the voice coil of a speaker. The speaker doesn't care if the music is distorted or not. To a speaker it is all just a combination of sine waves. A speaker cannot tell the difference between noise, distortion, or music. It doesn't care what kind of music it is or any thing else about it except how much energy is contained in the signal.

What Really Matters

The real problem for our speaker is that any amplifier can produce more than its rated power if it is allowed to distort and clipping is a sign that you have probably reached the point where you should turn things down. To get a real picture of what the speaker sees compare Figures 12 and 13.

These represent the power vs. time of a song with very steady energy. They are in stark contrast to the song in Figure 2 at the beginning of our article. Figure 12 is the output of an amplifier driven right to the verge of clipping but not quite. The amplifier was a MTX 4160 that is capable of producing about 219 watts (AM certified) per channel into 4 ohms. Even though the peak power to the speaker is over 200 watts the average power is only about 20 watts.

Figure 13 is the same song and the same amplifier, but the gain has been increased until the amp is clipping very severely. Notice that the dynamics (ratio between loud and quiet passages) have been compressed and that the average power level has increased by about 15 dB. Now the average power level is about 100 watts and the peak power is about 400 watts. This means that our speaker is getting five times the heating effect that it was receiving before we started clipping the amp.

Prevention

It is sometimes claimed that a good way to prevent speaker damage is to use a larger amp. This is true only if you don't turn the larger amp up until you exceed the average power to the speaker. A larger amp will allow for the larger transient peaks to be reproduced without damage to the speaker because the total heating of those short peaks is minimal anyway. This will almost always result in a cleaner sounding system. But if you continue to increase the power to the speaker until the average power is excessive, then it doesn't matter if the signal is clean or distorted ---- the speaker will overheat and burn up. Usually when this is done the speaker will begin to distort before the amplifier. If this happens this is a sure sign that you are feeding the speaker too much power.

When considering speaker power ratings and amplifier size it should never be forgotten that although amplifiers are usually rated with continuous sine waves, speakers rarely are. When matching amplifiers and speaker power ratings you are literally comparing apples and oranges. This should not normally be a problem as long as we remember how the ratings differ.

The most efficient woofers rarely exceed even 2%. If we put 100 watts of power into a speaker, this means less than 2% of the power is turned into sound and over 98% of the power is turned into heat. The next time you need to be reminded of how much heat 100 watts is just grab a glowing 100-watt light bulb. And just think, not all of the energy if alight bulb is turned into heat, in fact they are more efficient than most speakers!

If your speaker were fed 100 watts continuously it would have to dissipate this much heat from its coil. Not many speakers are really designed to handle 100-watt continuous sine waves, but when rated for dynamic music that is constantly changing (providing cooling periods), a speaker that can only handle perhaps 50 watts continuous sine waves can easily deal with the undistorted output of a 400-watt amp. That's because the average output of that 400-watt amp reproducing undistorted music is probably not over 50 watts.

If you can keep things under control you will be safe with the 400-watt amp. The same speaker that can handle 50 watts of continuous sine waves can probably deal with 100 watts of a totally distorted amp.

In nearly all cases when we see power ratings on a speaker, the manufacturer has taken into account that you are going to be playing music on that speaker. That is why we see ratings on some small tweeters as high as 100 or even 200 watts. This usually means that if the amp is about the same size and not allowed to distort, then the average power will be about right.

It is not that distortion hurts speakers; it is just that distortion is usually a sign that something is too loud. This is also why we can find examples of 300, 500, and even 1000-watt woofers. You can be sure these are not ratings for continuous sine waves, continuous music maybe, but not sine waves. Next time you think you have a 1000-watt woofer, think about the heat produced by your hair dryer and imagine that much heat building up in a small box in the back of your car ---- better get a fire extinguisher!

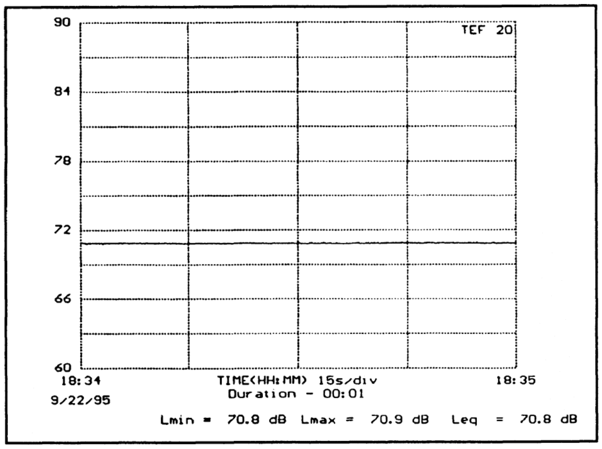

Figure 1: Constant power of a continuous sine wave

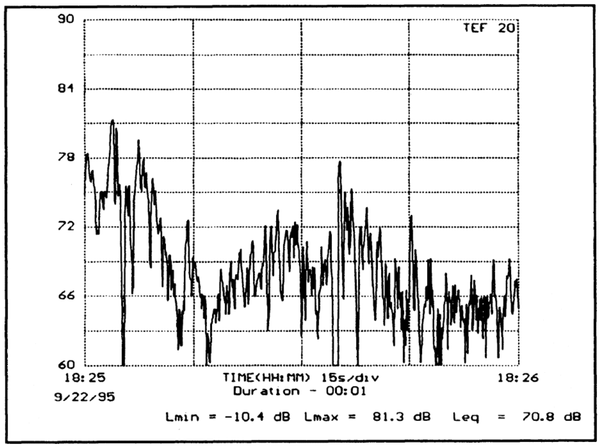

Figure 2: Typical dynamics of music program

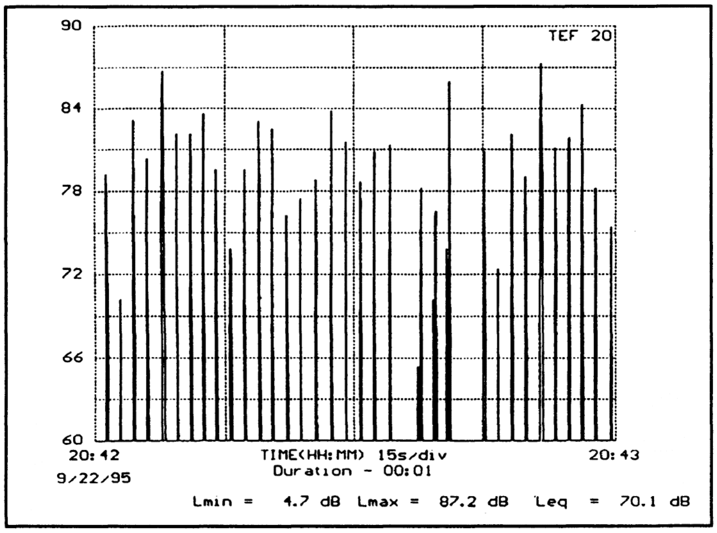

Figure 3: Playback of multiple snare beats

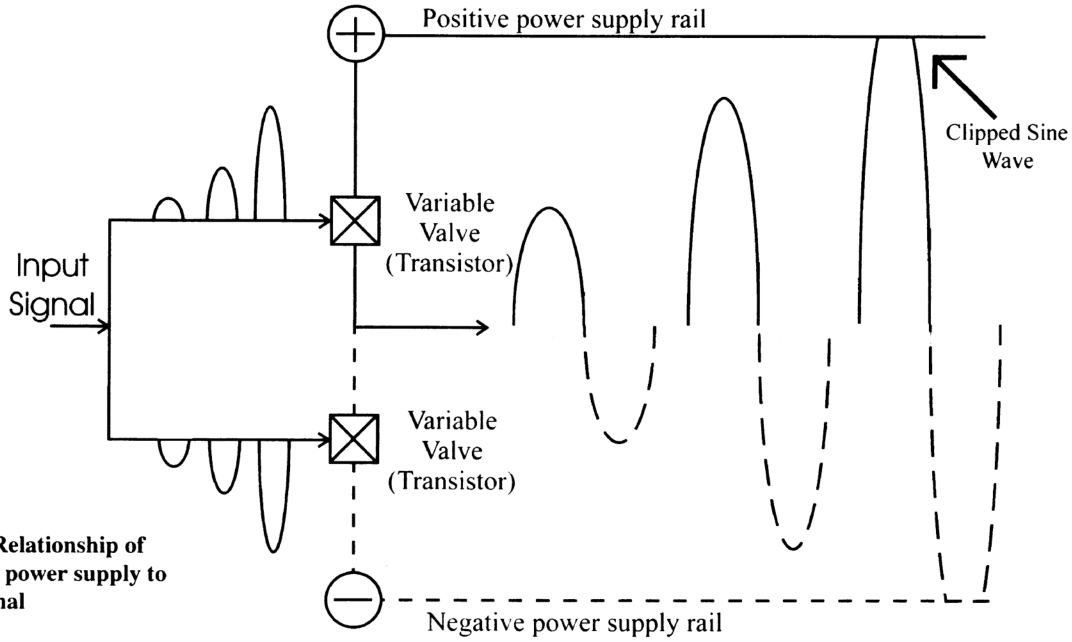

Figure 4: Relationship of amplifier / power supply to output signal

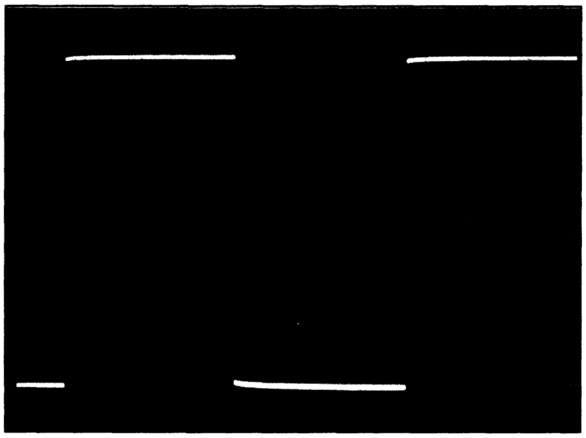

Figure 5: Sine wave with many harmonics resulting in a square wave

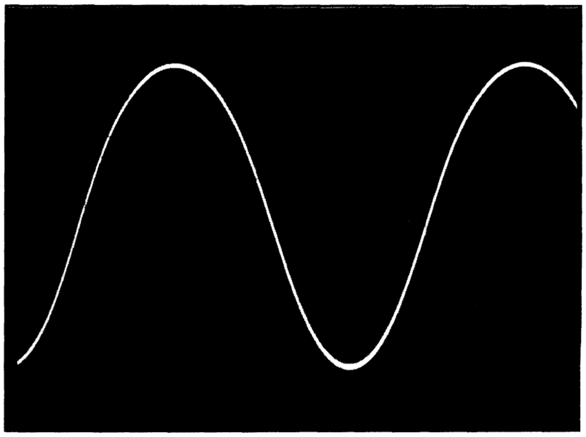

Figure 6: Fundamental frequency of a square wave

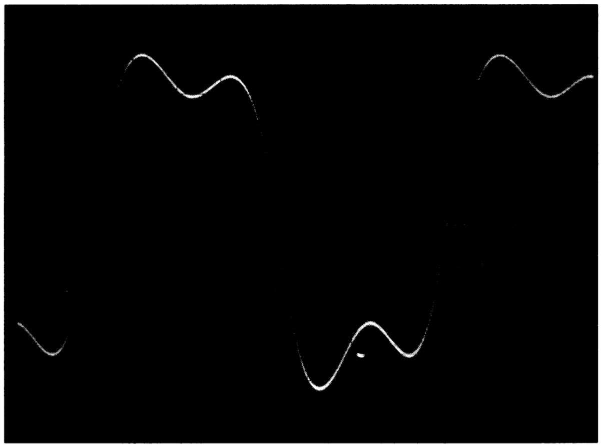

Figure 7: Fundamental and 3rd harmonic

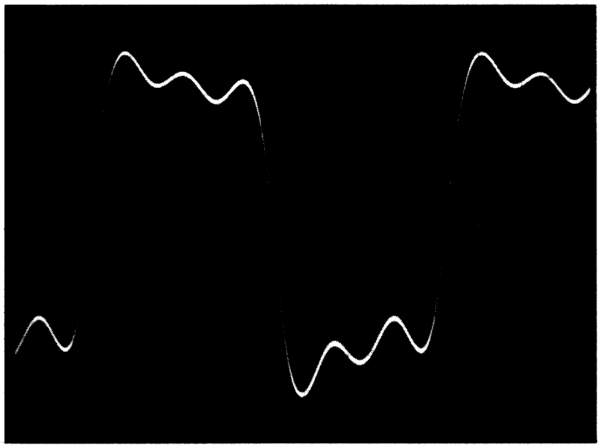

Figure 8: Fundamental and 3rd and 5th harmonic

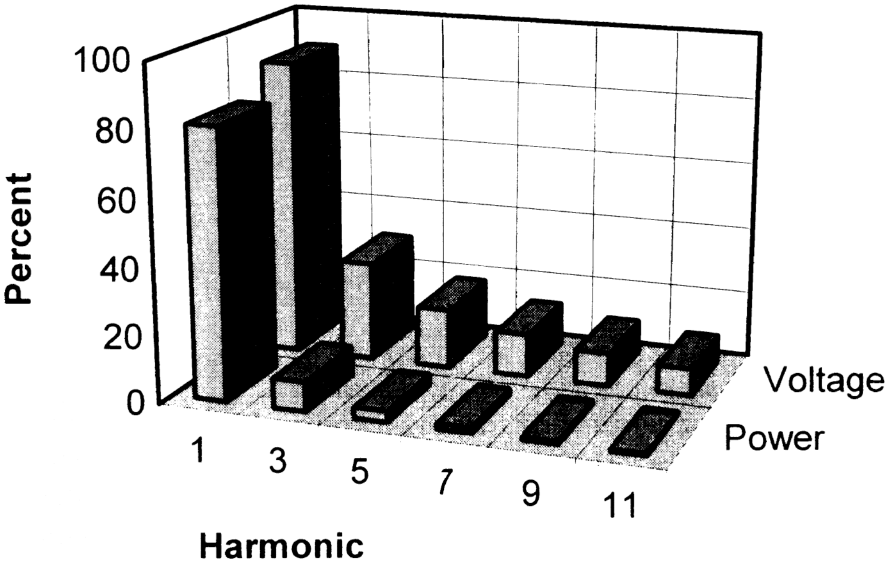

Figure 9: Relationship of Fundamental to harmonics of a square wave

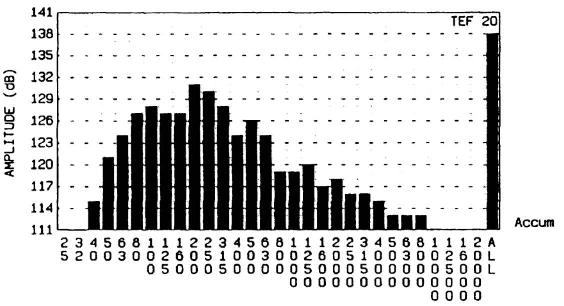

Figure 10: Spectral content of an undistorted music sample

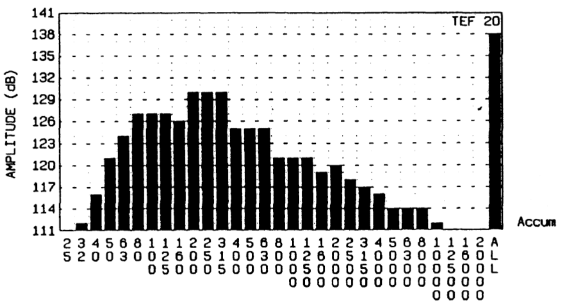

Figure 11: Spectral content of same music sample as Figure 10 -- totally clipped

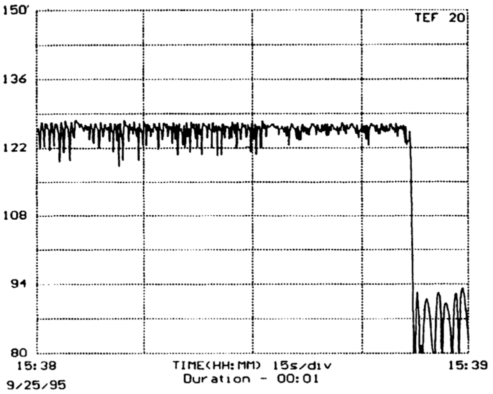

Figure 13: Compressed dynamics and increased average power level of distorted music

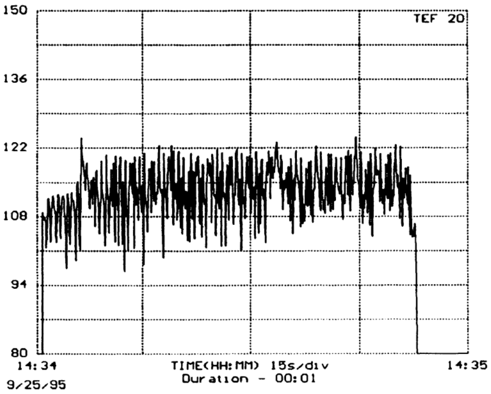

Figure 12: Dynamics of normal music sample