CHAPTER 1 SIGNALS AND SYSTEMS

System
composite makeup of components to perform a particular function.

Signal
a set of data representing some underlying meaning.

Home Theater (HT) is founded in two different but surprisingly similar technical areas—audio and video (AV). Both of these technologies have a common underlying technical basis and because these technical concepts are common to both areas, I have placed this material first. Therefore, I am going to start this text with the toughest material. I do this not because I enjoy torturing the reader from the outset, but because this material is fundamental and common to all the technology that makes up HT. Both audio and video are **systems** that manipulate **signals**. To be sure, the signals and systems in each of these areas are different, but the underlying concepts for dealing with them are virtually identical.

This means that I can cover the fundamentals of the two subjects with a single technical discussion—a tremendous savings of effort. The unfortunate part is that I must discuss these concepts generically, without specific reference to their applications since they can be applied in different situations in different ways. Admittedly, this creates a double-edged sword of conciseness versus elucidation. I hope that the reader will make the effort to follow the discussions since doing so will pay big dividends in later chapters when more specific applications are discussed.

1.1 Signals

Value
a numeric attribute of a physical quantity denoting its amount.

Signals are present all around us. A signal is any variation in a **value** (a numeric attribute of some physical quantity that denotes its amount, such as volts or pressure) over time which contains useful information. Voice is a signal, as is a doorbell, as is video, etc. Noise is not generally thought of as containing information, although there can be exceptions (such as when your car makes a “noise” that signals something is wrong). Thus, there are basically two classifications of time variations—signals and noise. One contains information—usually the desired information—and the other does not. Note that what constitutes a signal and what constitutes noise does not have an absolute distinction—one person’s signal can be another’s noise.

The concept of a signal is useful since it allows us to have a discussion of a data stream without the need to reference the underlying physical system that carries this data stream. For instance, we can talk about a signal and its characteristics without having to reference the actual physical quantity that we are talking about, such as the pressure or volts, etc. or even the intended end product, i.e. sound or video. We can let these signals flow from one system to the next with little concern about the details of the underlying physical system they are currently in. For example, when someone is talking on the phone, the acoustic voice signal is converted into an electrical current signal which travels along cables. This is often done as an optical digital signal. The signal then arrives at its destination ready to be converted back along an identical but reverse process into a sound signal at the receiver—human hearing. At every location the “signal” should be the same, but it will take on different physical forms (electricity, light, sound, even analog or digital) at every stage.

We hope, or assume, that the signal content remains constant along the entire path, but it never does. I will talk in later chapters about how a signal can be changed as it progresses along a transmission path, and I will show how the modifications to these signals are created and defined. How these modifications can change the human perception of these signals and the important question of when these modifications actually cause a change in the perception or not will also be discussed. Since all signals are changed by the systems that propagate them, knowing what changes affect human perception and what changes do not is a crucial distinction. I will attempt to define those signal changes which have been shown to be the principal contributors to deviations in perception. However, I will not be able to discuss all of the contributors down to the most inconsequential ones. The result of this limitation will be that I may not discuss aspects of systems and signals which are felt by the reader to be important. That does not mean that I don't agree to their existence or perhaps their importance, only that I may not agree with their priority.

1.1.a Signal Level

Scale
applying a number to the data value.

Even though a signal can exist without a reference to a specific physical quantity, we will need a way to define its **value** at any given moment in time, i.e. a way to **scale** it. Scaling is straightforward if we simply use the definition of the physical quantity carrying the signal, voltage in Volts or sound pressure in Pascals for instance, but it will turn out that another scaling method is far more convenient for AV signals. Since the concept of a signal usually implies that it has information content, it also implies a human perceptual interpretation. When human perception is involved a different scaling definition is usually desirable.

The reason for this different scaling is because human perception mechanisms for signals in nature, light, sound, and even touch or smell to a lesser extent, tend to respond equally to ratios of excitation rather than the actual level of excitation. For example, subjects will judge each doubling of the sound pressure as being a perceptually equal increment. This means that going from 1 unit to 2 units is perceived as the same perceptual change as going from 10 units to 20 units, even though the actual physical increments are one and ten respectively. There are biological reasons why humans react in this way, but that is a topic beyond the scope of my intended subject. Relationships which have this characteristic are defined mathematically as logarithmic. A more detailed description of the **logarithm** can be found in Appendix II and this appendix also discusses some important characteristics of the dB scale, which I am going to introduce next.

Logarithm
a mathematical relationship wherein equal percentage changes have equal numbers.

A method for scaling a logarithmic relationship, which has found almost universal application, is called the deci-Bel (dB, a tenth of a Bel, after Alexander Graham Bell). The dB is the most common unit of measure in both audio and video. The dB scale gives a signal scaling which is more in line with human perception. It is also a scaling method which has a number of useful features, most of which are discussed in the appendix.

1.1.b The Time-Frequency Relationship.

Frequency
the number of cycle, per second, that a waveform exhibits.

Pitch
the perceived tone of a sine waveform.

Amplitude
the value of a waveform at some point in time.

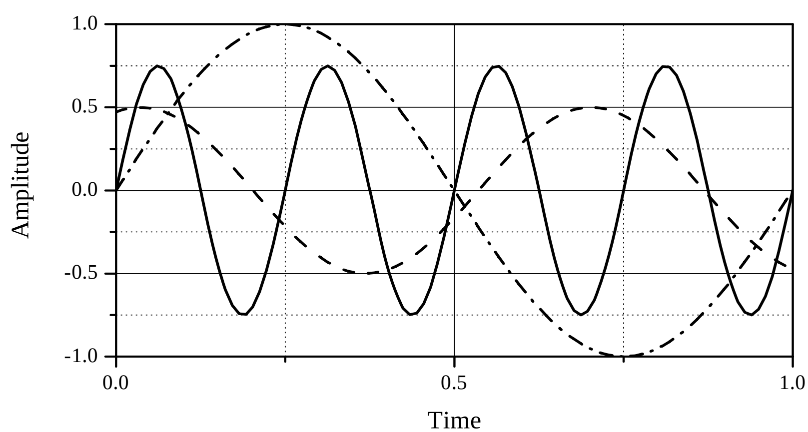
Hertz
the unit of frequency—one cycle per second.

Sine
a convenient trigonometric function for defining the simplest of waveforms
(See App. III).

One of the most important concepts in any discussion of signals is the relationship between the time and frequency domains. (In video, it is the space and frequency domains that we are interested in, but I think that the time domain is more familiar so I will focus on that one for the moment.) The time domain is something that we all have a basic understanding of even though it is hard to actually define. We typically have a good conceptual understanding of **frequency** because our daily lives contain all kinds of sounds and we can readily distinguish a high frequency sound from a low frequency one. The problem with this description is that what we perceive as frequency is actually called **pitch**, which is different from frequency. I don't want to get too deep into this difference other than to note that our common daily experience with "frequency" is really a perceptual experience with pitch. Once again it is basically a logarithmic relationship that connects the two.

Frequency is defined as the number of repetitions that the signal level undergoes in a given period of time. Figure 1-1 shows three **sine** waves of different frequencies and **amplitude** in the time domain, that is, time as the horizontal (x) axis. These waves have frequencies of 1 cycle per second or **Hz** (**Hertz** after the German physicist Henrich Hertz), 1.5 Hz and 4 Hz. If

Figure 1-1. Three sine waves of different amplitude, phase and frequency.



the time axis were in milliseconds (ms, 1/1000 of a second) instead of seconds then these waves would be 1000 Hz (or 1 kHz), 1.5 kHz and 4 kHz respectively. These higher frequencies are more common to us than the very low ones, so I prefer to show time in milliseconds since this time scale lends itself better to a more common every day experience.

Frequency domain
the technique of studying signals and systems by looking at their frequency characteristics.

Consider an alternate way of showing the data in Figure 1-1. Let me call this new way of looking at the data the **frequency domain**. In Figure 1-1, each of these waveforms can be described by an amplitude and a frequency—two simple numbers. If I now plot these numbers on a new graph with amplitude as one axis and frequency as another, then I will get a plot as shown in Figure 1-2—which is called the frequency domain. Note that time is no longer directly apparent in the frequency domain just as frequency is not directly apparent in Figure 1-1. The two plots do, however, contain exactly the same information.

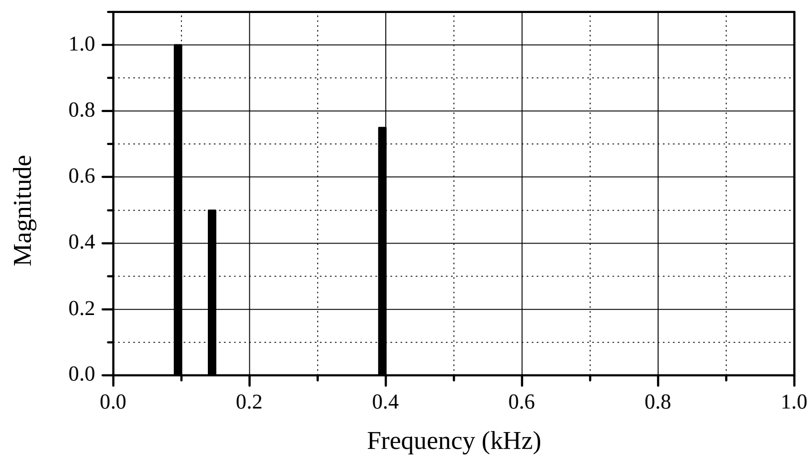


Figure 1-2. Frequency view of the three waves shown in Figure 1

Magnitude
the peak value attained by a sine wave independent of when it occurs.

RMS
the square root of the mean (or average) value of a waveform.

In the frequency domain, each waveform is located at its specific frequency and is drawn as a line whose height indicates the waves amplitude. The vertical axis is labeled as the **magnitude**, which is the peak value of the waveform independent of the point in time at which this occurs. The magnitude and the amplitude are slightly different things where the term amplitude is usually used to refer to the value of the waveform at any given time in the time domain and the term magnitude is used to refer to the waveforms maximum value. Sometimes, we might see magnitudes given as **RMS values** (Root Mean Squared). RMS is basically an effective value, a sort of average value for the waveform. The term magnitude is usually used to refer to frequency domain levels.

Phase
the starting location of a sine wave relative to its zero crossing or another sine wave.

In Figure 1-2, the 1.5kHz wave does not start at zero as the other two do, this wave has been shifted along the time axis. The starting point of the waveform is called its **phase**, completely analogous to the "phase of the moon". The time delay that causes a given phase depends on the frequency of the waveform. The important thing to note is that to completely describe the time domain data in the frequency domain, I also need to know its phase. Thus in the frequency domain the waveform is represented by its frequency and its magnitude and phase or equivalently its **complex magnitude**. It is complex because the magnitude can be described by two numbers known as the real and imaginary parts of a complex number.

Complex magnitude
the magnitude and phase of a waveform express as a two part—complex—number.

It's not really too important to note this "complex" aspect of the magnitude, but it is described in more detail in Appendix III - Complex numbers. The reason that I even acknowledge it is because mathematically the calculations are all done in complex arithmetic. Fortunately, I will almost never need to resort to this complication. I will plot magnitudes as a single real number and sometimes show the phase—the phase being of lessor (but not insignificant) importance. If it is desirable to show the phase, then it is usually done as a second plot or a second line on the same plot but with a different scaling. Note that the magnitudes as shown in Figure 1-2 are blind to the waveforms starting point, i.e. the phase.

Complex waveforms
waveforms composed of a multitude of sine waves of different frequencies.

The relationship between the time and frequency aspects of a sine wave can be extended to **complex waveforms**. Consider the waveform shown in the top half of Figure 1-3. This waveform might be exhibited, for example, by a musical instrument since it is periodic. With a period of 1 ms it would have a base frequency of 1000Hz, or 1kHz. The period of a repetitive signal is the length of time that passes before the signal exactly repeats itself. By using a mathematical technique called the **Fourier Series**, I can decompose the complex upper waveform into a set of pure sine waves. I have shown this decomposition in the bottom half of the figure. This series—it is called a series because it is the sum of a series of independent waves—is named after the French mathematician "Fourier" who first studied its properties. In the case shown here, all of the sine wave are related by integral frequencies. The longest waveform in the series is called the **fundamental** and the higher frequency (shorter) ones are called **harmonics**. The integer relationship is important because it means that only waves which are *n* times the fundamental, where *n* is an integer, are allowed in this series. This requirement is a direct result of the fact that the Fourier Series repeats itself, it is always periodic. Its values continuously repeat in any time interval that is outside of the time **period** of the fundamental—1.0 ms in Figure 1-3. Any signal represented by a Fourier Series must repeat itself on exactly this period and so only waves which have frequencies that are integer multiples of the lowest frequency and, hence, synchronous to it are allowed.

Period
the time for a periodic waveform to repeat itself.

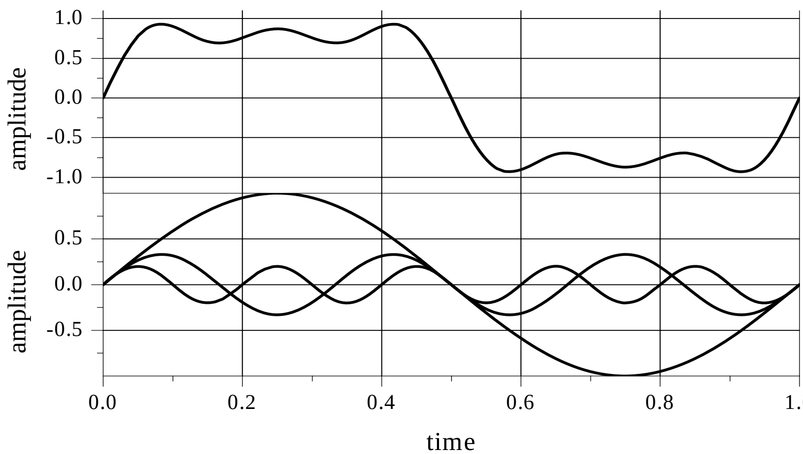


*Figure 1-3.
A complex waveform and its constituent parts.*

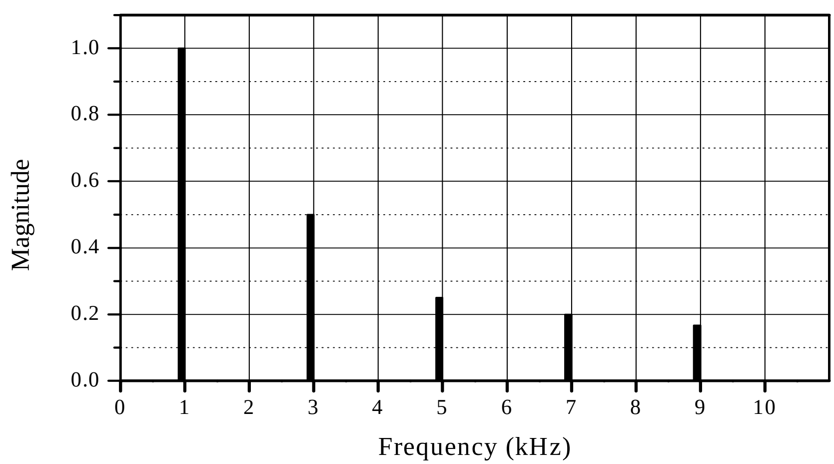


*Figure 1-4.
Frequency domain representation of a square wave.*

Figure 1-3 also has a frequency domain representation as shown in Figure 1-4. In this figure, I have plotted the components by their order (the integer of their multiple) which is of course proportional to their frequency. There is no need to plot the phase of the components since it is obvious that they are all in phase. I have shown the first to the fifth order for the Fourier Series decomposition of a square wave. Figure 1-3 only shows the first three components of this series, but it should be clear that adding more terms would lead to a more squared off waveform.

The reader should take note of the close relationship between the discussion of Figure 1-1 and Figure 1-2 and the Fourier Series component representation shown in Figure 1-3 and Figure 1-4. They are basically the same thing. The Fourier Series shows us how to map signals from the time domain to the frequency domain in a concise mathematical framework.

The Fourier Series is ideal for decomposing harmonic waveforms as would appear from a single instrument playing a continuous periodic tone, but this is hardly the most general form of a signal. The world is composed of harmonic, in-harmonic and transient signals as well as music, which is made up of a multitude of individual instruments with all of these signals being present simultaneously. In-harmonic signals come from instruments like cymbals and drums which have waveforms which do not have an integral relationship between their components and almost all musical instruments that have both transient and steady state signal components. In order to be able to decompose a completely general (real life) waveform, I will need to extend the Fourier Series concept to a close relative which allows signals which are not necessarily periodic.

Window
the limited time frame in which we look at a waveform.

Consider an impulse waveform as I have shown in Figure 1-5. For now, let's simply ignore what this waveform looks like outside of the "**window**" that we are currently looking at ($0 - .3$s) and go ahead and find its Fourier series components. The result of this exercise is shown in the top of Figure 1-6. Note that the components are still all harmonics, but of the lowest frequency component, defined by the window, at 4Hz The principle waveform period is seen to be at 24 Hz (the peak value) which corresponds to a period of about .04 s.

Consider the **time base** of Figure 1-5, the length of time shown in the plot, the window, which I will also assume is the length of time over which I take the data. In this example it is .3s. If I let the time base become much

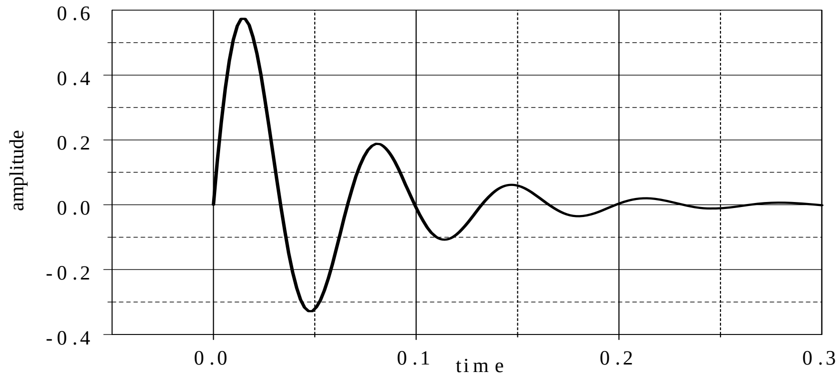


Figure 1-5.
A non steady state signal—an impulse.

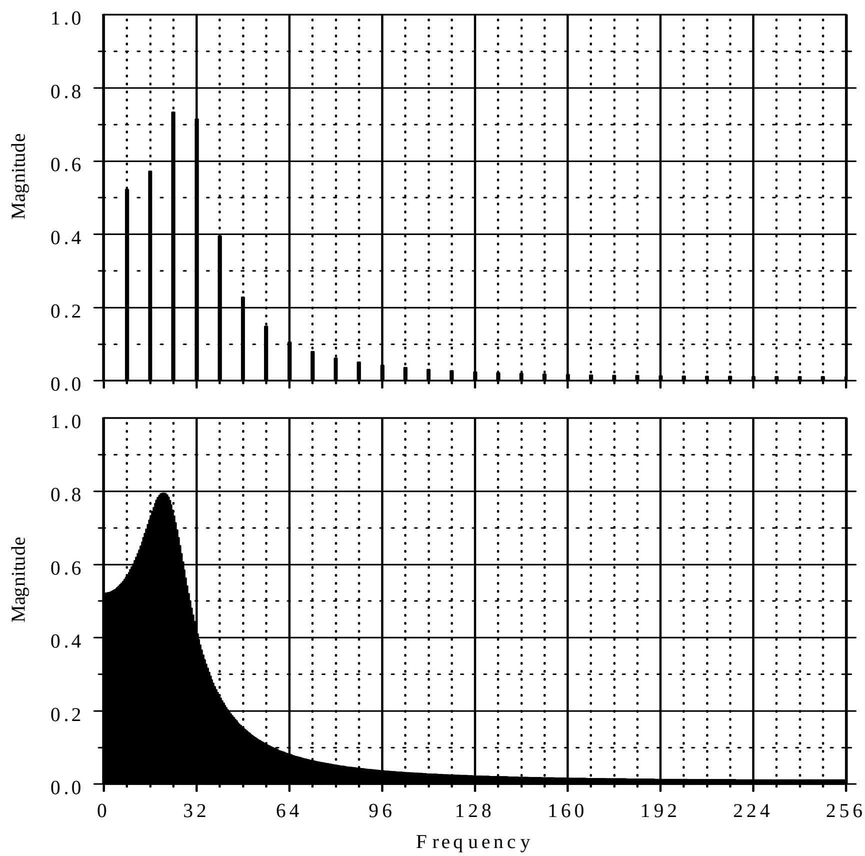


Figure 1-6.
The Fourier Series terms in the expansion of Figure 1-5 for two different time windows.

larger, say 8 times longer or about 2.5s, then I will get the results shown on the bottom of Figure 1-6. Both curves are plotted with discrete lines to represent the frequency components, but in the lower plot these lines have become very dense. The lines are still all harmonics of the lowest period, but this period has become much longer and the lowest frequency is now 4/8 = .5Hz. The peak value is still at 24Hz however. The results have not changed, only the resolution of those results has increased.

If I were to let the period in Figure 1-5 go to infinity, then the discreet line structure of the representations in the above figures would become infinitely dense and would create what is called a continuum—a continuous curve not discreet lines. In this later case we usually drop the filled-in area under the curve and draw a line from point to point. The continuous version of the Fourier Series—the one where the time base goes to infinity—is called the **Fourier Transform**. It is called a transform since it transforms data from one continuous domain—time—into another continuous domain—frequency. For our immediate purposes, we will limit our discussions to transformations between the time domain and the frequency domain although the Fourier Transform (as I will show later when I talk about video) has a much broader applicability and holds for transformations between many different kinds of variables (domains) and also in multiple dimensions such as space.

Fourier Transform
the continuous frequency version of the Fourier series where the time window becomes very long.

Applying the Fourier Transform to a time domain signal results in its frequency **spectrum**—basically a frequency domain plot like that shown in Figure 1-6. A spectrum is defined by two values at every frequency. The two values can be given in two equivalent sets, as **real** and **imaginary** values (a **complex number**) or the far more common terms, magnitude and phase. I discuss these quantities and the relationship between them in Appendix III - Complex numbers.

Spectrum
the continuous frequency representation of a signal.

In audio it is common to view the spectrum in dB with a logarithmic frequency scale (making it a log-log plot). This plotting standard is useful because the picture created closely represents what we would actually hear (perceive)—the pitch being approximately logarithmic and the level scaling in dB. Appendix II - Logarithms shows some of the important relationships that apply in the log-log domain. Exponential relationships appear as simple straight lines in this type of plot which can be a significant simplification in visualization and calculations.

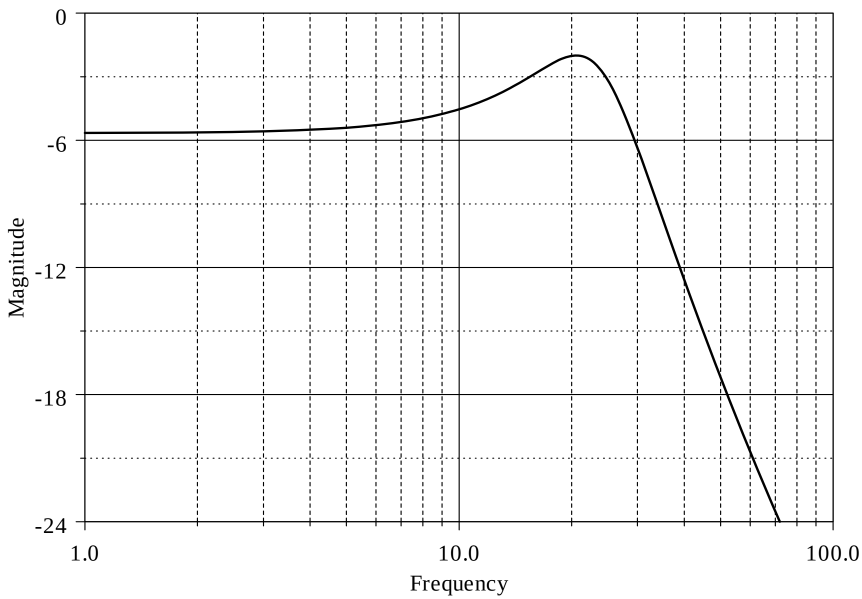


Figure 1-7.
The dB-log spectrum of impulsive signal in Figure 1-5

Figure 1-7 shows the same data as shown in Figure 1-6 but using the more conventional dB-log scale. Figure 1-7 is exactly the kind of graph that I will be showing often in this text—the spectrum of a signal (the signal in this example being shown in Figure 1-5). In the next section, I will describe how if this spectrum is found at the output of a system when the input was a unit **impulse**, then it is known as the **impulse response** or equivalently the frequency response. The frequency response of a system being, perhaps, the single most important characteristic of an audio component.

Impulse
a very large amplitude but very short time duration signal where the total area is one.

Impulse response
the time domain signal that would be seen at the output of a system if there was an impulse at the input.

FFT
an efficient computer algorithm for performing the Fourier Transform.

The examples that I have shown in this chapter were all generated on a computer using what is know as the **Fast Fourier Transform** or **FFT**. The FFT is a computer algorithm that aims to numerically approximate the Fourier Transform—a mathematical construct that I described earlier with an unrealistic time window of infinity. If properly implemented, the FFT will yield a close approximation of the actual Fourier Transform that we desire. Usually, any errors that can occur with the use of the FFT approximation are benign, but sometimes these errors can actually obscure what we are trying to see. In Appendix IV - Acoustic Measurements, I discuss some of these issues in more detail. For our purposes here, however, we can think of the FFT as simply being the “real world” implementation of the Fourier Transform.

At this point I would like to review the basic principles discussed in this section because they are key to the reader’s understanding of almost everything that I will do in this text. These principles are described below.

- *A time domain signal has a completely equivalent representation in the frequency domain.*
- *The rules for the moving between the two descriptions in the time and frequency domains come directly from the requirements of the Fourier Series and the Fourier Transform.*
- *The FFT is a computer algorithm that attempts to implement the Fourier Transform numerically. It creates a good approximation of the true spectrum, but not without some error.*

1.2 Systems

System
a fundamental assembly block which processes input signals to create output signals.

Time invariant
a system whose properties do not vary in time.

Linear
a system whose input/output transfer characteristic is a straight line.

Now that we know how to decompose a complex signal (one that is composed of multiple tones) into a composite of simpler single frequency waves and how to scale these signals in a meaningful way, we can move on to a discussion of systems. A **system** is a symbolic block which receives data as an **input signal** and acts on that data to produce an **output signal**. It is important to limit our discussion to several restricted sub-classifications of systems. The first is that the system under consideration is **time invariant**— that is the system does not change over time, and the second assumption is that the system is **linear**.

The time invariant requirement is self-evident and simply means that the properties of the system do not change with time, or at least not significantly during the time frame that we are looking at them. If the system were not time invariant, then any statement made about it would be invalid only moments later. Linear means that when a complex signal passes through a system, each component in its spectrum is acted on individually, i.e. that the individual components do not interact with one another. This non-interacting characteristic is also called superposition, the principal that two signals can be superimposed on one another without affecting each other. For instance, when the 100 Hz component of an input signal passes through a linear system, its output level depends only on the input level of the 100 Hz component. It is not affected by any other components of the input signal, i.e. 200 Hz, 101 Hz, etc. This is an extremely important restriction, for without linearity, virtually all of the system theories that I use in this text fail to be valid or applicable.

Dynamic range
the range of signal levels over which the system can operate effectively.

The problem is that no system can be linear for any arbitrary signal level at its input or output. A system is said to have a **dynamic range**—i.e. the limits on the range of signal levels over which it is linear. The lower limit of this range is almost always the noise floor, which in some systems (like digital ones) is often a linearity issue. On the other end, at some signal level the system will saturate and cease to be able to correctly output signals or accept larger input signals. Good audio systems have a dynamic range of 90 dB or more (or a ratio of highest level to lowest level of about 30,000). Achieving this in electronics is relatively straightforward, but, as we will see, for acoustic signals in real rooms this is a major challenge. Despite the inevitability of all systems becoming **nonlinear** (not linear), it is convenient to assume linearity and move on from there. Later, I will step back and take a deeper look at the implications of the nonlinearity in typical audio system components. Linearity also comes up in the context of video signals, but it does not play as central a role in video as it does in audio.

Convolution
a mathematical technique for calculating an output signal from an input signal in the time domain.

What one usually wants to know about a system is how it acts on an input signal. This can be done in the time domain, but believe me, if math intimidates you, you wouldn't want to do it this way. The process involves a mathematical technique known as **convolution**, which is all I am going to say about it. Fortunately analyzing a system in the frequency domain, especially in dB, is relatively straightforward.

If we assume that a system is linear, then individual sine wave components in the input spectrum can be treated independent of each other. This allows a system's response to a complex waveform to be determined by simply evaluating the response of each of the individual input components exclusive of any of the other components. The components are then recombined at the output to form the output signal. This feature of a linear system is the principal motivation for moving into the frequency domain since no such simplification occurs in the time domain.

Gain
the ratio of the signal level seen at the output and the input.

Each input component is affected by the **gain** and phase of the system at that frequency. The gain is the ratio of the magnitude of the output signal to the input signal and the phase is the difference in the phase of the signal at the output relative to the phase at the input. In dB terms the gain is the difference, in dB between the input and the output (because division of two numbers is the same as the difference in their log values see Appendix II - Logarithms)

Transfer function
the gain and phase of a system versus frequency.

A plot of the gain and phase factor for all frequencies (of interest) is called a **transfer function**, $T(f)$ which is also called the frequency response of the system. The transfer function is defined as the ratio of a systems output spectrum $B(f)$ to its input spectrum $A(f)$ as

$$T(f) = \frac{B(f)}{A(f)}$$

If, in this equation the input signal were to contain all frequencies (of interest) at a unit amplitude, i.e. $A(f)=1$ then $T(f)=B(f)$—i.e. the transfer function $T(f)$ would simply be the spectrum seen at the output, $B(f)$. For example, if Figure 1-7 was the spectrum seen at the output of a system when the system had a flat spectrum $A(f)=1$ at its input, then this curve would be a plot of the systems transfer function—its frequency response. It is a classic example of a low pass filter with unity gain at low frequencies.

From the above example we can see that it would be useful to have at our disposal signals which simultaneously contain all frequencies, because placing these signals at the input to a system yields a spectrum at its output that is its frequency response. There are actually several signals with this feature. I have shown two of them in Figure 1-8. It has always amazed me that white noise (random noise with a flat spectrum) and a unit impulse have exactly the same flat magnitude spectrum. The only difference in these two signals is the phase, which is random for the white noise and a straight line for the impulse. This occurrence is a classic example of why we should always consider the phase, because looking only at the amplitude spectrum these two dramatically different signals would be indistinguishable. The term **white noise** comes from white light, which contains the full spectrum of colors, equally represented. White noise contains the full spectrum of sound of uniform amplitude. **Pink noise** has a slight emphasis on the low frequencies—the color red—hence the color pink.

White noise
a random signal with a flat spectrum.

Pink Noise
a random signal with a spectrum that falls at 3dB per octave.

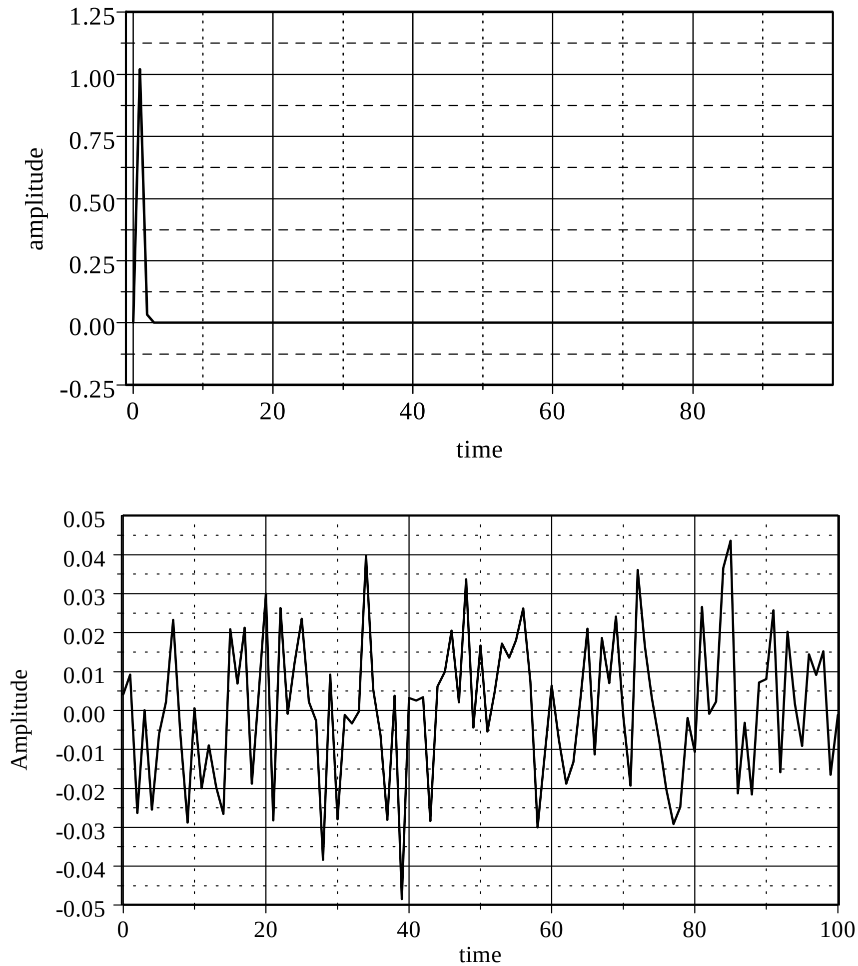


*Figure 1-8.
Two different signals with identical amplitude spectrums.*

A third signal with a nearly flat spectrum is a swept sine wave. The details of using swept sine waves is too complex to get into here, but they are one of the most common and useful of all of the measurement signals and techniques.

Figure 1-8 shows an important characteristic of noise versus an impulse: the instantaneous level of the impulse is very high since all of the energy is located at a single point in time while the noise spreads the energy uni-

Power spectrum the frequency distribution of the power—no phase consideration.

formly across time. The impulse has a very large level for a very short period of time while the level of the noise is low for a long period of time. Both signals have the same total energy and **power spectrum** (the spectrum of the signals independent of its phase), but the peak value of the impulse is about fifty times greater than the noise. If I send either of these two signals through a linear time invariant system, the output spectrum will be the frequency response of that system. If the system is nonlinear, then it will react quite differently to these two signals which is why the linearity assumption is so important.

The response of a system to an impulse signal is logically called its impulse response. The impulse response, in the frequency domain, is the system response to all frequencies, which is its transfer function. Thus, the impulse response completely defines the linear operation of the system. If we know the impulse response of a system then we can calculate its effect on any signal that we feed this system. Interestingly, using the response to the white noise signal can also give us the impulse response, although not directly. Since the noise is random we need to do a statistical correlation of the output to the input and to get a good estimate of the spectrum, we need to take many averages. Characteristics which are not true of the impulse. The impulse, however, is likely to overload the system if not handled properly. As I mentioned, there is also the simple swept sine wave technique which can also give us the frequency response of a system. In practice, each of these methods has pros and cons as actual test protocols. The thing to remember is that if one wants to describe a system, any system, then the most important characteristic to consider is its impulse response—regardless of what signal is actually used to obtain this response. And, if done correctly, they should all yield exactly the same answer.

1.2.a Filters

Filters are one of the more common types of system manipulations that occur. This subject in its entirety is enormous and I will only touch on some of the basics.

Generally, systems do not pass all frequencies equally, whether desirable or not. Often they have a low frequency limit of usefulness (they don't always pass DC signals) and always have some high frequency limit. Systems can also be made to intentionally pass specific sets of frequencies and

Filter
a system which passes some signal frequencies while blocking others.

to reject others. Systems with these characteristics are called **filters**. Filters come in many shapes and sizes which can take many descriptors (parameters) to describe.

The first characteristic of importance to a filter is its type. A filter that passes only high frequencies is called a **high pass** filter, one that passes low frequencies a **low pass** filter and one that blocks both high and low frequencies is called a **bandpass** filter. There is also a **band reject** filter which passes both highs and lows but cuts out certain frequencies in the middle. It should be apparent that only high pass and low pass filters are unique— bandpass filters are made up of cascaded high and low pass filters while band reject filters are made up of parallel high and low pass filters. In a bandpass filter the low pass filter is set higher than the high pass filter and in a band reject, the opposite is true.

Order
the order of a filter denotes the steepness of its slopes.

Cutoff frequency
the frequency where a filters response has dropped to half power, or –3 dB.

The next most important characteristic of a filter is its **order**. The simplest filter—first order—can be described simply by its **cutoff frequency**, the frequency where the response is down by 3dB, or one half power. Higher order filters have sharper slopes in their cutoff region. The order of a filter is determined by the number of simple filter stages in its makeup. Each simple filter stage achieves a 6dB/octave slope which can be either positive or negative depending on the filter type. Equivalently, it has either an f or $1/f$ slope in the linear (non-dB) domain. This can be understood by recalling that the slope of a line in a log-log plot (see Appendix II - Logarithms) is the same as the power of its variable. Since sequential filters (systems) add their responses in the dB domain (multiplication in the linear domain) the filter slope increases by 6dB/octave (or a power of f) for each filter section. Thus a three section cascaded low pass filter (order 3) would have a cutoff slope of –18dB/octave (or $1/f^3$).

Filter Q
the shape of the filter near its cutoff frequency.

Two stage (order two) filters are described by a well-known set of descriptors; cutoff and **Q,** both of which are defined in the frequency domain. Cutoff specifies the location of the filter's slope change form its passband to its stopband and Q defines the shape of this transition. Higher order filters are often referred to by names, like **Bessel** or **Butterworth** or **Linkwitz-Riley** since these higher order filters would otherwise take too many parameters to describe. I would like to point out, however, that "named" filters are nothing more than specific cases of the general filter form, which is actually much more flexible. These named filters have usu-

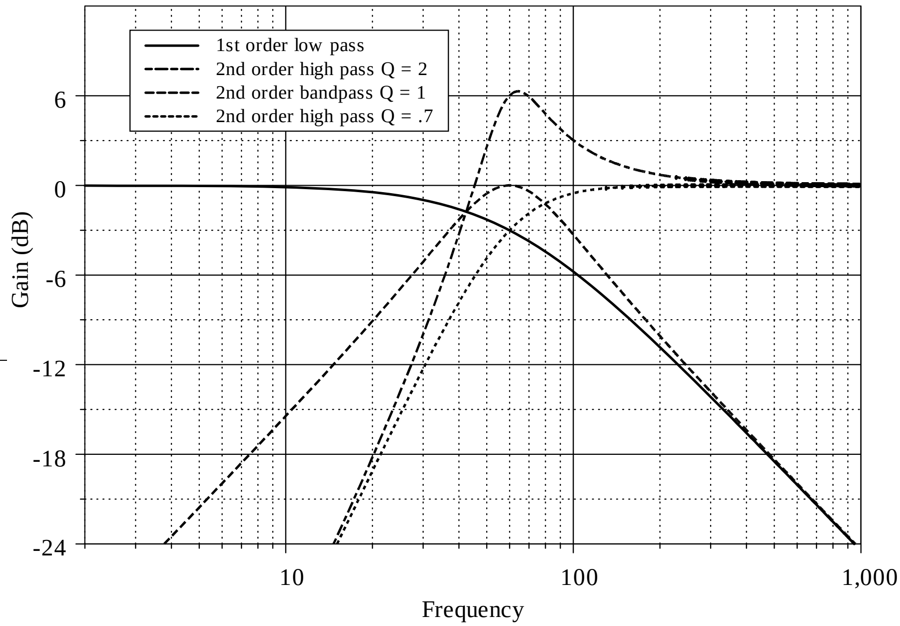


Figure 1-9. Filter function with various orders and shapes.

ally been developed with a particular goal in mind when their unique set of parameters is determined. They may have the minimum delay or be maximally flat, but otherwise there is nothing magical about them.

Some typical examples of filters are shown in Figure 1-9. These filters all have the same mathematical cutoff frequency (60 Hz) which shows that the higher order high Q filters do not actually have the –3 dB point at their theoretical cutoff, although if Q = .7 then the cutoff will be at –3 dB. Bandpass filters are defined by the sum of the high and low pass orders and are centered on the cutoff frequency with a bandwidth defined by their two –3 dB points (about 60 Hz in the above example).

1.2.b Amplifiers

Probably the most common system that we encounter after the filter is the simple amplifier. There is an enormous mystique about amplifiers, types (tube or solid state), class, etc. I think that there is a lot of evidence that amplifiers can sound different, but there is also a lot of evidence to say that any good amplifier is certainly adequate.

As a system there are usually four parameters to consider, which place the limits on its capabilities two different dimensions. The first dimension is frequency and the limits are its upper and lower –3 dB response points. As I said, no system can have an infinite bandwidth and so one needs to know what these frequency limits are. For most high quality audio amps these values are virtually always sufficient, say 20 Hz – 20 kHz, or more.

The other two parameters frame the upper and lower signal levels that can be utilized in this system component—the dynamic range. The upper level is the most commonly referenced one and is usually expressed in Watts, which unfortunately, is not a very useful number. What we really want to know is at what voltage the amplifier will clip the output signal or sound objectionable. The clipping voltage is usually fixed, while the Wattage capability depends heavily on the speaker's impedance, which is complex—literally. Always remember that clipping is a significant sound quality problem. One should never allow an amp to clip in practice for any signal as this will usually result in poor sound quality.

The final parameter is, I believe, the most important and that is the noise floor. One has to be careful here since by noise floor I mean two different phenomena. The first is the actual random (thermal) noise that appears at the output which is usually specified as a ratio of the maximum signal level to the noise level as a dB number. But, there is a second criteria, which is far more difficult to determine, and that is how well the amp handles very small signals. These small signals may be above the noise floor, but still small enough that they are affected by zero crossing errors. This type of problem can usually be seen as a rising distortion value for small signal levels. Reject any amplifier with a rising distortion for small signals as these types of distortion can be shown to be highly objectionable.

Simply put, the amplifier is, in my opinion, not the place to spend an inordinate amount of money. Any well designed amplifier will relegate this piece of equipment to the status of "not the weakest link". For example, cutting the expenditure on speakers to buy expensive amplifiers is just plain foolish. You'll only end up with a lot more power to drive your speakers into even more distortion. If you have lots of money and love the looks of a cool amp (and let's face it, this is one of their biggest appeals) then go right ahead. But please don't waste money on an amp if that forces you to cut costs somewhere else.

1.3 Video Signals and Systems

Video signals and the systems that process these signals are completely different than audio signal and yet they are amazingly similar. The first obvious difference is that a video image is a two dimension signal. However, the electronics that process these signals are basically one dimensional systems like those that I have been talking about. It is possible to do two dimensional signal processing directly using lenses or complex electronic array systems, but these are not so easy to understand and their usage is not very common.

Spatial frequency
a 2D concept where the light variations in space are analyzed as a frequency distribution.

The key concept that one must understand is that of **spatial frequency**. Once the reader has a grasp of this concept they will find that they already have the necessary background to understand image processing.Figure 1-10 shows a spatial frequency in the *x* direction of three. Compare this figure

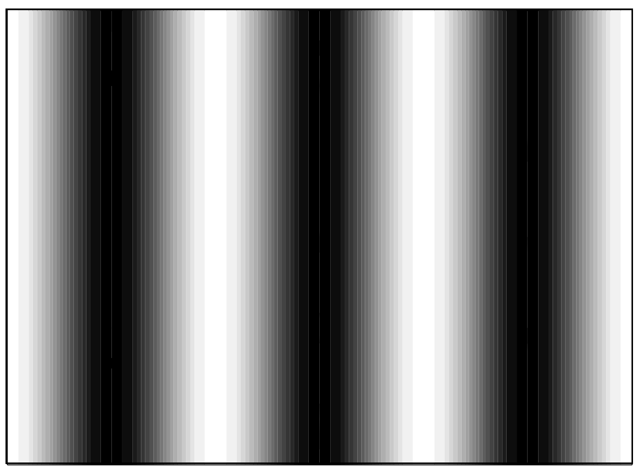


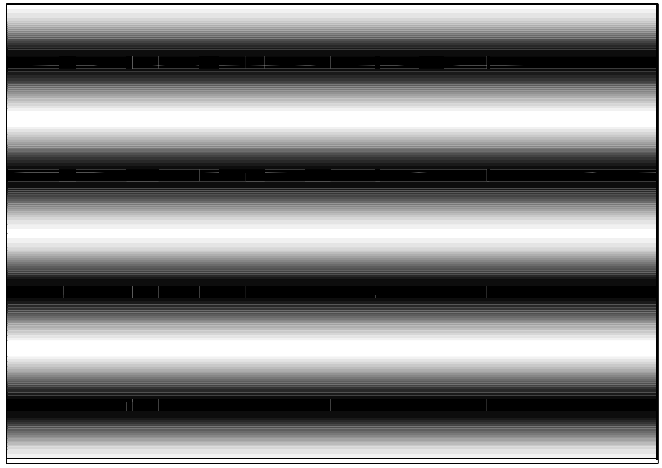
*Figure 1-10.
A spatial frequency of 3 in the x direction, there is no y variation.*

with Figure 1-1 for the time domain. While the two figures do look different (the first figure has three waveforms), they are basically the same thing along any horizontal line in Figure 1-10. Note that there are exactly three repeats of the color (black) bars. In a video signal, on a computer for example, the values in this figure would range from 0 to 255 (for a 8 bit number in gray scale).

Wavenumber
the number of wavelengths in a given length.

Figure 1-11 shows a spatial frequency of four in the *y* direction. It is common in optics to use the **wavenumber** *k* and denote the *x* and *y* compo-

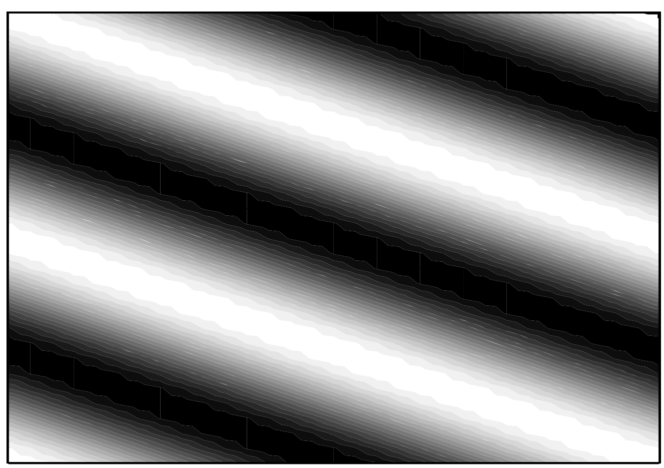
Figure 1-11.
A spatial frequency of 4 in the Y direction.



nents by k_x and k_y. The wavenumber is the number of wavelengths in a given distance. If the distance is defined to be the image width and the image height then the wavenumber are the same as the frequency.

A combination of the two wavenumbers is also possible and one such image is shown in Figure 1-12. Note that along the *x*-axis there is only a single wave and that along the *y*-axis there are two. Along the diagonal there are three

Figure 1-12.
An image with a spatial frequency $k_x = 1$ and $k_y = 2$. The total spatial frequency is then $k = k_x + k_y = 3$



Once one has grasped the concept of spatial frequency, it should come as no surprise that the most important tool for dealing with images is the two-dimensional Fourier Transform. This transform takes any image into the spatial frequency domain and in fact can be operated on by the same sorts of filters that we have already been talking about. This is because the two-dimension FFT is nothing more than a bunch of one-dimensional ones all strung together as a single steam of data. To be more accurate for image processing one usually uses the a cosine transform which is nothing more than a simplification of the Fourier Transform which makes the processing a bit faster.

I will return to this topic in Chapter 7 when I talk more about image processing.

1.3.a Conclusion

This has been a long and challenging chapter, but the principles that I have discussed here will be used time and time again throughout this text. If on first reading the material is unclear, then perhaps a second reading is in order. I am always amazed when people think that if they don't understand something the first time they read it then they won't ever get it. From my experience this is not at all the case. There have been many texts that I have had to read twice to understand and I know of certain sections in texts that I have read four or five times to truly comprehend. There is no doubt that technical subjects can be difficult to follow, but diligence always pays big dividends in the end.