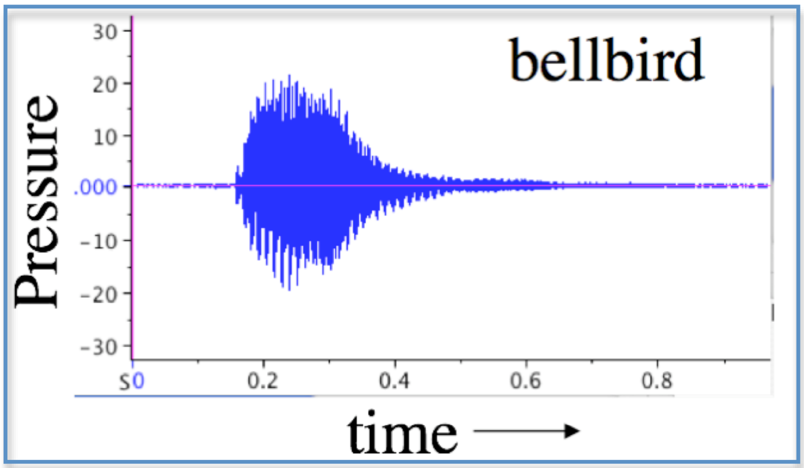
A Primer on Analyzing Animal Sounds

Jack Bradbury-The Macaulay Library

Basic Principles

A. Waveforms: Tape recorders track the rise and fall of pressure as sound waves pass the recording microphone. A plot of how the pressure of a sound changes over time is called its **waveform**. Here is an example:



B. Time Domain vs Frequency Domain Descriptions of a Sound: The simplest waveform is a pure sine wave of constant amplitude. Any other waveform can be decomposed into a set of pure sine waves which when combined with the proper amplitudes and relative phasing would recreate the original sound. The original waveform is called the **time domain description** of the sound. The set of frequencies composing that waveform and their relative amplitudes provide the **frequency domain description** of the sound. Usually, it is necessary to examine both descriptions to fully understand a given sound.

1. **Frequency measures**: The frequency of a sine wave is measured in the number of complete cycles per second (a unit called the Hertz).

2. **Amplitude measures**: Amplitudes of sine waves are measured as the difference in pressure (or the voltage generated by pressure in a microphone) between the highest and lowest values (peak-peak), or as a geometrical average of deviations from ambient during the wave (called **rms** measures). One rarely gives the absolute value of amplitude because it is the relative amplitudes of different frequency components in a sound that are biologically important. Accordingly, amplitude is usually given as the log of the ratio between the sound of interest and some reference amplitude. The decibel (abbreviated dB) is 20 $\log_{10}$ (amplitude of sound of interest/amplitude of reference sound). The reference is usually SPL, the softest sound that a human can hear.

C. Fourier Analysis: This is the mathematical process by which we decompose a complex waveform into its frequency components. Modern computers use an accelerated method for doing this called a **Fast Fourier Transform** (**FFT**). The results of this analysis may be displayed as a:

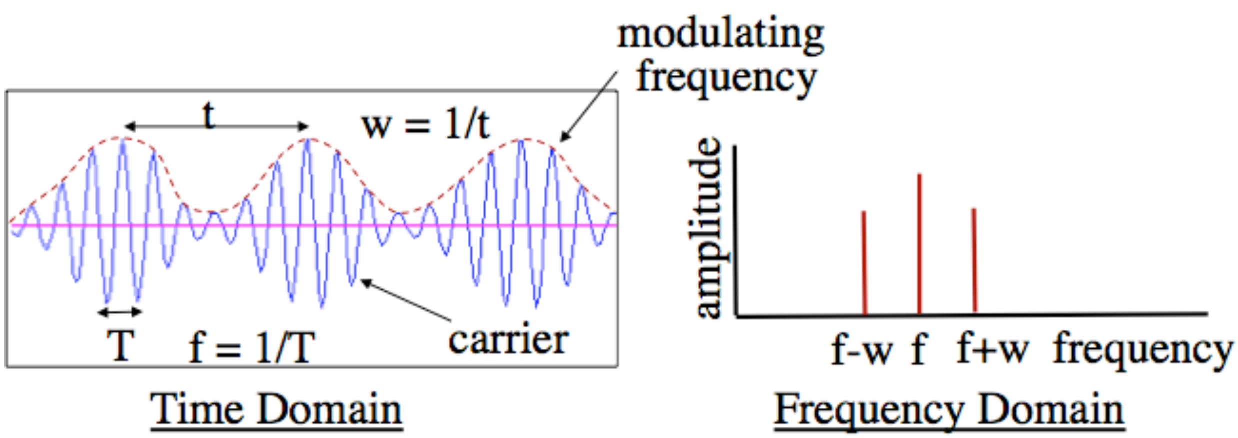
1. **Power Spectrum**: This is a plot of amplitude on the vertical axis and frequency on the horizontal axis that shows how much energy is present at each possible frequency within a segment of waveform.

2. **Spectrogram:** This is a plot of frequency on the vertical axis and time on the horizontal axis that shows loud frequency components as dark regions and faint frequency components as light regions. It can be thought of as a sequence of power spectra lined up along the time axis.

Fourier Guide

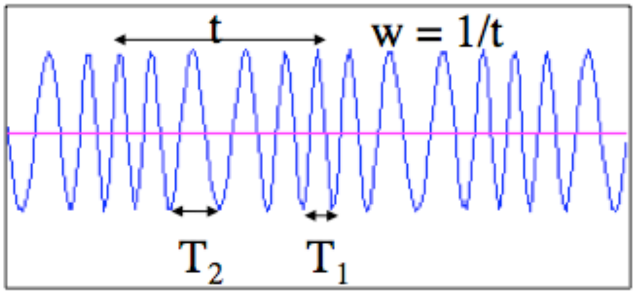
A. Fundamental Models: There are three fundamental types of sound waveforms. Nearly any sound you will encounter will have a waveform composed of one of these types or a combination of them. Learning the Fourier rules for each type thus prepares you for interpreting nearly any spectrogram or power spectrum that you will see. Each type of waveform has its own predicted power spectrum and spectrogram. The three basic types are:

1. **Amplitude Modulated Signal (AM):** Consider a sine wave of frequency **f** (called the **carrier**) that is amplitude modulated in a sinusoidal way **w** times a second. The power spectrum of this sound is a band of energy for the carrier at frequency **f** and a sideband on either side of the carrier band, one at a frequency **f+w** and the other at frequency **f-w**. The two sidebands will always have less energy than the carrier. The amount of energy in the sidebands increases as the amplitude variation in the original waveform is increased.

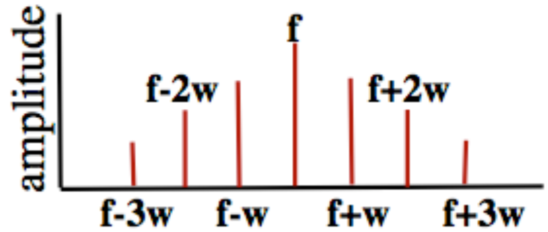


2. **Frequency Modulated Signal (FM):** Consider a sine wave of average frequency **f** (again called the **carrier**) and fixed amplitude that is frequency modulated in a sinusoidal way **w** times a second. In other words, the frequency of this signal rises and falls in a sinusoidal way around the average value **f**. The power spectrum of this sound is a band of energy at **f** and a set of sidebands on either side of that carrier band

at frequencies **f±w**, **f±2w**, **f±3w**, etc. Usually, the further the sideband from the carrier, the smaller its amplitude. However, if the carrier is sufficiently highly modulated, the amplitude of the first few sidebands can be greater than that of the carrier.

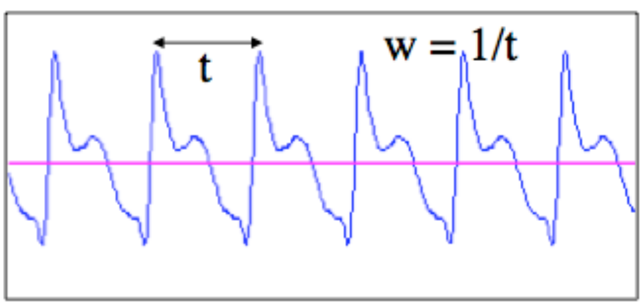


Time Domain

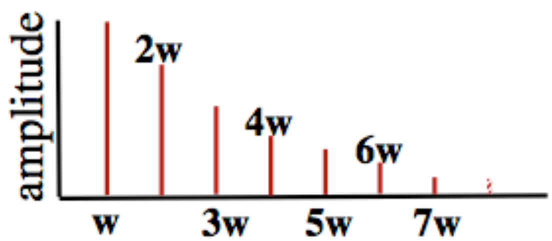


Frequency Domain

3. **Periodic Non-sinusoidal Signal:** Consider a signal that repeats regularly at some rate **w** times/second, but is not a sine wave and is neither the AM nor FM cases noted above. Most frogs, mammals, and non-song birds make such sounds. The power spectrum of this signal will be a harmonic series with energy at frequencies **w**, **2w**, **3w**, **4w**, **5w**, etc. The first frequency component, **w**, is called the **fundamental**. The rest are called the **second harmonic**, **third harmonic**, etc. Usually, the amplitude of the components decreases as one examines successively higher harmonics. If the repeating waveform unit is symmetrical (so that the first half looks like an inverted version of the second half-called "half-wave symmetric"), then only the odd harmonics (**w**, **3w**, **5w**, **7w**, etc.) will be present. Here is an asymmetrical example:

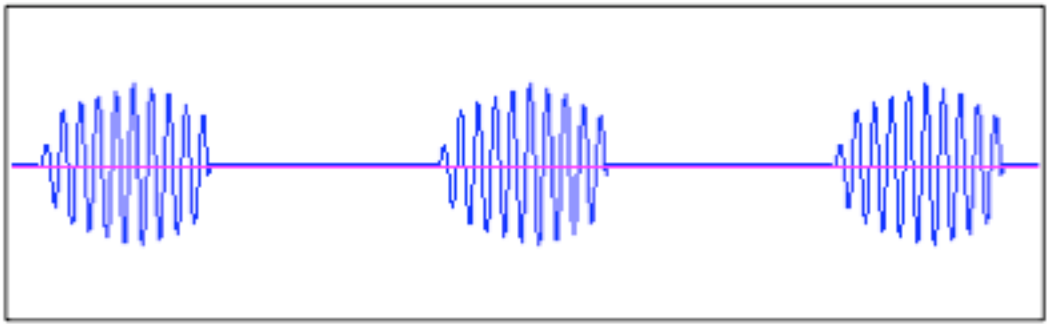


Time Domain

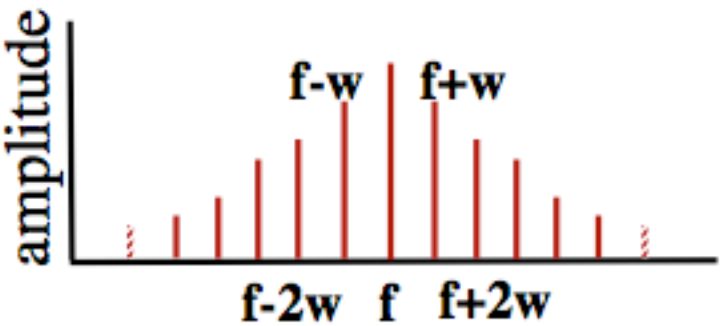


Frequency Domain

B. Compound Signals: Most signals that you will encounter are mixes of two or more of these basic waveforms. The mixture might be simultaneous (e.g. amplitude modulation of a periodic non-sinusoidal signal) or successive (the first part of a signal shows AM and then shifts to FM in the latter parts). As an example of a compound signal, some frogs produce short bursts of a frequency **f**. This is amplitude modulation by a modulating waveform that is periodic but not sinusoidal:



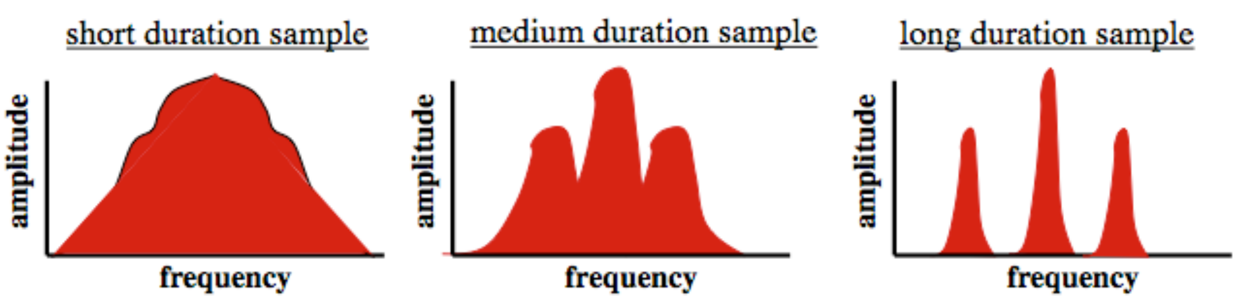
Suppose the bursts occur at a rate **w** times/sec. Then the modulating waveform consists of a harmonic series. Each of these harmonics in the modulating waveform will appear as a pair of sidebands positioned around the carrier at **f**:



Bandwidth and the Uncertainty Principle

A. Ideal Analyzer: An ideal sound analyzer would be able to take any waveform we record, cut it into infinitely tiny adjacent segments (allowing us to monitor very accurately any changes in its Fourier structure over time) and identify all the frequencies present in each segment with great accuracy.

B. Uncertainty Principle: To measure any frequency accurately, an analyzer needs to count how long it takes to complete several successive cycles of the sound. The more successive waves that can be counted, the more accurate the computation. If we cut a waveform into very small segments, there will not be enough cycles in each segment to accurately measure the rate at which the wave repeats. If a segment contains several components that are fairly similar in frequency, the analyzer will then be unable to recognize them as distinct components, and instead will record one large band of energy covering a range of similar frequencies:



Frequency resolution, (the ability to distinguish two similar but different frequencies) thus decreases as segment duration is decreased. If one wants to distinguish those frequencies, one has to cut the sound into fewer and longer segments. However, making segments longer reduces the analyzer's ability to track rapid changes during the sound. The temporal resolution of an analyzer gets better as one makes the segments smaller. Thus there is a tradeoff in sound analysis between getting good temporal resolution on the one hand and good frequency resolution on the other: you cannot have both at the same time. As a general rule, the smallest difference between two frequencies that can be resolved by an analyzer, Δf, and the smallest duration between two successive events that can be recorded by an analyzer, Δt, are related by the rule:

$$\Delta f \bullet \Delta t \approx 1.$$

C. **Bandwidth:** The bandwidth of a sound analyzer is the value of Δf that one has decided to use. If Δf is smaller than the difference between the harmonics in a periodic non-sinusoidal signal, or the difference between the carrier and sidebands in an AM or FM signal, then the resulting power spectrum and spectrograms will show the fully decomposed frequency domain image of the sound (e.g., all of the component bands will be visible and separate). But if Δf is larger than the difference between adjacent frequencies, then the power spectra and spectrograms will be unable to break the waveform into its components, and they will have to plot some sort of time domain image of the sound. Some practical guidelines:

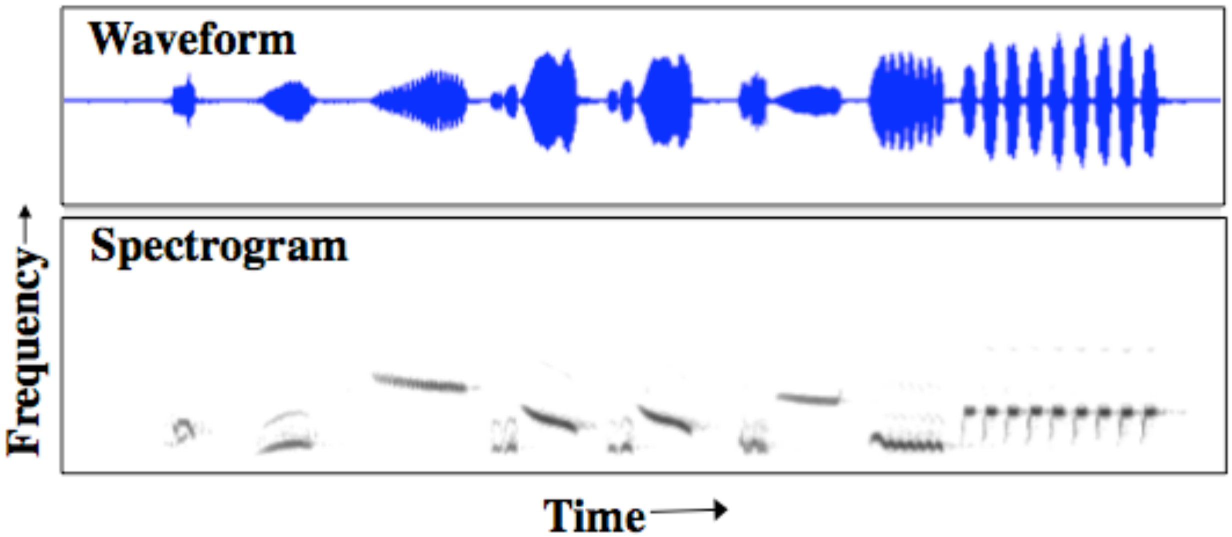
1. **AM Signals:** For most animal sounds, one wants to know what the carrier frequencies are in a signal, but does not want to see AM sidebands. Here one sets Δf so that it is bigger than the amplitude modulation rate **w** (which is usually easy since animals tend to amplitude modulate signals at much slower rates than the bandwidths used to separate harmonics).

2. **FM Signals:** Similarly, one usually does not want to break an FM signal into a big set of sidebands. Instead, if a bird or mammal is frequency modulating a sine wave, we want to see how the frequency changes and to measure the maximum frequency, the lowest frequency, how fast it changes, etc. So again, one usually sets the bandwidth Δf so that it is *larger* than the FM modulation rate. This is typically easy since most birds modulate frequencies relatively slowly.

3. **Periodic Non-sinusoidal Signals:** Because the shape of the repeating unit for a periodic non-sinusoidal signal can be anything, and thus hard to describe, this is precisely the kind of signal where breaking the sound into its harmonic series gives one the ability to describe it quantitatively. So in this case, we usually set the bandwidth Δf so that it is *smaller* than the repeat rate of the periodic signal. This will ensure that the analyzer breaks the signal down into its harmonic components.

4. **Optimal Δf:** The optimal bandwidth is thus one that breaks rapidly repeating periodic non-sinusoidal signals into harmonics, but leaves AM and FM components un-

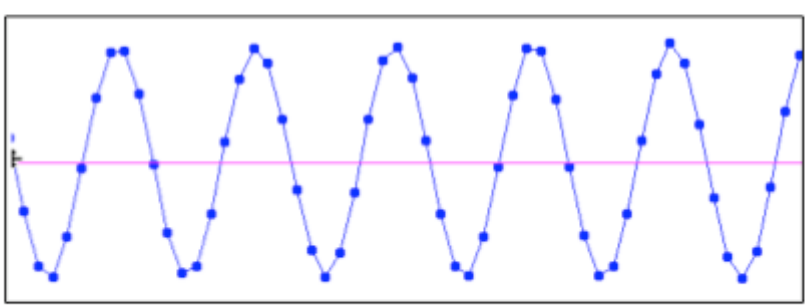
decomposed. There are of course exceptions for specific needs, but this is a good general goal. Here is an example of optimal settings for a bird song:



5. **Interpretation:** It is important to know how the bandwidth you have chosen relates to time domain structure. For example, the same sound might be described as a harmonic series or a pulse train depending on where the bandwidth was set on the signal analyzer. The same is true for FM and AM patterns: the same signal can look very different on a spectrogram depending on how the bandwidth is set. The safe solution: always examine BOTH the waveform and the spectrogram of any signal and make sure you have considered how any conspicuous patterns in one domain are translated into the other domain. If you have an analysis program that lets you change the bandwidth continuously, try different values to see how sensitive your spectrogram is to the bandwidth value. Usually, there will be some intermediate value of bandwidth below which the spectrogram has one appearance, and above which it suddenly shifts in appearance. Other than finding this cutting point and being sure you are on the side you need, bandwidth may not be too critical to your analysis.

Digital Sound Analysis

A. Digitizing Sounds: Analog tape recorders and sound analyzers preserve a continuous record of a sound waveform. Digital recorders and analyzers, in contrast, sample an incoming wave at a given rate and store the amplitude of the waveform for each sample in a file:



Such files thus just long strings of successive numbers with some header information on sampling rate and other details. Where all the samples are stored, the file is said to be "lossless". File formats that are lossless include **.wav** and **.aiff** files. Other formats try to save file space by "compressing" the data. There is some reduction in fidelity but most human listeners will not notice. Examples include the popular **.mp3** format.

B. Sampling Accuracy: When a digitizer samples a sound wave, it stores the amplitude that it measures with a limited accuracy. Cheap sound digitizers use 8 bits of storage per sample which only provides moderate amplitude accuracy; commercial CD's store each sample with 16 bits of storage which permits much more accurate records of amplitude. Sound archives typically use 24 bits per sample. Clearly, the more bits you use per sample, the larger the sound file that has to be stored. MP3 and similar protocols are ways to compress sound files to reduce storage needs yet retain reasonable quality. MP3 quality is usually lower than that for commercial CD's.

C. Sampling Rates: The rate at which sounds are being sampled in a digital analyzer is usually stored in the header of the sound file and the file itself is then just a list of numbers (each number being the amplitude of the waveform at that sample point). Digital sound files are an incomplete record of a sound when compared to an analog tape recording: clearly, the intervals in the original signal between samples have been lost during digitizing, Thus the higher the rate of digital sampling, the more accurately the digital data file characterizes the original waveform. Music on commercial CD's is digitized at 44.1 kHz (e.g., there are 44,100 samples stored for every second of waveform). Such a sound file can get very large for a long duration sound. One can economize on storage needs on the analyzer (usually a computer) by using lower sampling rates, but usually at the expense of the quality of the sound. Some people (and probably many animals) can tell the difference between original sounds and CD digitized copies, so sound archives typically digitize sounds at 96 kHz or higher.

D. Nyquist Frequency: A digitizer must be able to sample a sine wave at least twice per cycle before it has even a slim hope of characterizing its frequency. Measuring more cycles is of course better, but 2 separated samples per wave is the absolute minimum required. This means that one must digitize sounds at a rate equal to **twice** the highest frequency in the sound. Put another way, if you digitize at some rate **R**, the highest frequency that should be in the sound sample, called the **Nyquist frequency**, is **R**/2. If there are sound components present with frequencies greater than **R**/2, these can produce artifacts in power spectra and spectrograms called **aliasing**. Usually, the components above the Nyquist cutoff get "folded" over the cutoff and appear below it inverted from what they would have been had one used a higher R. To avoid aliasing, we usually use a high enough **R** that there are likely to be no noticeable components in our sounds with frequencies greater than **R**/2, or we put a filter between the tape recorder and the computer to filter out any frequencies above the Nyquist cutoff. Note that if you are using a digital recorder, it may be too late to add a filter after recording. Thus many of these recorders have filters built in to remove frequencies greater than half the digitization rate before they store the sounds.

E. Digital Bandwidths: Most sound analyzers that use digital data do not give the user the option of picking a particular bandwidth, **Δf**, but instead ask the user to specify a segment length (usually called "FFT Size) in numbers of samples/segment. This is also called "frame length" or "transform size". What this means in terms of real time (in seconds) depends upon the original digitization rate **R**. As a rough guide, divide the number of samples/frame by the sampling rate **R**. This is the fraction of a second that the segment represents. The reciprocal of this fractional time is an estimate of the bandwidth in Hz. The more samples/segment that you specify, the better the frequency resolution of the subsequent power spectra and spectrograms, but the worse the temporal resolution.

F. Windowing: Cutting a continuous sound into segments actually converts the original sound into a set of new sounds. Where originally one segment flowed smoothly into another, there are now abrupt starting and stopping points. The Fourier composition of a sharp start or stop is a broad band of frequencies. Broad bands will thus be present at the beginning and end of each segment that were not in the original signal. To try to reduce these artifacts, most sound analyzers do not cut a sound into segments with instantaneous starts and stops, but instead, taper the onset and offset of each segment gradually. There are different opinions about how to do this, and the result is a number of different analysis "windows" such as the Hanning window, the Hamming window, the Blackman window, etc. A "rectangular window" is one without any tapering at all and thus has the sharp onset and offsets. One rarely uses a rectangular window, but instead one of the others. Which one depends a bit on taste, as they are largely similar in result.

G. Spectrograms vs Power Spectra: A power spectrum (or spectrogram "slice") shows the amplitudes of all frequency components present in a segment of signal that can be discriminated given your bandwidth setting. All components present in that segment are there. In a spectrogram, strings of successive segments are connected so that you can see how power spectra change during the course of the signal. Showing linked spectra as a three-dimensional plot usually results in too much information and thus one simplifies the graph by hiding the lower energy frequency components. Most spectrogram analyzers provide controls that allow one to determine the lowest amplitude component that will be shown and how much contrast (range of marking from white to black) is to be used to discriminate between the remaining components' amplitudes. Adjusting these parameters to give optimally clear spectrograms is an art that requires practice and a good eye.

H. Scaling and Measurement: Once you have a power spectrum or a spectrogram, most programs provide settings that let you adjust the time, amplitude, and frequency scales to optimize the view. In addition, there are usually tools that allow you to make measurements of any of these parameters off of the screen using cursors or selection mechanisms. There are methods such as Spectrographic Cross Correlation that compare entire spectrograms in a quantitative way for statistical analysis. The toolkits for utilizing spectra and spectrograms are growing all the time. It is a powerful entryway for sound analysis and comparison.