Additive Synthesis

Additive Synthesis creates complex sounds by combining simpler sounds, usually sine-waves of varying frequencies and amplitudes. The following pages describe how, as well as being a useful tool in the production of electronic sound, Additive Synthesis is also nature's chosen method of making herself heard.

This document is an extract from the documentation for **AdsynDX**, an implementation of Additive Synthesis for Windows XP.

How sound works

Sound is the process whereby the physical variations occurring in one object are propagated through an intervening medium (the air) to another object that responds in some way to these variations.

The ways that real objects vibrate, and the way those vibrations are transmitted through air, can be described and explained by relatively simple theories of mathematics and physics. An understanding of this theory leads to a greater appreciation of how musical instruments work, and how their sounds can be synthesised.

The physics of sound

All physical objects have an inherent 'springiness', so that when energy is imparted to them, e.g. by striking or bowing, they respond by vibrating in a spring-like fashion. Most objects, when struck, do not produce what we would consider a musically interesting sound. This may be because the object is made of a material which is not very springy, so that any vibrations induced in the object are quickly damped out (which is why striking a block of wood produces a 'dead' sound). Other objects may vibrate at so many different frequencies that we can't detect a definite pitch in the sound; the 'splash' of noise from a cymbal, for example, is only useful in a percussive, or rhythmic, context – you can't play a tune on a cymbal. In musical instruments, resonant materials are used, but these are usually constrained in some way so that their resonant frequencies are limited, and the instrument is tunable and can produce sounds with pitch.

In many musical instruments the vibrations are initially induced in either a stretched string (of metal or nylon), or in a column of air enclosed in a cylinder. The physics of vibrating strings and air columns is quite simple (and elegant). In both cases the vibrations occur in the form of *standing waves*, the patterns you see when pluck a string that is held firmly at both ends, or that appear on the surface of a glass of water when struck. The waves are described as *standing* because, although the waves are travelling along the length of the string (or air column), they appear to be stationary because they are reversed and reflected at the fixed (or closed) end. In the case of a string, the standing waves are formed by the movement of the string; in an air column the waves result from movements of air and the changes in pressure along the length of the column.

The simplest way a string, fixed at both ends as in a musical instrument, can vibrate is shown in Figure 1.

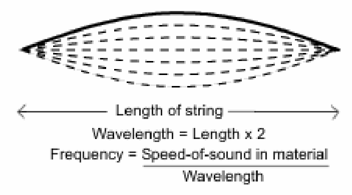

Figure 1: the *fundamental* vibration mode of a fixed string.

This shape is actually one half of a cycle of a sine-wave, with a wavelength twice the length of the string, and the wave is described as being *sinusoidal*. This is the longest wavelength (and thus lowest frequency) at which the string can vibrate, and is called the *fundamental frequency* of the string. The actual frequency of the vibration (i.e. what we would perceive as its pitch) is related to the string's length, and the speed of sound in the material of the string (frequency = speed / wavelength). The speed of sound in the material is related to its mass, which itself is related to the tension of the string (because the higher the tension the thinner the string and the lower its mass). This is why the four strings of a violin, although of the same length, produce different pitches: they are of different thickness, i.e. mass, and at differing tensions.

As well as the fundamental mode, a string can also vibrate in other modes, although these will have less energy (i.e. are at lower amplitude) than the fundamental. The frequencies of these other modes are constrained by the fact that their starting and ending amplitudes must be zero to coincide with the fixed ends of the string, as is the case for the fundamental. These other possible modes turn out to be integer multiples of the fundamental frequency. The first 2 of these modes are shown in Figure 2; the first is a vibration at 2-times the fundamental's frequency, and is known as the 2nd *harmonic*. The next mode is 3-times the fundamental, or the 3rd harmonic.

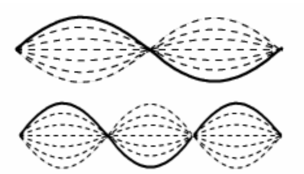

Figure 2: the 2nd and 3rd *harmonics* in a fixed string.

Some of these harmonics, specifically the 2nd, 4th, 8th etc., are equivalent to a doubling of the fundamental frequency. A sound with a frequency that is twice that of another frequency is perceived by us as being an octave higher in pitch. Other harmonics do not have this simple relationship with the fundamental; the 3rd harmonic, for example, is an octave plus a musical fifth above the fundamental. The 5th harmonic is at a frequency 2 octaves plus a major third above the fundamental. Remember, these harmonics are not the result of playing different notes on the string; they are present to some extent in each individual note played. They contribute greatly to the overall tone (or *timbre*) of an instrument. Other factors affect the tone: if the string is struck or plucked there will be sound generated by the plectrum strike; if the string is bowed, it will tend to be pulled by the bow in a saw-tooth fashion, which adds to the timbre. The string itself is acoustically coupled, i.e. its vibrations are transmitted to, the body of the instrument (by the *bridge*), so that the resonance of the body and the air it contains contribute to the instrument's tone. However, all these additional components of the timbre can be thought of in terms of the same sinusoidal vibrations described above.

In a wind instrument the equivalent to the vibrating string is a vibrating column of air contained within the instrument. This is made to vibrate by the energy imparted to it by blowing into or across the mouthpiece. The physics of vibrating columns of air are similar to those of the string described above. The standing wave is present in the form of alternating areas of high pressure *nodes* and low pressure *antinodes* along the length of the air column, and the modes of vibration are limited by the *boundary condition* at the ends of the column. In some instruments, such as the flute, both ends of the tube are open. This means that at these end points the pressure of the air in the column will be equal to that of the outside air, i.e. at atmospheric pressure. Some vibrational modes in a column of air are shown in Figure 3. Note that in the following diagrams the red traces indicate the amount by which the air is being squeezed; where the lines meet the air is at maximum pressure (a node), and as they move apart the pressure decreases towards atmospheric pressure.

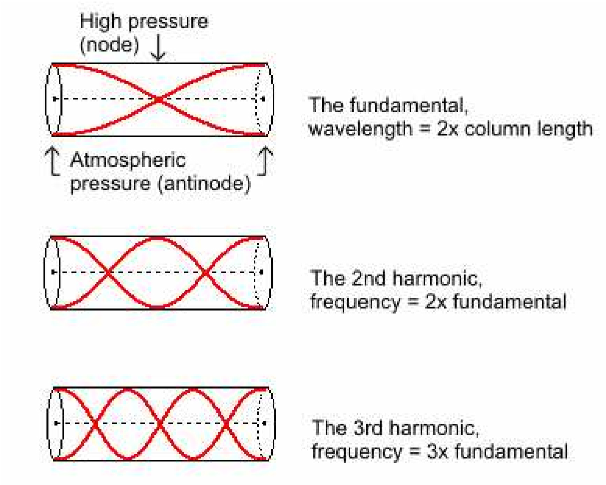

Figure 3: the Fundamental, 2nd and 3rd *harmonics* in a column of air open at both ends.

In other instruments, such as the clarinet and the brass family, one end of the air column is closed. This changes the constraints on the possible harmonics because the air at the open end will be at atmospheric pressure (an antinode), while at the closed end there will be a high pressure node. This is shown below.

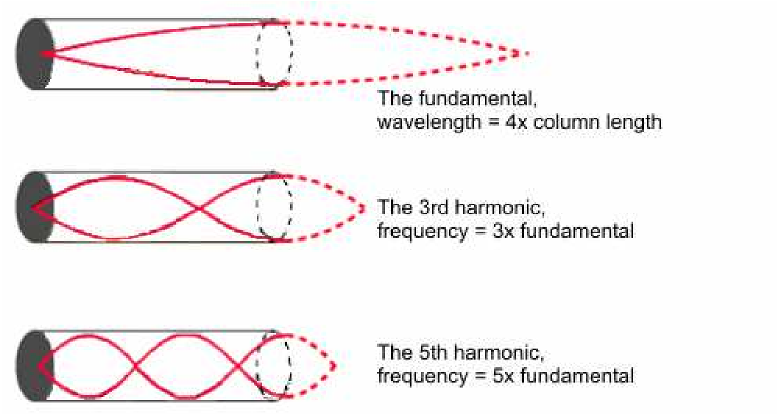

Figure 4: the Fundamental, 3rd and 5th *harmonics* in a column of air closed at one end.

The closure of one end of the column has some interesting consequences. Firstly, you can see that the wavelength of the fundamental is twice that of an open-ended column of the same length. The fundamental is thus an octave lower. Secondly, the even-numbered harmonics aren't present. This is because they would require a high-pressure node at the open end of the column, which is not possible.

In an actual instrument, e.g. a flute, different musical notes are achieved by the player uncovering holes along the length of the tube enclosing the air column. This creates points open to the atmosphere, so that the standing wave in the air column reconfigures itself to include an antinode at this point, changing the frequency (pitch) of the vibrations.

As in the description of the vibrating string, the above describes the behaviour of an idealised air column, and ignores factors present in a real instrument, such as the noise of the player's breath, variations in the shape of the column and the material of the containing cylinder, all of which contribute to the sound of the instrument (but are themselves produced by the same vibrational mechanisms described above).

In fact, anything that we perceive as sound (with the exception of 'white' noise, which consists of vibrations at all possible frequencies) can be described in terms of the sum of a set of sinusoidal vibrations. Our ears have evolved to respond to a range of frequencies suited to our survival (from about 15 to 20,000 cycles per second), so that we can distinguish the sound of a mouse from that of a lion, and recognise a decent tune. If we didn't filter out other frequencies we would hear the sound of earthquakes striking the earth (with frequencies from about 1 to 30 cycles per second) and the subsequent bell-like ringing of the Earth itself, with a fundamental frequency of about 1 cycle per hour.

The mathematics of sound

Figure 5 shows a sample of a recording (about ¼ of a second long) of a note played on a real instrument. If you've read through the previous section on the physics of sound, you'll know that this is the result of the various sine waves created by the resonances of the instrument being played, combining together to form this complex waveform.

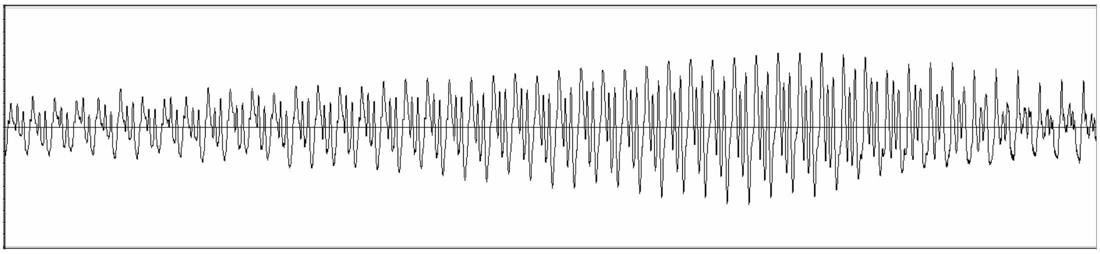

Figure 5: part of a recording of the sound of a real instrument.

Could we, by examining this waveform, determine anything about the instrument that created it, e.g. what type of instrument it was, or what note it was playing? There is some regularity in the wave, so perhaps we could count the major peaks, which are probably attributable to the fundamental, and (if we knew the duration of the sample) we could calculate the fundamental frequency. In fact the waveform was created by a cello, but to work this out from the waveform we would need some way of decomposing it into its constituent parts, and then comparing this 'mix' to tables containing typical partial mixes (harmonic spectra) of various instruments.

This of course is exactly what our ears and brains do continuously, and very quickly, all the time. How our ears and brains achieve this is discussed a little later. But it is possible to use mathematical procedures to analyse the sound; this is particularly relevant now that we can record sound digitally and process it with the power of computers. This is the realm of *digital signal processing*.

Imagine that someone plays a long, sustained Middle C on a cello. We know that this note will not be a pure tone at 261.63 cycles per second (the frequency of Middle C), but will contain other components at multiples of this frequency (the upper partials). If we had a tuning fork, tuned to Middle C, and placed it in front of the cello, we'd see that the tuning fork resonated strongly with the cello's sound, whereas tuning forks tuned to the pitches either side of C (i.e. B and C#) would not resonate, or only relatively weakly. If we had lots of tuning forks, tuned to the pitches of all the possible harmonics of Middle C, and some way of measuring the amplitudes with which each one resonated with the cello note, we could analyse the harmonic content of the sound. This is the basis of the mathematical procedure called the *Fourier Transform*, named after the 19th century French mathematician Jean Baptiste Fourier who, with others, developed theories and methods that could analyse systems, such as vibrating strings, which can be regarded as being made up of a series of sine waves.

The Fourier Transform allows us to take a section of a sound signal, such as the one in Figure 5, test it mathematically against a set of sine waves (equivalent to the tuning forks above), and measure the amount of each harmonic's contribution (if any) to the sound. With this information it becomes possible to recreate the sound (by adding together the required amounts of digitally or electronically generated sine waves) and to process it in various ways (such as by filtering out or emphasising some of the partials).

The transform as originally proposed by Fourier involved comparing the signal with an infinite series of sine waves, infinitesimally close together in frequency. The intervening 200 years have seen refinements to the transform so that today we can use a form of the transform which tests the signal against a discrete set of sine waves (the *Discrete* or *Finite Fourier Transform*) and uses some manipulation to limit the amount of calculation involved, the *Fast Fourier Transform*.

In making the transform faster, and thus making it suitable for use in digital signal processing, some compromises have been made. The transform operates on a set of samples, numbers representing the amplitude of the input signal at regular intervals (the *sampling rate*). In the Fast Fourier Transform the number of samples in the set must be a power of two, i.e. the transform can analyse 128, 256, 2048, 32768 etc. samples, but doesn't work with 300 or 30000 samples in the set. The result of the transform is a set of numbers of the same size as the input sample set, with each number representing the amount present in the signal of a defined (discrete) frequency.

This means that the larger the size of the sample, the more accurate the result (in terms of the number of discrete frequencies analysed). If we analyse 128 samples we will end up with a power spectrum of 128 frequencies (spread between zero cycles per second, and the sampling rate used). Double the sample size and we can look at 256 frequencies. However, there is a trade-off; the larger the sample set size, and the better the resulting frequency resolution, the lower the resolution in time with which we can analyse the sound. Suppose, for example, that the sample in Figure 5 is just one of a continuous series of samples from a longer recording, sampled at CD quality, i.e. 44,100 samples per second. If our sample set size is 4096 samples, each set is just less than $1/10^{th}$ of a second in duration. This means that we can analyse the sound into 4096 frequency bands ten times per second. If we increase the set size to 32,768 samples we will get a more accurate analysis of the harmonic content of each set (into 32,768 frequency bands), but we will be able to analyse the signal less than one and a half (i.e. 44,100 / 32,768) times per second. If we increased the sample set size to include the whole recording (or the nearest power of two thereof) we could get a very accurate picture of the harmonic content. However, each transform only measures the average amplitude of the frequencies over the length of the sample set. In practice this means that if we are analysing a signal in which the harmonics change rapidly (speech, for example), we need to use a relatively small sample set size, to catch the changes, and live with the consequent inaccuracy in the frequency resolution. On the other hand, if we are examining a long, sustained chord sung by a choir, we can use a larger sample set size to capture more of the individual frequencies involved, and accept that we may be missing, (or rather, averaging out) the changes occurring in the dynamics over the course of a sample set.

Note: the Fourier Transform as implemented in AdsynDX has user-adjustable settings for various parameters of the transform so that, by trial and error, the best results can be achieved.

The Ear

Having seen that, with enough computer power, and using quite complex mathematics, it is possible to analyse an audio signal into its harmonic components, this raises the question of how we are able to do the same thing, so quickly and effortlessly, with our own ears and brain. When we hear a chord and are able to identify the individual notes, are we performing the same kind of computations described above? The answer is that the ear is, in effect, a biological Fourier transformer, operating in the analogue, rather than digital, domain. The mathematics described above have been built into its structure by evolution.

There are many web-based resources that describe the intricate construction of bones, membranes, fluids and hairs that make up the ear. Much of this structure is concerned with focusing and amplifying incoming sounds. The final stage of a sound's journey through the ear occurs in the *cochlea*, a shell-like structure of the inner ear. In the cochlea is a tiny structure called the *organ of Corti*; this is where the incoming sound is finally converted into the electrical nerve impulses that are transmitted to the brain. This *transduction* is performed by a series of intricate hair cell structures positioned along a membrane, the *basilar membrane*, that responds in a wave-like fashion to the pressure variations of the sound. In 1961 the Hungarian biophysicist Georg von Békésy was awarded the Nobel Prize for theorizing, and demonstrating with a mechanical model, that the structure of the cochlea and the membrane, and the distribution of the hair cells along the membrane, was such that each hair cell was positioned so that it would respond to a different frequency component of the sound, and transmit an appropriate impulse to the brain. The brain receives from the ear a constant, real-time and very accurate analysis of the frequencies present in the incoming sound.

What the brain does with this information is beyond the scope of this introduction. In reality, not much is known. Some sound processing done by the brain is fairly basic, in evolutionary terms, and happens subconsciously without the need for our awareness. We react instinctively to the possible danger signaled by sudden loud noise, and only later pay attention to the details of the sound to determine whether it came from an ambulance siren or a police car. One thing we can say about our conscious experience of sound, for example in respect of the so-called *cocktail-party effect*, our impressive ability to extract and listen to just one voice amongst many at a crowded party, is that the fineness of the analysis done by the ear, and the brain's ability to process this information, far exceeds what can be done with algorithms and silicon.

Andy Bridle, 2007