The Digital World

Digital Everything

Digital clocks and digital watches, digital cameras and digital camcorders, digital television, digital thermometers, digital toasters and digital steam irons and digital coffee-pots, personal digital assistants….

What does it mean for something to be "digital"? Narrowly speaking, it means that somewhere along the line, whatever sort of signals the device is using are encoded as sequences of bits and manipulated using the kinds of electronic circuits that are found in computers.

More broadly speaking, digital devices encode whatever data they are working with using a discrete set of symbols, and manipulate these symbols. (That's kind of a mouthful, and it doesn't really tell you anything until you see the examples. Read on!)

We have already seen how numbers and text are encoded as sequences of bits, but that's just replacing one symbolic representation by another, closely related one. It's more interesting to see how we do the digital trick with images and sounds and the like.

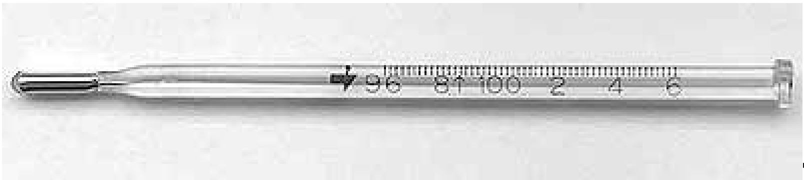
The Analog World of my Youth

The opposite of "digital' is "analogue", or "analog". Analog devices start out with one kind of continuously varying data and change it into another kind of continuously varying data. Sometimes the words *analog* and *digital* are used to contrast different sorts of clocks and watches: those that display the time as something like "6:45" are digital, those that have hands moving around in a circle are analog. But this is a particularly misleading use of the terms, since practically all present-day clocks and watches use the same sorts of digital integrated circuits to produce the one-second pulse that keeps the time. The difference is just in the part of the device that counts the pulses and displays the count.. For that matter, old-fashioned mechanical clocks,

including pendulum clocks, deserve to be called digital: The pendulum is used to break the continuous flow of time by a regular series of discrete pulses, and the mechanism that drives the hands counts the pulses. If you want a *real* analogue clock, try an hourglass, or a sundial.

A digital clock?

A much better example of an analog device is a mercury fever thermometer, the kind that used to come out of the medicine chest whenever I told my mother that I wasn't feeling quite up to going to school.



These have now become illegal, not because they're not digital, but because the mercury in them is poisonous.(The biggest concern is that people will toss them out in the trash, and the mercury from the broken thermometers will released into the environment.)

How do these work? When the bulb at the end containing the mercury heats up, the mercury expands and advances along the hollow tube inside the thermometer. One continuously varying quantity (temperature) has been converted by a physical process into another (length of the column of mercury). To be sure, there is a digital element here as well: the calibration marks on the thermometer's scale translate the length into one number in a small set of available values. In fact most of the devices we'll look at have a combination of analog and digital elements.

Analog Sound Recording

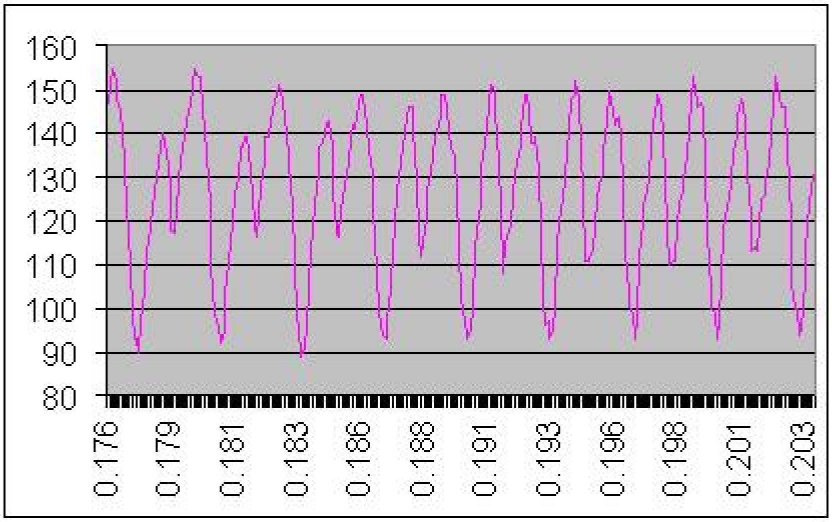
The most analog thing I can think of is the phonograph.

We have a pretty good idea of what Shakespeare looked like, and a very good idea--thanks to the many portraits produced by skilled painters and sculptors--of what George Washington looked like, in spite of the fact that neither of them was ever photographed. If, by some miracle of time travel, we were to see Washington walking down the street, we might well recognize his face. But what did they *sound* like? The voices of these people and their contemporaries, and everyone who came before them, are lost to us forever.

Before the invention of the phonograph, there was no sound recording. For many years after, until the advent of magnetic recording on wire and tape, phonograph records were the only sound recordings that existed, and until quite recently they were the most common medium for listening to recorded music in your home. (Sales of CDs first exceeded those of phonograph records in 1988, the year many of you were born. You probably don't think of the year you were born as "quite recently", but I do.)

Before we consider how phonographs work, we'll need to discuss how sound itself works. The sensation of sound is due to rapid changes in the pressure of the air near your ear. These cause your eardrum to vibrate, and the vibrations are passed through some elaborate machinery in your ear before being sent to your brain for sorting out.

If you plot the pressure variations on the vertical axis versus time on the horizontal axis, you get something like this:

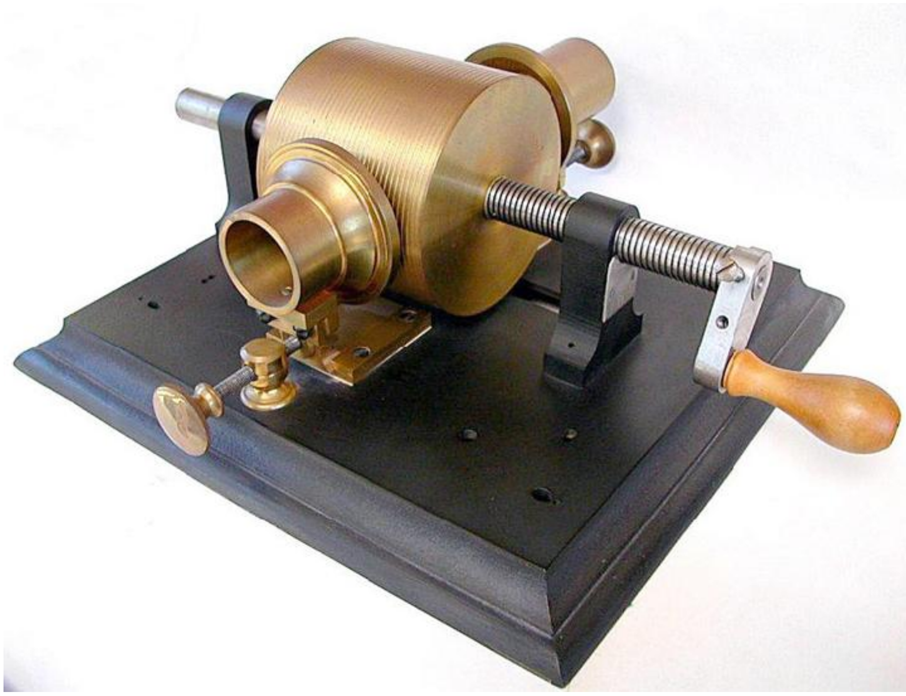


In fact, this chart is taken from a brief recording of a French Horn. The numbers on the x-axis are in units of seconds from the start of the clip. (I'm only showing you a very brief portion of the clip, about 1/40 sec. in duration.) The units along the y-axis, which measures the intensity of the sound, are not especially relevant for now.

The very regular pattern in the vibrations is typical of musical notes. The larger the distance between the peaks and valleys of the successive waves, the louder the sound will be. (The actual height of the wave within the plot area is irrelevant, it's the distance between the high points and low points—the *amplitude*—that counts.) The faster the vibrations—the higher the *frequency*--- then the higher the pitch will be. The extract shown above makes about 8 complete vibrations in 0.025 sec, which comes out to a frequency of 320 cycles per second, or 320 *hertz*. (If you're a music type, that's about E above middle C.)

Now anything that makes your eardrum vibrate in exactly this fashion will sound exactly like the French horn. If you could make a record of this amplitude-versus-time relation, and later use the record to produce vibrations with the same amplitude-versus-time pattern, you ought to hear something very much like the original sound.

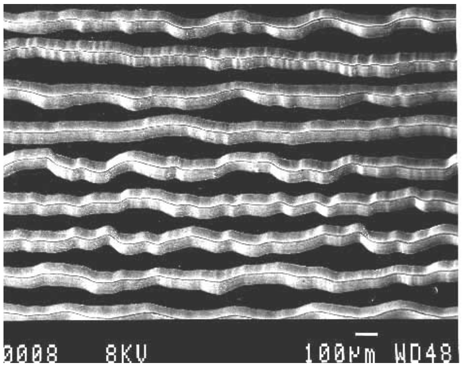
In 1877, the American inventor Thomas A. Edison built a device that consisted of a cylinder wrapped in tinfoil, along with a diaphragm attached to a sharp needle that just touched the surface of the cylinder. Edison spoke into the diaphragm while turning a crank that rotated the cylinder. The sound of his voice caused the diaphragm to vibrate, which in turn caused the needle to engrave an indentation into the surface of the tinfoil. The greater the amplitude of the vibration, the deeper the indentation was. When he finished speaking, he placed another needle-and-diaphragm unit into the indentation produced by the first one and turned the crank. The new needle rode up and down in the indentations, these motions were transmitted to the diaphragm, and Edison heard his voice. The first sound-recording device literally engraved a a chart like the one above into the surface of the cylinder.



Replica of Edison’s 1877 Phonograph

Although Edison struggled to keep his cylinder recordings competitive, they were soon supplanted by phonograph records on flat disks, which are easier to mass-produce and store. In a disk recording the needle moves from side to side rather than up and down.

An early disk phonograph record.



Microphotograph of grooves in a phonograph record. The area shown is about 2 millimeters (1/12 inch) wide.

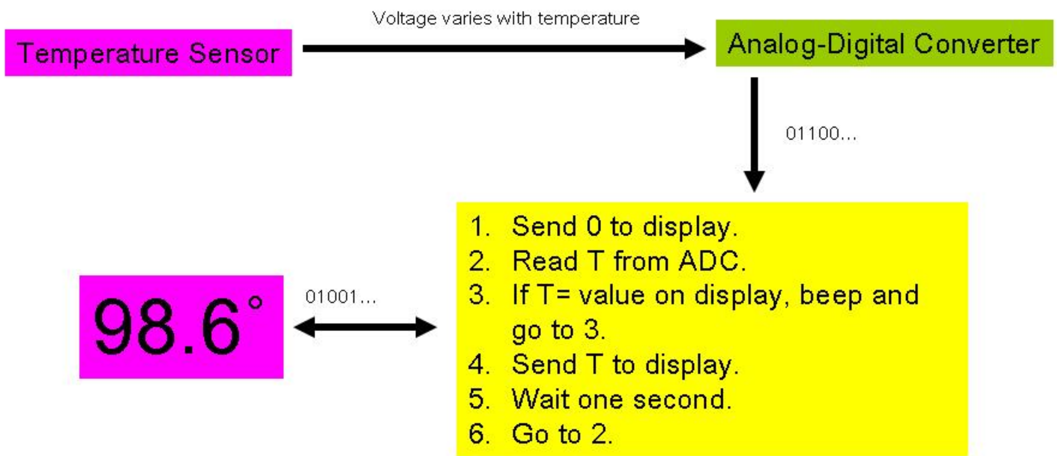
For the first fifty or so years of their existence, phonographs were completely *mechanical* devices. No electricity was involved in the sound reproduction---a fancy horn was used to amplify the faint sound made by the needle wiggling in the groove. It is astonishing that something so beautifully simple can work so well. Later on, mechanical sound reproduction was replaced by a process in which the needle's vibrations were converted into a varying electrical voltage, which was amplified electronically. But, mechanical or electrical, it's still an entirely analog process.

Digital Sound and Images

Digital Thermometers

Yes, we're going to talk about how digital sound recording and digital images work, but let's not leave that thermometer behind just yet.

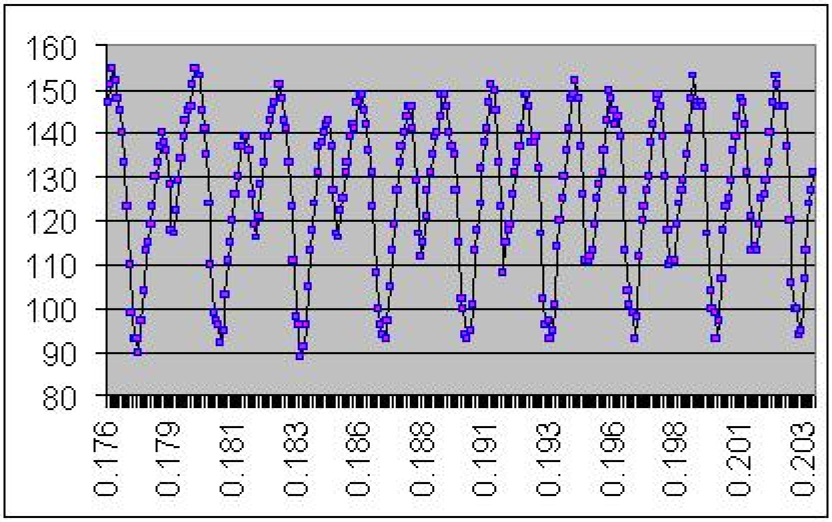
Here is a rather simplified diagram of the functioning of a digital thermometer



At the upper left of the diagram is an analog device that senses temperature, and produces an electrical signal whose voltage varies with the termperature. The temperature corresponding to this voltage is encoded as a series of 0's and 1's by an *Analog-to-Digital Converter.* The display device at lower left stores a string of bits and displays the corresponding temperature. But the heart of the thermometer is the large box in the middle. This is a *program*, a series of instructions for manipulating the bits that appear on the Converter and that are stored in the display. *Digital devices manipulate* ***information****, and typically use complex programs to do this.*

Digital Sound Recording

Let's go back to that chart showing the sound wave of the French horn. Suppose you measure the height of the chart at very frequent intervals and note down the *numbers.* This would be like taking one of Edison's phonograph records and measuring the depth of the indentation at many different locales, then throwing away the record and just keeping the measurements. You could use that list of numbers at a later time to reconstruct the chart and then reproduce the sound.



155,152,148,145,140,133,123,110,99,93,90,93,97,…….

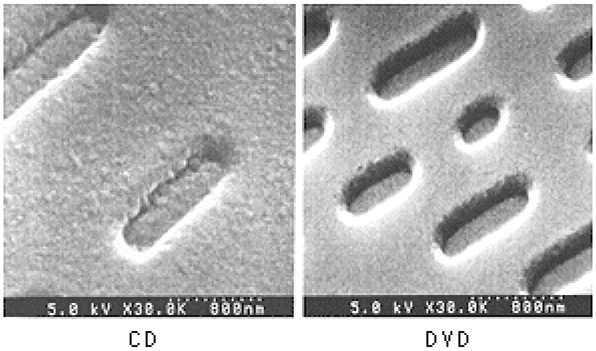
That's what digital sound recording does: The start of the process is the same---. a microphone is used to convert the vibrations into a varying electrical voltage. Then an analogue-to-digital converter measures this voltage very frequently (thousands of times per second) and encodes the voltage as a sequence of bits. What makes it digital is the conversion of the continuously varying signal into a discrete numerical code.

There are several considerations that affect the quality of sound that is recorded and reproduced in this way. First, you need to measure the amplitude frequently enough to capture all the ups and downs in the orignal chart. (For instance, if the original sound repeated the same 1/1000 second cycle over and over again, and you only measured 1000 times a second, then you would get the same amplitude measurement with every sample. The result would be silence.) In other words, the higher the sampling rate, the better the result. Second, you need to record the measurements accurately enough to capture very small variations in the amplitude. This means that you have to use a sufficient number of digits, or bits, to record each sample---the more bits per sample, the better the sound quality.

CD-quality recordings sample at a rate of 44100 hertz (44100 times per second). Each sample value is represented by a 16-bit (2-byte) string, which means that up to 2^{16}=65536 different amplitude levels can be distinguished. Now since CDs are stereo recordings, there are actually 88200 2-byte samples for every second of music. This means that one hour of recorded music will require

2 channels x 44100 samples/sec x 2 bytes/sample x 3600 sec = 635,040,000 bytes.

This will all (barely) fit on one CD. A compact disk is an amazingly dense medium for storing information (and a DVD is even denser). To record data on them, a powerful laser burns a tiny pit in the surface of the disk for each bit of the encoded data equal to 1. (If the bit is zero, the surface is not altered.) To recover the recorded bits, a weaker laser is fired at the disk's surface: the pits scatter the light and the unpitted areas reflect the light.

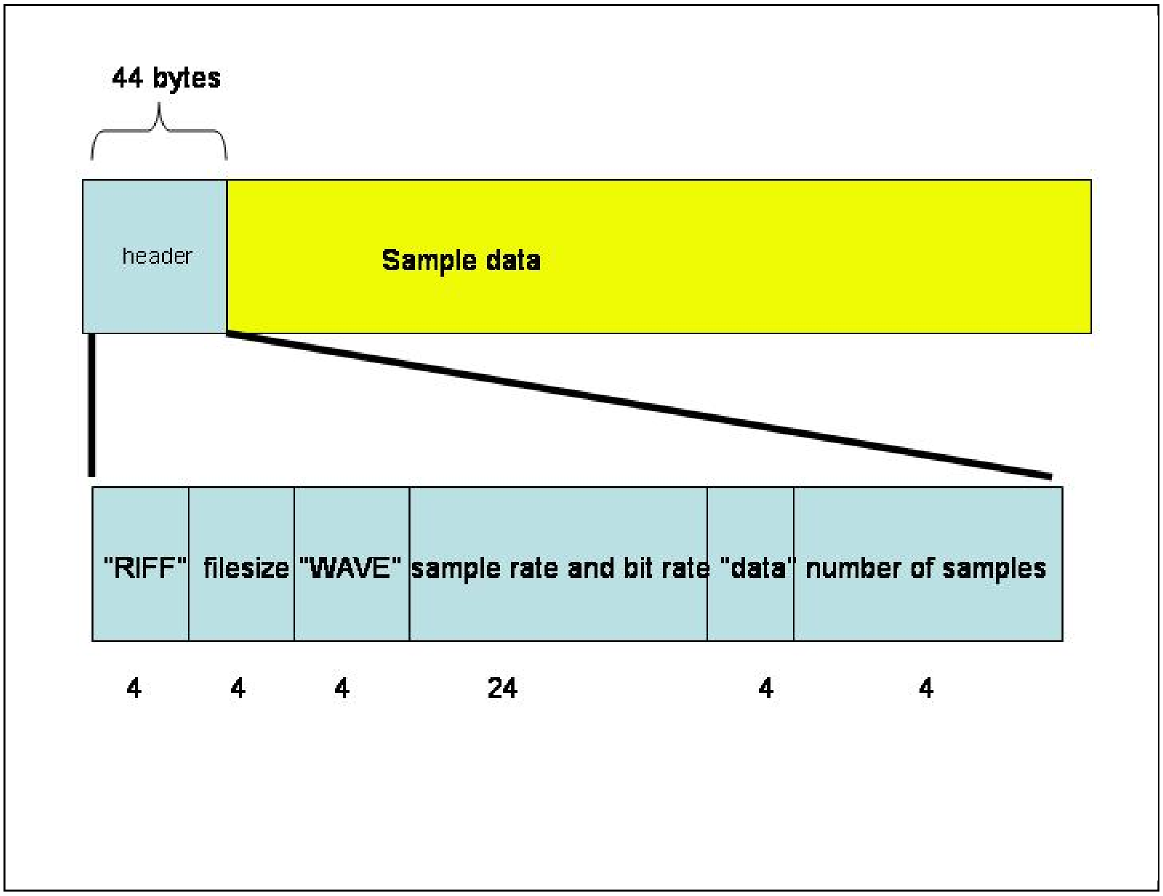


Pits on the surfaces of a CD and a DVD .

Audio Files

The digital sound recordings that you will experiment with in class are monaural (rather than stereo), sampled at 11025 hertz, using 8 bits per sample. (By the way, that explains the numbers on the y-axis of the French Horn wave. The chart was plotted from the data in a digital recording of the horn. Each sample value was encoded by one byte, and hence as a value between 0 and 255.) Since we have half the number of tracks, one-quarter the sampling rate, and half the number of bits per sample as a CD, an hour of music will take up only one-sixteenth the amount of storage a CD-quality recording requires, about 40 MB. The sound quality is noticeably lower, but still quite reasonable.

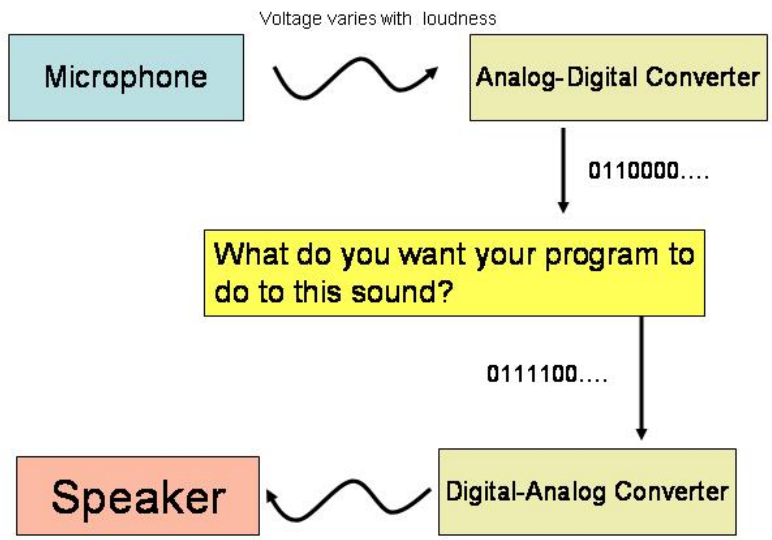
These sequences of bytes are stored in "wave files", whose names have the extension ".wav" in your computer's file system. Each wave file begins with a brief "header" that contains information about the sampling rate and number of bits per sample. The header contains both ASCII text (the words "RIFF", "WAVE" and "DATA") and integers encoded as 32-bit sequences. The sample data contain integers encoded as 8-bit sequences. But, like all information contained in a computer file, it's all bunch of bits.



Format of a .wav file, with an expanded view of the 44-byte header.

Of course, to get any of this to work, you need equipment that can turn the sequence of bits back into a fluctuating electrical signal. This is the job of a digital-to-analog converter: your CD player has one, as does the "sound card" in your computer.

As was the case with the temperature reading in our digital thermometer, once the audio signal is translated into numbers, you can use a program to manipulate these numbers so as to alter the sound: speed it up, slow it down, remove noise, boost the bass, etc.



For instance, the program could read a sequence of sample values, like

23,23,27,29,32,28,24,20, etc.

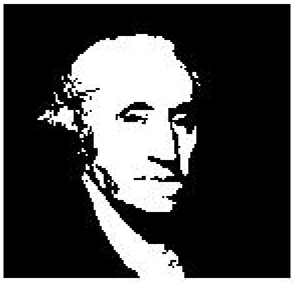
and insert the average of each pair of successive numbers in between them:

23,23,23,25,27,28,29,31,32,30,28,26,24,22,20

The effect, on the chart of the sound wave, is to stretch it horizontally by a factor of two. The resulting samples, when played back, would be a slowed-down, lower-pitched version of the original sound. You could play the sound backwards by just feeding the numbers to the Digital-Analog Converter in the reverse order. You could create synthetic sounds without a microphone, simply by generating the appropriate numbers, perhaps by some mathematical formula. *You can manipulate the sound by manipulating numbers.*

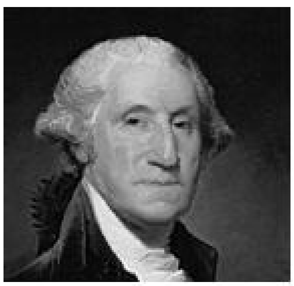
Digital Images

A typical computer display is divided into several hundred thousand tiny rectangular patches called *pixels*. An image is rendered by painting each of these pixels, so a digital image can be formed by sampling the original image or scene at points corresponding to each of these pixels and encoding the measurement as a string of bits. The simplest thing we can do is encode each pixel by a single bit, which has the value 1 if the corresponding place in the scene exceeds a certain brightness threshold, and has the value 0 otherwise. When we reproduce this image on the screen, all the 0 bits are painted black, and all the 1 bits are painted white:



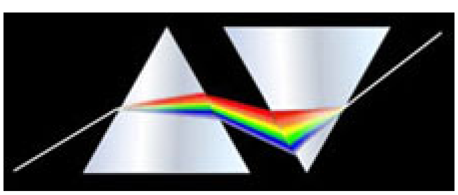
The illustrated image is 144 pixels wide by 140 pixels wide.

A far more satisfactory result is obtained if we specify the brightness more precisely than just "bright" versus "dark". If we encode the brightness of each pixel by a byte, rather than a single bit, we get 256 distinct brightness levels. This is called a *grayscale image*:



How should we encode the *color* of each pixel?

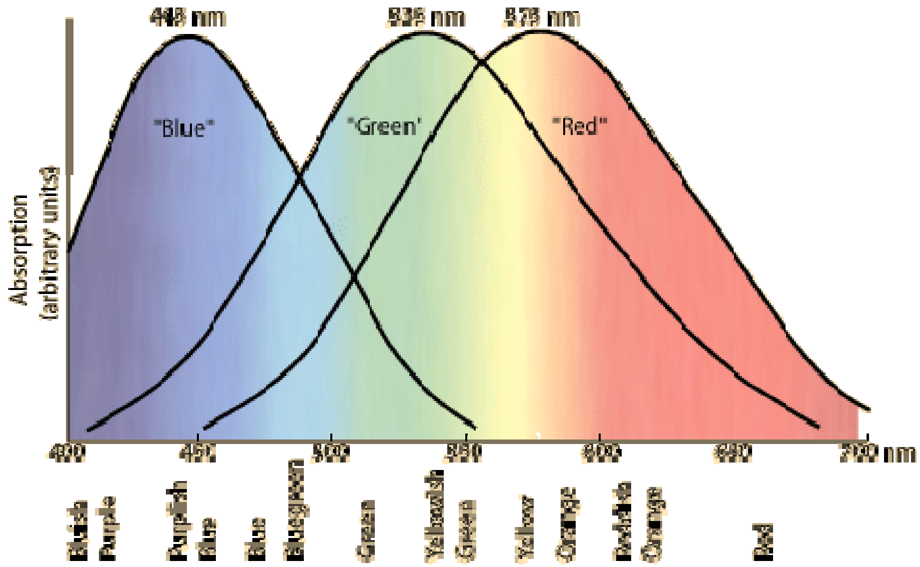
It's helpful to understand a little bit about how color vision works. White light from a light bulb, or the sun, is a mixture of light at many different colors, or *wavelengths*. You've probably all seen the experiment in which white light is separated by a prism into its component colors---this occurs because different wavelengths of light are refracted at different angles. The diagram below shows a version of the experiment in which the light is sent through a second prism that recombines the different colors into white light.



Now let's do a little thought experiment. Suppose we blocked the light leaving the first prism so that all that was allowed to pass was a narrow band in the yellow part of the spectrum. The beam emerging from the second prism would appear yellow, as it is made entirely of light whose wavelengths lie in that little band.

If instead we blocked the first prism so that only a narrow strip of green and a narrow strip of red were allowed to emerge from the first prism, the beam emerging from second one would still appear yellow! The yellow light produced in this second experiment looks the same as that produced in the first, but they are different, as can be proved by adding a third prism to the experiment, oriented just like the first one. The second beam of yellow light will now be separated again into its red and green components, but the first remains yellow.

Why does this happen? Specialized cells in the retina of your eye, called cone cells, have different responses to different wavelengths of light. There are three kinds of cone cells. One kind is most sensitive to a certain wavelength of red light---the more that the wavelength of the incident light differs from this ideal value, the less the response of the cone cell. The chart below shows the responses for each of the three types of cone cells.



Pure yellow light, with a wavelength of about 570 nanometers, provokes a large response from the red-sensitive cone cells, a somewhat smaller one from the green-sensitive cone cells, and none at all from the blue-sensitive ones.. But we could produce a very similar physiological response by stimulating the two cells with a mixture, in the proper proportions, of green light at 525 nm and red light at 650 nm. (It won't be exactly the same as the spectral yellow light, because our mixture will also cause a slight response from the blue cone cells.)

This suggests that we can simulate all the colors of the spectrum, or very nearly so, by mixing appropriate proportions of red, green and blue light. That's not entirely true, because of the way the three curves overlap---some spectral colors cannot be synthesized exactly using three "primary colors", no matter how you choose the primaries. But we can come very close.

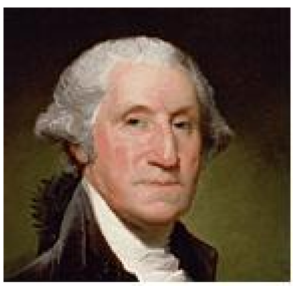
Thus color-imaging systems, whether analog or digital, simulate the eye's response to an image by passing light from the image is passed separately through red, green and blue filters, and three separate brightness measurements are recorded.

Bitmap Image Files

In digitized images, each of the three brightness measurements is encoded by a string of bits. The result is a "bit-mapped" color image. In the files that you will work with (Windows .bmp files) in the labs, each pixel is encoded by three bytes. Each byte represents the brightness of one of the three components of the pixel. A 144 x 140 image, for example, will thus require

$144 \times 140 \times 3 = 60480$

bytes of storage. There are certain conventions about which of the three bytes associated to each pixel corresponds to red, which to green, and which to blue, as well as how the sequence of pixels encoded in the file is oriented in the picture. (Does the first pixel represent the upper-left corner of the image? Are the pixels arranged column by column, or row by row?) You will investigate these in your experiments. In addition, the file has to contain some information about the dimensions of the picture, since, for example, a 144x140 image contains the same number of pixels as a 240 x 84 image. Odd distortion results if you display the image using the wrong dimensions.



The original image

The same bitmap displayed at 240 x 84 pixels

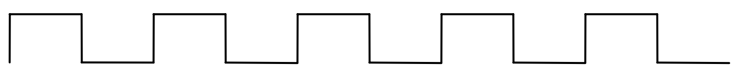
Because all digital data is encoded as bits, there is nothing intrinsically about the values in the file to tell you that this is a picture, or a sound, or something else. (In fact the first few bytes of both wave files and .bmp files are ASCII text.) You could, if you wanted, treat the red, blue and green intensity values of pixels as audio samples, and play the result (which will surely sound like noise).

Something may be troubling you at this point. If you use an iPod or an MP3 player and check the sizes of the music files, you will find that an hour of music takes up far less space than the 635 MB we calculated above. And if you download image files from the Internet, you are probably aware that tiny color images like the ones shown above take up considerably less than the 60KB we find in a .bmp file. In the next set of notes we will take up this question of how information is compressed.

Notes.

Digital toaster, digital steam iron, digital coffee pot. Yes, they all exist. I found them on Google, offered by the usual big appliance manufacturers. About fifteen years ago, a colleague of mine who was teaching a hardware-design course showed me an exam problem he had written involving a computer-controlled toaster. I thought it was a joke. Of course, when *you* hear “digital coffee pot”, you might be asking, “What other kind is there?”

Are all watches and clocks digital? This footnote will make a little more sense after you’ve studied the material on digital circuits. At the heart of all watches and clocks made today is something called a crystal oscillator. This produces a fluctuating electrical signal, whose voltage, if you plotted it, would look like this:



Typically, the frequency of this signal is 32768 hz., exactly 2^{15} fluctuations per second. You can think of the signal as a sequence of bits: 1010101010….The signal is processed by a simple circuit that executes the following program, keeping track of two bit values.

1. Set Bit1 to 0
2. Set Bit2 to 0
3. Wait for the next incoming bit.
4. Change the value of Bit2.
5. If Bit2 is 1, change the value of Bit1.
6. Output Bit1.
7. Go to 3.

The result is the sequence 110011001100…… If we pass the resulting signal through 14 more divide-by-two phases, we get a nice 1-hertz pulse. Old-fashioned mechanical clocks take the 1-hertz pulse from a pendulum or a spring and use toothed wheels to divide by sixty in two successive phases, producing a once-per-minute and once-per-hour pulse to drive the hands of the clock.

Invention of phonograph. A French poet and inventor named Charles Cros, independently had the idea of the phonograph, and described a design very similar to Edison’s in a letter addressed to the Academy of Sciences of Paris in 1877. Cros never built his machine.

Why is the label on the disk recording so big? The groove on a disk recording is a very long spiral that starts at the outside edge of the record and travels inward. The motor turning the disk rotates at a constant rate, but the inner tracks of the record have, of course, a smaller circumference than the outer tracks. This means that the fluctuations in the inner tracks have to be packed closer together than those on the outer tracks, and recording very close to the center of the disk is impractical. The labels got smaller by the time I started listening to records (the one in the illustration is from 1904). Edison's cylinder records did not have this quirk.

Incidentally, the music on a CD starts at the center of the disk, and finishes at the edge. The motor in the CD player turns at a variable rate, so the spacing of the pits can be the same everywhere on the disk.

Pitches and Frequencies. The table below gives the frequencies, in hertz, of standard musical pitches. Middle C is the C in octave 4 (261.26 hertz)

Note	Octave 0	Octave 1	Octave 2	Octave 3	**Octave 4**	Octave 5	Octave 6	Octave 7
C	16.351	32.703	65.406	130.813	**261.626**	523.251	1046.502	2093.005
C#,Db	17.324	34.646	69.296	138.591	**277.183**	554.365	1100.731	2217.461
D	18.354	36.708	73.416	146.832	**293.665**	587.330	1174.659	2349.318
D#,Eb	19.445	38.891	77.782	155.563	**311.127**	622.254	1244.508	2489.016
E	20.061	41.203	82.407	164.814	**329.626**	659.255	1318.510	2367.021
F	21.827	43.654	87.307	174.614	**349.228**	698.456	1396.913	2637.021
F#,Gb	23.124	46.249	92.449	184.997	**369.994**	739.989	1474.978	2959.955
G	24.499	48.999	97.999	195.998	**391.995**	783.991	1567.982	3135.964
G#,Ab	25.956	51.913	103.826	207.652	**415.305**	830.609	1661.219	3322.438
A	**27.500**	**55.000**	**110.000**	**220.000**	**440.000**	**880.000**	**1760.000**	**3520.000**
A#,Bb	29.135	58.270	116.541	233.082	**466.164**	932.326	1664.655	3729.310
B	30.868	61.735	123.471	246.942	**493.883**	987.767	1975.533	3951.066

Digital thermometer block diagram. Of course this is oversimplified. Even if we didn't compare the current temperature read to the last one, There still has to be some sort of program for processing the bits coming from Analog-to-Digital Converter before sending them to the display. (Among other things, the output has to be converted from a binary encoding of the voltage to a decimal encoding of the corresponding temperature.)

"This would be like taking one of Edison's phonograph records and measuring the depth of the indentation at many different locales, then throwing away the record and just keeping the measurements." This sounds like some sort of retro conceptual art, but it's actually been done to recover the audio from old recordings without damaging the record! Here's the URL of an article in the San Francisco Chronicle describing this research---the newspaper article contains links to the researchers' papers.

http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2004/07/12/MNGJP7JRC21.DTL