Digital Music and Data Files

What is this MP3 stuff, anyway?

Barry Johnson – July 2008

Sound Waves

Before we can talk about digital music, let's establish some foundations – let's talk about sound.

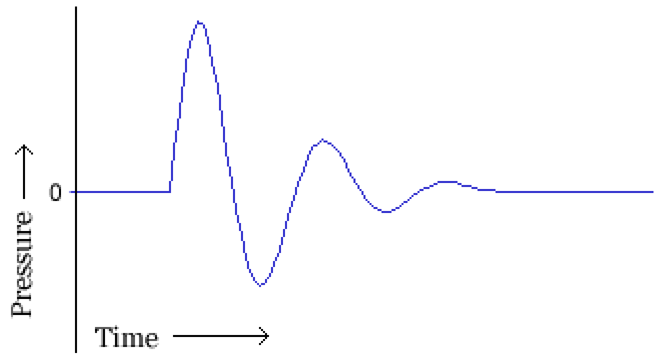
As you know, sound is waves of compressions in air (or water or any other medium). It's hard to show pictures of compressed air, so most people use an analogy to describe sound waves: They'll talk instead about the kind of waves you might see on water:

We can see these waves starting from one point in the middle and flowing out in all directions. If we were to drop a buoy into our water at one place and graph the height of the water at that one spot over time, we would get a picture that looks somewhat like the surface of the water: a curve smoothly connecting alternating high and low points.

Sound works in much the same way: Something like a speaker or handclap causes air to be compressed, and that compression wave travels outward in all directions. If we measure the pressure at one point, we'll see that pressure change over time: it will vary smoothly from high point (more compressed) to low points (less compression), and a graph of those pressure changes as time goes on can also look much like the surface of the water.

Let's look at a graph of sound compression waves from a handclap:

Questions, Comments and Corrections are appreciated!
You can reach me at callerbear@gmail.com



At the far left, the air is at its normal atmospheric pressure – for sake of simplicity, we'll just say it has "zero compression". As time moves on, we see a sharp rise in the amount of the compression (the highest point in the graph, corresponding to the actual clap itself), then a sharp fall to decompression (the lowest point), and then a bit of a bounce following until the air settles back down to its normal pressure.

If you wanted to produce the sound of that handclap using a speaker, you'd have to first push the cone of the speaker quickly towards the listener (compressing the air), then pull it backwards (decompressing), and then "bounce" the cone a little bit.

In fact, the graph of the pressure wave can also be viewed as a graph of the position of the speaker cone over time: we'd push the cone out and back creating the high and low pressure compression waves as time goes on – and the distance of that motion corresponds exactly to the amount of air compression that the speaker is producing.

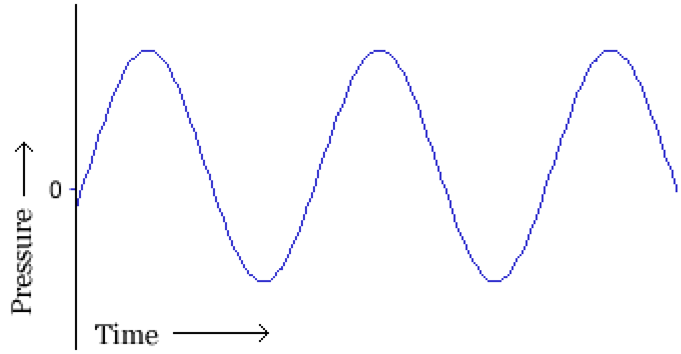
The pitch of a sound is measured by the frequency at which sound waves crash against your eardrum: as the waves get closer together (hitting your ear more frequently), we consider that to be a higher note – as they stretch apart, we think of it as a lower note. The height of the wave corresponds to the volume of the sound – the stronger the compression, the louder the sound appears to us. Correspondingly, a small or short wave represents a soft sound.

Digitized Sound

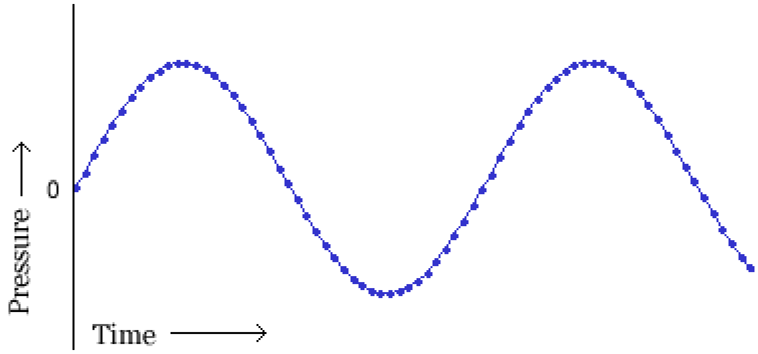
These sound waves are very "smooth". Going back to our water wave analogy, no matter how closely we look at the surface of the water, we'll never see any "steps" in the water – it's always a smooth slope. Sound waves are the same: the compression smoothly changes from one level to another: No matter how close we look, we won't see "steps" in the wave. The amount of compression can change *quickly*, but if we look close enough, we'll see that it always changes *smoothly*.

Electronic equipment that can accept or produce very smooth signals like this is called "analog" equipment – the electrical signal being used (usually voltage level or current flow) becomes an "analog" of the original sound. Unfortunately, analog information is virtually impossible to store electronically, so today's computers make a "digital" (or "digitized" version of that analog information.

Let's look again at a series of sound waves:



To convert this analog wave into a series of digital numbers, an "analog to digital" circuit will measure the value of the wave at many points in time:



Each pressure measurement is recorded as a numeric value, and this series of numeric values is stored to represent the waves. Other specialized circuitry ("digital to analog converters") will accept these streams of numbers and produce an analog wave that looks as close to the original as possible.

There are two important things that affect the quality of this conversion from analog to digital and back: The precision with which you measure the values along the original analog wave, and the number of points you measure each second. If, for example, you measured the wave using only the digits 0 through 9, each point could be off by as much as 5% from its true value. On the other hand, if you measured using values 0 to 999, then each value will be within .05% of the true value – a much closer measurement.

CD-quality sound is measured using values from 0 to about 65,000, so each point is fairly close to its true value. This range of possible values is called the "sample size". It takes 16 binary digits of data to hold the values 0 to 65,000, so this is called a "16-bit sample size". Audacity supports the CD-quality 16-bit samples, as well as 24-bit samples (accurate to one part in 8 million) and a "32-bit float" format that sacrifices a little quality but supports extremely large ranges in volume.

The frequency at which you sample the sound affects the quality. If you "sampled" the sound only once every second, then you would likely miss any waves that last less than a second – and if you measured 100 times per second, you could still miss any waves that last less than 1/100th of a second.

The most sensitive human ears can hear sound frequencies from about 20 waves per second (20 "hertz"[1], often abbreviated to "Hz") to about 20,000 waves per second (20,000 Hz). If we don't want to miss out on any waves that a human could hear, we need to measure sound at least 20,000 times per second – in fact, to be sure, we need to measure at least twice that frequently.

For technical reasons, the engineers that designed the CD audio format chose to sample the sound 44,100 times per second – about 10% more than twice as frequently as the highest audible sound waves.

This turns out to be a great deal of information: If we use 16 bits (in computer speak, this is 2 "bytes") for each measurement, and we have 44,100 measurements every second, then we are creating 88,200 bytes of data every second, or about 5.3 million bytes of data each minute. A typical singing call lasts about 3:50 – which means each song we store on our computer would take just over about 20 million bytes (which works out to about 19.3 MB in computer-speak).

The digital information on an audio CD is basically this uncompressed stream of two independent tracks (stereo right and left tracks) of 16-bit samples performed at 44,100 times per second. Since a raw CD can hold about 670MB of data, this works out to about 70 minutes of audio on a standard CD.

Compression Techniques and MP3

One technique to store sound information in a disk file is to simply write out the long string of 16-bit sample values, one after another after another. Sound contained in "WAV" format files is stored in essentially that format. WAV files quite large (about 5MB per minute per track), but are the easiest (and most accurate) file format for software programs to use.

Back in the late 1970's when the standards of a 16-bit sample size and 44,100 sample frequency were being developed, digitized sound was a *huge* amount of data -- far more than could be comfortably stored in computers or transmitted across wires. One of the approaches to reducing the amount of data was the idea of "compression": being able to take data and "squeeze it down" to a smaller size. For example, with a large number of samples being taken per second, the average *difference* between sample values is fairly small: if one recorded a starting value, then perhaps it would only take 1 byte of data to represent the *change* from one sample to the next (instead of 2 bytes to hold the full value

[1] Named after the German physicist Heinrich Hertz, who was the first to experimentally transmit, receive and reflect radio waves. When asked about his results, he said "It's of no use whatsoever… we just have these mysterious electromagnetic waves that we cannot see with the naked eye. But they are there." Shortly afterwards he became seriously ill and died a few years later of blood poisoning at the age of 36. He never knew that he had developed the basis for today's radio, television and radar.

of the sample) – meaning that we could represent the same values in about one-half of the number of bits.

Many different and sophisticated techniques for compressing sound have been developed. The most popular compression algorithm is called MP3 (very short for "Moving Picture Experts Group Standard 1 Audio Layer 3"). This algorithm can squeeze a typical 20 MB song down to about 2.5 MB in size, but introduces some (usually imperceptible) "fuzz" into the sound when it's played.

Think of a damp sponge: you can squeeze the sponge down a bit without losing any water. But if you squeeze it too much, some of the water will drip out of the sponge, so that the sponge isn't quite as wet when you release it. "Lossy" compression techniques, like MP3, can lose a little bit of the detail in the music if you compress it too much.

If the sound is very rich – perhaps it has very sharp notes, especially if they've very high or very low – it's more like a sponge that's very wet: you can only squeeze the sponge a little bit without loosing water. When you compress vibrant music with MP3 these very rich areas can become slightly distorted, inserting unwanted "artifacts" into the music. One type of artifact is a "pre-echo" – you might hear a faint clink of castanets just a moment before the castanets actually appear in the music. Another form of artifact is a loss of very high-frequency sounds – the waves are just too close together for MP3's compression algorithm to properly handle.

On the other hand, fairly simple sounds (like a person speaking) can be compressed quite far without noticeable loss of quality.

Because of this, MP3 compression allows you to control the amount of compression it will do. This is known as the "compression bit rate", and is usually a value like 64 Kbps (64,000 bits per second), 128Kbps (128,000 bits per second), or 192 Kbps (192,000 bits per second). 192 Kbps is generally considered to be CD-quality sound, but most music purchased digitally and downloaded across the Internet is compressed to 128 Kbps. 64 Kbps is probably too compressed for use in square dance calling.

Values of 16 Kbps and 256 Kbps are also available – 16 Kbps is far too compressed for our music (it would sound like you're playing your music across a telephone). The difference between 192 Kbps and 256 Kbps is very slight – only the most trained ears, listening carefully, would likely be able to tell the difference in quality.

In the last few years, MP3 encoders and players have begun to support a variable bit rate, where the compression bit rate varies as the dynamic range of the music changes. Variable bit ratios generally result in smaller files while preserving good quality, and are fairly well supported on today's computer systems, as do the majority of MP3 players sold since late 2006. However, older MP3 players may not support this variable rate, so you'll want to test your devices.

Several companies claim to hold patents on the principles used in MP3 compression. While these companies do not attempt to collect royalties from people using MP3 files, they *do* attempt to collect royalties from folks that write software that will be used to *create* MP3 files. (There is no intention to collect royalties from the users of such software, only from the developers of the software itself). This outstanding patent issue is the reason that the Audacity music editing program doesn't directly support creation of MP3 files – requiring you to separately download the LAME MP3 encoder.

Other Compression Types (WMA, AAC, OGG and M4P)

As digital music matured throughout the 1990's, newer compression techniques were developed. Microsoft created their proprietary "Windows Media Audio" (WMA) technique. A successor to MP3 known as "Advanced Audio Encoding" (AAC) was adopted as a standard[2] and used by Apple for their iPod product line and iTunes store.

A group of software developers created yet another type of compression algorithm that expressly is *not* restricted by software patents. This compression goes by the unusual name of "Ogg Vorbis", and files on Windows systems in this format will have an extension of OGG.

At lower bit rates like 64 Kbps, these other standards produce considerably better quality of sound than MP3 compression, meaning that you can have smaller files while maintaining the same quality. At 128 Kbps the case is less clear – different independent studies have come to conflicting conclusions on quality. At 192 Kbps, the differences become virtually impossible to hear, even for skilled listeners, but the OGG, WMA and AAC encoders will still produce somewhat smaller files than MP3 encoders.

WMA and AAC have both been extended to support "Digital Rights Management" – software support for the concept that copyrighted music should not be freely distributed by individuals. Apple's DRM-extended version of AAC uses a file format called "M4P" – most songs downloaded from iTunes will have a file extension of M4P[3].

What Compression Should I Use?

The various formats have different benefits and compromises.

WAV files have the best quality of all because they're completely uncompressed. Each square dance record will take about 20 MB of disk space – that's between 5 and 10 times larger than the other formats. On the other hand, today's computers have so much disk space available that 20MB isn't nearly as formidable as it was just a few years ago. If

[2] AAC is actually Part 7 of the broad MPEG-2 format, which eventually became the standardized format for encoding DVDs. While AAC is a standard, it too includes algorithms that are patented by various companies, which means that software developers have to pay someone in order to use it.

[3] The names just get worse. The MPEG-4 standard has 23 different parts covering video, audio and many related types of media. Part 14 of the MPEG-4 standard defines the MP4 file format, which is often used for storing movies as files. However, Apple decided that when they sent out MP4 files that contained only audio tracks, they would use an extension M4A – even though it's exactly the same as MP4. Apple then decided that their MP4 files protected by DRM would be called M4P. Do they do this just to confuse us?

your computer has, say, 20GB of space available, you could still record 1,000 different songs – far more than you can practically use on a day to day basis.

WMA and M4P are formats designed to be used with copyright-protected software purchased from the Internet. Several of the software packages commonly used by callers (such as SqView) do ***not*** support the copy-protected versions of these format. (See the later section of this paper for techniques to convert a song from these formats to something SqView can use.)

Ogg Vorbis compression produces the best results and the smallest space, but it's somewhat obscure and isn't supported in a number of audio editing programs (although both Audacity and SqView support it). In particular, many of the programs available to create audio CDs (allowing you to play your digital music on a CD player) don't support the OGG format – so you may have to convert your files to a different format before burning a CD.

MP3 is a compromise solution: it offers reasonable compression at reasonable quality and it's very well supported throughout the industry.

Personally, I recommend using MP3 at the 192 Kbps compression rate as the best overall choice.

Buying Music over the Internet

Be sensitive to quality issues when purchasing music over the Internet. Many of the big-box online stores (such as iTunes) will sell their lowest-priced music compressed at the 128 Kbps level, which is somewhat worse than CD quality. With some of the online stores, one can sometimes pay a slightly-higher price to get an "unlimited usage" version of the music – and incidentally, these higher-prices downloads are usually in the CD-quality 192 Kbps compression setting.

Either way, many of the songs you purchase on-line are often available only in a "protected rights" format such as WMA or M4P. When you download these protected songs, you also download a small file that indicates you have the right to play that music on your computer. If you copy the music file to another computer without transferring the rights as well, the music won't play on the other computer.

The most common software programs used by callers will not play music in the AAC format or DRM-protected music in the WMA and M4P formats[4]. On the laptop itself, Windows Media Player will play WMA files, and iTunes will play AAC and M4P files. Most MP3-player devices (like iPods) will play either format as well. But in all these cases, you loose the ability to change tempo or pitch, or to automatically "loop" the music for longer hoedowns.

[4] SqView will play WMA files, but not ones that include copy protection.

Converting from WMA or AAC to MP3

As callers, we are (or should be) ethically committed to the protection of copyrighted music. But when the DRM protection prevents us from using music we have purchased, our ethics allow us to convert the music to a format we *can* use, as long as we don't then take advantage of that conversion and give (or sell) copies of music to other.

The simplest and most reliable technique of converting protected music to MP3 files is to "burn" it to an audio CD, then "rip" the track back off the CD. If the CD burner in your computer can create "R+W" CDs, then you can reuse a single CD for doing this several times. Otherwise, you may wish to do several songs at once at on a single CD, just so you aren't throwing away a CD every time you convert every song.

There are some free software programs that will convert files in these formats to MP3 without using a CD. As you hunt for them, watch out for "shareware" versions – they'll often have a free "demo" version that will only convert a portion of your music (like the first 10 percent of a song), and only convert the entire song if you pay for the software.

Digital Sound Summary

In summary, there are several important points to know:

- Digitized sound approximates true audio by breaking the smooth analog sound waves up into 44,100 tiny steps per second, and recording the "height" of each step. Specialized hardware in computers will re-create the original sound waves from these sequences of numeric values.

- Digitized sound requires relatively large amounts of data: a mono track of audio generates about 5MB of data per minute, while a stereo recording generates twice as much data.

- Specialized compression techniques (such as MP3, WMA and AAC) can "compress" the data for music to a much smaller size, but do so at the loss of some quality within the music. The tighter you compress the music (so it takes less space on disk), the more you will compromise quality.

- When used with a CD-quality compression setting (192 Kbps), MP3 compression produces good quality sound while using only about 20% of the disk space of uncompressed music. Use of a variable bit rates can produce even smaller files while still preserving audio quality.

- Some music stored in WMA, AAC and M4P files may not work with your music player. These files can be converted to MP3 by creating an audio CD then ripping that CD.