The Guide to Computer Audio and Video for Non-Experts

by Deskshare

We all want to stay in touch. How much communication now is done through digital multimedia? If you want to persuade, to demonstrate, or just to amuse, the odds are you'll want to use audio, video, or both.   If you have to go beyond the basics, it's a good idea to learn something about the underlying technology. Everyone making an MP3 file with Digital Media Converter doesn't need to be an audio engineer—or even if you're just editing your home movies with Video Edit Magic, a basic knowledge of file formats can help you share your masterpiece with others. This guide begins by explaining how video is stored and compressed, and then audio. If you have any questions or comments on this document, please let me know at writer@deskshare.com.

Contents

Contents

Introduction:

Moving pictures go back to the 19th Century. For over 100 years, making a movie meant using expensive, hard-to-use equipment and a team of cameramen, technicians, editors, and many others.

These days, video can be created using an inexpensive camcorder and edited on your PC. Moving pictures belong to everyone! You've probably made, or thought about making, your own films. If you're making video on a computer, it helps to understand something about how computers produce and store moving pictures, so you can make the correct choices when it's time to distribute your masterpiece.

Background

All moving pictures, from the old zoetropes through celluloid films and on to the screen on a video iPod, work on the same principle. The device projects a series of still pictures, and changes from one picture to another very quickly. If the static images are swapped quickly enough, the human eye and brain interpret what they see as movement. Computer video at its most basic is a series of images stored in a single file. (See MJPEG).

How fast do the images have to update in order to create an illusion of movement and avoid flickering and jumpy motion? The commonly accepted number is 16 updates per second. (This is normally written as "frames per second" or "fps".) Most commercial video systems update significantly faster than this. Theatrical films change images 24 times per second, European TV at 25 fps, and US and Japanese TV at about 30 fps. Computer video is often distributed with a lower frame rate, as low as 5 fps, depending on the requirements.

Computer Video

So, computer video is just a series of still pictures. However, if that were all there was to it, the files would become enormous. One uncompressed still picture from a medium-quality camera can be 24 megabytes in size. Record 30 per second for 10 minutes (600 seconds) and you have a 14 gigabyte file! Clearly, there is a need to compress these files.

One compression method (called intraframe compression) is just to compress each still picture. For example, a file compressed using MJPEG (Motion JPEG) consists of a series of still pictures, but each picture is compressed using the JPEG system, the same method used when making the .JPG files that are so common on the Web. JPEG compression can make each image only 10-25% of the original size (depending on the exact settings). Essentially all digital video uses some sort of intraframe compression. However, this can still leave you with very large video files.

The second powerful technique in video compression is what's known as interframe compression. Interframe compression relies on the fact that most video frames are very similar to the frame before. For instance, if the camera is focused on a woman sitting in a chair, talking, the part of the picture that shows the chair, her legs and body, and the wall behind her, might not change for several seconds at a time, which could represent over 100 frames of video. Instead of encoding 100 separate still pictures, advanced computer video can store just one still picture, the first frame of this section, and then store only the differences between each frame that follows. The differences are tiny compared to the whole picture, so this can make files much smaller. Almost all computer video uses some form of interframe compression, because it offers such a big advantage in terms of compression ratio.
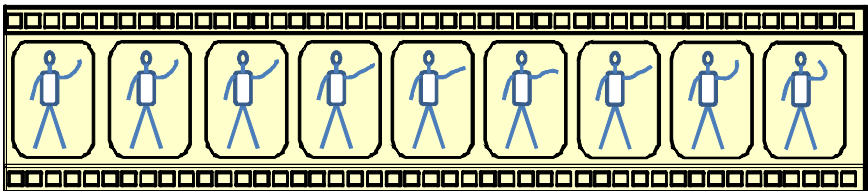
Mr. Abstract waves to you
This is the way film, uncompressed AVI, or MJPEG video files store a movie.
Every single frame is stored completely.
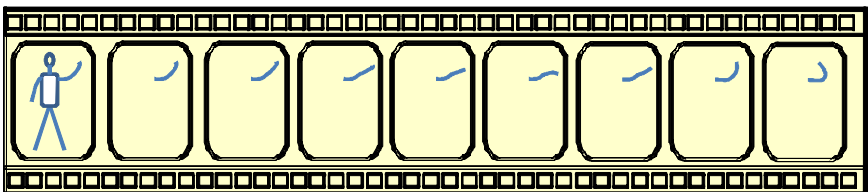
This is Mr. Abstract compressed using Interframe Compression. After the first "full" frame, all that need to be stored are the **changes** between frames, in this case Mr. Abstract's arm moving.

The disadvantages of interframe compression come from the dependency on the small number of full frames that are stored. If a full frame is lost in transmission, there is no way to show the following frames, until another full frame is sent. Interframe compression can make jumping to a particular spot in a video file difficult. If you jump to frame 1229, but the last full frame was 1187, the program has to backtrack, one frame at a time, from 1229 until it finds the full frame, then reconstruct every frame in between, in order to rebuild 1229. For these reasons, most compressors insert a full frame fairly often. (Video on a DVD generally has a full frame every 15 frames.) These full frames are known as key frames.

The graphic below shows the structure of two video files, one using intraframe compression and one using both intraframe and interframe compression. (For simplicity, we aren't showing how these files also store audio information.)

Video Streaming

Downloading a long video can take a long time, ranging from a few minutes to hours, even over a very fast network connection. When you watch TV, you don't switch to channel 43 and then wait 10 minutes for the show to start. It just starts.

Video streaming lets you start watching without waiting for the file to download. It's a very simple idea: as soon as the file starts coming in, begin playing the frames you already have, without waiting for the end of the file.

There are a few refinements, tricks that make the viewing experience better. One is buffering. Internet connections are not totally consistent. Sometimes there are brief pauses. If your computer displayed every frame of video as soon as it came in, whenever the connection paused, the video would pause. Instead, the media player lets the downloaded frames "get ahead" of the playing, keeping a few seconds of extra video and audio in a "buffer". If five seconds are buffered, even if your Internet connection freezes for four seconds, the video plays seamlessly.

Some formats are specifically designed to work well as streams, including Windows Media (WMV), MPEG4, and Flash Video. Some others are very rarely streamed, such as AVI.

Container vs. Compressor

Video can be stored in several different file formats, and also compressed several different ways. Unfortunately, the two often get mixed up. A container format is a file format. Examples are AVI, ASF, and QuickTime (MOV). Actual compression methods are mathematical descriptions of ways to reduce video and audio file sizes. Examples of this are Cinepak, MPEG 2, and Sorenson Video. The confusion comes because sometimes a compression method is used as a filename, or the same standard defines both the file format and the compression method. For instance, the popular audio MP3 format is both a file format and an audio compression method. (The name "MP3" does not mean "MPEG 3". It comes from the place where the compression method is defined: Part 3 of the MPEG 1 document, describing audio layer 3 compression.) Sometimes a file format (container format) and a compression method have similar names, which can cause confusion. For instance, "Windows Media Video" is a file format (usually called "WMV files"), but "Windows Media Video 9" is an advanced method for video compression. Generally, once you understand the concepts of container and compressor it's easy to tell the difference.

Compressors

Video

In this section we'll discuss some popular video compression systems for video files. After reading this, you should be able to make decisions about what compression system to use for your own files.

MPEG 1 and 2

The Moving Picture Experts Group has created a series of standards for movie files. The first one, MPEG 1, is still used. While video compressed with MPEG 1 doesn't have the quality of MPEG 2 or later standards, the actual task of decompression is relatively easy, making this format suitable for older or slower equipment. MPEG 1 is the compression method for VCDs, and can be used on DVDs (but very rarely is). MPEG 2 is the standard compression method for DVDs and is one possible method used on Blu-ray Disc and HD-DVD.

MPEG 1 and 2 compress each full frame using a JPEG-like system, and for all other frames store only the differences from the last frame. There are two major differences: MPEG 1 can only store two sound channels (stereo), while MPEG 2 supports multi-speaker systems like Dolby Audio (6 speakers). Also, MPEG 2 files can produce higher-quality video than MPEG 1. On the other hand, MPEG 2 video compression requires more processing, so it can be slower and/or require more computational resources. MPEG 1 video is more suited to older or slower equipment.

Both MPEG 1 and 2 produce larger files than newer standards like MPEG 4. On the other hand, because they are so old and because they are used on standard media like DVD, both are exceptionally compatible and MPEG 1 and 2 files can be viewed on almost any computer.

Digital Media Converter includes specialized profiles for DVDs and VCDs, which can't be modified. This guarantees that you won't accidentally create files that a DVD player can't play back.

Uses

1. DVD: all standard DVDs use MPEG 2 video and audio compression.

2. VCD: this older optical disc standard uses MPEG 1 compression.

3. MP3 files: these very popular audio files for portable devices and streaming are defined by the MPEG 1 standard.

4. Cable TV On-Demand Video: most cable and satellite TV providers use MPEG 2 video compression.

5. Digital Video Recorders: both TiVo and Microsoft's Media Center use a version of MPEG 2 compression to store recorded TV.

MPEG 4 / H.264

This newer set of standards from the Moving Pictures Experts Group improves on older MPEG standards in several ways. (Note: the MPEG 3 standard was never completed or used.) It offers better compression (smaller file sizes), excellent quality, and several new capabilities, including built-in support for streaming.

The document that defines MPEG 4 compression and files is technically very complex, with 23 parts covering everything from video compression to font formats. Two parts of the document deal with video compression. Part 2 is a video compression codec, also known as the Advanced Simple Profile (ASP). Part 10 is the Advanced Video Codec (AVC), another compression codec that gives better video quality and file size, at the cost of being more complex and slower to implement. It's correct to speak of a video file compressed using either ASP or AVC as an "MPEG 4" file. However, files that end in "MP4" are always compressed using a variation of the Advanced Video Codec. (See the section on file formats.)

The MPEG 4 standard describes compression methods, but it doesn't force programmers to use the standards in any particular ways. There are several programs that use MPEG 4 compression, which are not identical and may not produce the same quality video and audio, even though they're all called "MPEG 4" compressors.

Different organizations have implemented these two standards under different names. DivX and Xvid, for instance, are codecs that implement the Advanced Simple Profile, letting Windows computers compress and decompress video. On the other hand, Apple's QuickTime includes its own method for using the Advanced Video Codec, and video downloaded from their iTunes service is encoded using this system.

Another name for the Advanced Video Codec is H.264. The standard is jointly maintained by two authorities (the Moving Pictures Experts Group and the International Telecommunication Union). They each gave the same set of rules their own name. ("H.264" is the ITU's term for it.)

MJPEG / "Motion JPEG"

Most computer users are familiar with the JPEG file format, used to store still pictures, especially photographs. Many digital cameras create files in this format. A Motion JPEG video file (usually an AVI file) is literally a series of JPEG images. There is no interframe compression. Each frame is compressed and stored individually. Because of the lack of interframe compression, MJPEG files tend to be much larger than other compressed video files.

There are two primary reasons to use MJPEG compression even though it produces large files. First of all, because it doesn't do as much processing on each frame (on average) as an MPEG or other more advanced compressor, it can be useful when a video stream must be viewed or compressed as it comes in (with no time for a computer to convert the video in advance), because it's easier for slower processors to "keep up" with the stream of images. This is the reason that many digital video cameras produce MJPEG files, and that many webcams produce MJPEG streams.

The other advantage of MJPEG files is that, because all of the frames are fully stored, it is easy to "jump" to any frame. There will never be a need to reconstruct multiple frames in order to view (or cut at) a particular point in the video clip, as is needed when editing MPEG files. This is why some editing programs use MJPEG files.

As computer hardware gets faster and cheaper, the need to use MJPEG will diminish.

There are several different MJPEG standards created by different manufacturers, so there is no guarantee that a file created using one program or camera will work with another.

DV AVI or DV

The Digital Video AVI (DV AVI) compression system works very much like MJPEG. Each individual frame is compressed, but there is no interframe compression. Generally, DV-AVI files are smaller than an MJPEG file of the same video.

There are two types of DV AVI file. Type 1 files are slightly smaller, but not compatible with older versions of Windows that use the 'Video for Windows' system. Type 2 files are larger, but marginally more compatible. In practice there is little reason to pick one or the other unless you are using software that works only with one type. DeskShare's Digital Media Converter lets you select either, defaulting to Type 1.

In general, DV AVI is used for the same type of application that MJPEG is: digital video cameras and some editing software.

Sorenson

The Sorenson Video Codec was developed by Sorenson Media. It is included in both Apple QuickTime and Macromedia Flash. It seems to work much like H.264/MPEG 4 Part 10. (The details have not been revealed by Sorenson Media.)

Windows Media

Microsoft has created its own audio and video compression codecs.   Their newest video standard, VC-1, is supported by all Blu-Ray and HD DVD players. Microsoft's video codecs are otherwise almost entirely used in Microsoft products (including Windows Media Player, the Xbox 360, and their Zune portable media players).   Windows Media files are very suitable for streaming, and offer a good balance of quality and file size.

Other Codecs

There are many different video compression techniques available. In this article we haven't discussed such widely-known codecs as Intel's Indeo, Cinepak by Radius, and RealVideo, but this listing covers the majority of the video files being created and watched today.

File Formats

This section briefly describes some of the most commonly used file formats for digital video.

Audio Video Interleave (AVI, .avi)

The Audio Video Interleave format was introduced by Microsoft in 1992. The name comes from the fact that an AVI file combines audio and video streams. AVI is a very flexible format and can use many different compression systems for both audio and video. AVI files can be streamed, but in practice this is rarely done.

The AVI standard doesn't work well with modern video codecs like MPEG 4, tending to produce unnecessarily large files. For this reason, there has been a trend to use other formats like WMV or MP4. However, AVI files are still extremely common. The AVI format is also well-suited to editing, and is used by many programs during the editing process, even when the final movie file will be in another format.

DivX (.divx)

DivX, Inc., recognizing that AVI files using their compression codec were inefficient, introduced its own .divx file format, which is an enhancement of AVI. However, this format has not gained much popularity.

MPEG 1 and 2 (.mpg)

The MPEG standards define file formats as well as compression methods. Files using either of these formats will end in ".mpg". It is expected that the media player will detect what type of file it is. Video files on DVDs are MPEG 2 files, but have the extension .VOB.

QuickTime (.mov)

QuickTime video files can also use several compressors, including Sorenson and H.264. QuickTime files are suitable for streaming, and often balance file size with quality well.   The main limitation on .mov files is that to create or play them can require installing special software (the QuickTime player).

MPEG 4 (.mp4)

The MPEG 4 standard defines a file format, as well as compression methods. MP4 is based on QuickTime's file format. As you'd expect, MP4 files are compressed using MPEG 4 (although some use other audio compression systems). There are various specialized forms of MP4 file, for everything from audiobooks to files for the Apple iPod or Sony PlayStation Portable. The 3GP format used by third generation mobile phones is a variation on the MP4 file.

Advanced Systems Format (.asf)

Developed by Microsoft, this file type can be used with any audio and video compression methods. Originally named Advanced Streaming Format, it was specifically developed to be highly compatible with streaming and web servers. The commonest use of the ASF format is with Windows Media codecs.

Windows Media Video (.wmv)

If an ASF video file uses Windows Media compression, that file is called a Windows Media Video file and ends with .wmv. Windows Media is a very flexible file format. Recent versions of WMV can include advanced compression comparable to H.264 (MPEG 4 Advanced Video Codec), or use H.264 itself with high picture quality and small file sizes. While WMV is a proprietary format, almost any computer can play WMV files.

Flash (.swf/.flv)

A video format that has become popular recently is Adobe's Flash video. Flash video is unique, because there are generally two files involved. SWF files are programs, which are run inside web pages. Flash video sites have a programmer design a SWF program that plays video, with the usual controls (Play, Pause, Stop, etc.). The SWF player is automatically downloaded to your PC when you load the web page that includes it. Once it is loaded, the player then connects to a server and begins streaming the actual video file, which is stored in Flash Video (FLV) format. Flash video is designed specifically for streaming (and is rarely seen in non-streaming applications).

Here's a description of a particular video file, showing some of the characteristics:

File:	My House Party.mp4
Type:	Mp4
Video Compression:	MPEG 4 Part 23, also known as H.264 or Advanced Video Codec (AVC)
Audio Compression:	MPEG 4 Part 3, also known as Advanced Audio Codec (AAC)

Metadata

Video (and audio) files can also include other information. Generally, any information in the video file which isn't actually video or audio is called metadata. "Meta" is Greek for "with" (among other meanings). The metadata in a video file usually describes the video and/or audio.

For instance, here's a media player playing a video with metadata.

"When Fish Attack" is the title of this CNN video, and it was included in the file as metadata. If you've ever played an MP3 file you've seen the audio file version: the title and artist information displayed by your player is stored in the File as metadata.

Digital Rights Management / Copy Protection

In almost every nation in the world, the law recognizes that the creators of any piece of audio or video have certain "intellectual property" rights to it. With the coming of digital media files, copying audio and video have never been easier. Many companies that own intellectual property would like to keep the audio and video they have rights to from being copied without their permission. Software and hardware techniques to prevent unauthorized copying and playing are known as Digital Rights Management (DRM). An older term was Copy Protection, but the modern DRM gives much finer control than just "Don't copy this!"

For instance, Apple's standard iTunes downloads of commercial music are covered by Digital Rights Management. If you buy a song from iTunes, you may copy it to any iPod portable music player, and play it on up to five computers. iTunes will refuse to authorize playing on a sixth computer. You can also use the song to make an audio CD. Because of the way Apple's "FairPlay" DRM system works, songs can only be played on computers that run iTunes software—Windows and Macintosh computers.

Various forms of DRM exist for audio and video files. In general, older formats didn't include DRM, so MP3 or AVI files will not be protected.   Examples of file formats that were specifically designed to support DRM are the AAC music format used by iTunes, and the DVR-MS format used by Microsoft's Media Center to record TV.

DRM is often controversial. People who purchase a song or movie may feel cheated when they realize that the seller has restricted what they can do with it, for instance by not letting them turn the movie into a DVD. Companies that use DRM argue that they continue to own the movie, and they can sell only the rights they choose to sell.

Selecting A Video Format

So which one is best? Any of them. The answer depends on what your requirements are.

Sometimes the choice is easy. If you are making a video file to play on your iPod, you should make an MP4 file. That's what iPods play. If you're making a DVD, you need an MPEG 2 file. That's what the DVD standard calls for.

Sometimes things are less clear. If you're going to let people download a video file from your web page, should it be an AVI file, MPEG 1, Windows Media, QuickTime, or MP4? What are the factors to consider here?

1. Compatibility—that is, will your audience be able to play the video? Using QuickTime can limit this. Not every Windows computer has the QuickTime Player installed. Do you want to ask your viewers to install software? On the other hand, some Mac users will have trouble Playing some AVI and Windows Media files. MPEG 1 may be the most widely-compatible format, but see point 2.

   Another point about compatibility: if you plan to stream your video, you should create it in the format supported by your streaming server. Apple's server streams QuickTime, Microsoft's only Windows Media files, and others will have different limitations.

2. Size—the larger a video file, the more it costs to store on a server (most hosting companies charge by the megabyte) and the more bandwidth it takes to send it. If you are limited to a certain bandwidth, big files mean that fewer people at a time can watch. Even on free services, there's usually a size limit (for instance, 100 megabytes Maximum file size). All other things being equal, you want your files as small as possible. MP4, Windows Media, QuickTime, and FLV formats all use similar compression methods and can produce files of comparable sizes. The older compression systems like MPEG 1 and 2 or Cinpak (or older versions of the new formats) will produce much Larger files.

Let's discuss an example. We'll use the rates for DeskShare's Video Desk service (www.videodesk.net) to show the advantage of a smaller video file.

Disk Space	Data Transfer	Monthly Price
5 MB	100 MB	Free
100 MB	5 GB	$ 19.95
200 MB	10 GB	$ 29.95
1000 MB (1 GB)	Unlimited	$ 179.95

If you have a training video that all 1500 people in your company must watch in November, and it is 100 megabytes in size, you could use the $19.95 account to store it, but the transfer (download volume) involved would be

100 MB x 1500 = 15 GB

This would force you to upgrade to the $179.95 per month account. If you could reduce the video size to 50 MB, the transfer involved would be only

50 MB x 1500 = 7.5 GB

This would let you use the $29.95 per month account, a substantial savings.

3. Quality—the newer compressors, like Windows Media, give higher video quality than the last generation, even at higher compression rates.

4. Special Requirements—sometimes video is created for a specific reason, and has unusual requirements. For instance, MJPEG AVI is an excellent choice for video that you will be editing, because it includes every frame as a full static image.

 When creating video, remember that you control the results. You can trade smaller size for lower quality, for instance, or create somewhat fuzzier video but in a much shorter time. Don't be afraid to experiment, because the proper settings can make a huge difference.

Computer Audio

Thomas Edison used wax cylinders. Then came vinyl discs called "records". Now, most recorded sound spends some time as computer data. This article will give you some idea of how sound is turned into computer files, how those files are compressed to save space, and what audio settings to use when creating your own audio files.

What Is Digital Audio Anyway?

Sound as you hear it is analog. That means that it isn't broken into little specific pieces each of which can be recorded as a number. Instead, sound waves in the air vary smoothly over time, like this:
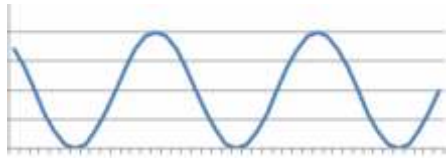
Figure 1 : A simple sine wave

Almost all modern computers are digital. At its most basic level, that means they handle numbers,. A digital computer can't show a smooth curve, like the one in Figure 1. It has to turn it into a series of numbers, like Figure 2. (In electronics this is called an analog to digital conversion.)   The commonest method of doing that is called Pulse Code Modulation (PCM). Basically, the level of the signal is measured every fraction of a second, and those numbers are stored. Figure 2 shows the PCM representation of the signal from Figure 1.
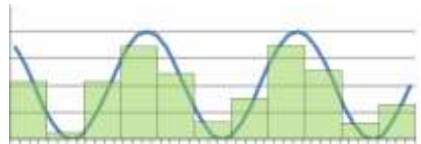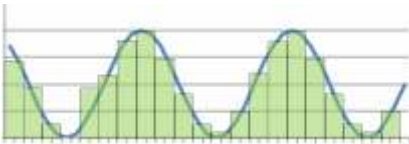
Figure 2: Sample Rates.
note that only the height of the green rectangles is included in the audio file.

You can see that the second graph, sampled 4 times per second, is a lot less like the original curve than the first, sampled 8 times per second. In general, the more times per second you sample, the closer the sound is to the original. The sampling rate is one of the two key numbers in determining fidelity, which is just how close the replayed sound is to the original.

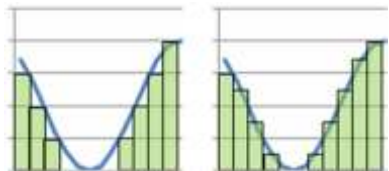The other major factor is bit depth.
Look at Figure 3:

Figure 3: 2-bit and 3-bit audio bit depths

Both graphs have the same sampling rate, but the left graph has a bit depth of 2—each sample can only have one of 4 values (0, 1, 2, 3). The right graph has a bit depth of 3, meaning that each sample can have one of 8 different values (0, 1, 2, 3, 4, 5, 6, 7, or 8). (Like many computer-related things, this uses powers of 2, so 3 bit samples have 23=8 different possible values.) As you can see, the right-hand graph is much closer to the original curve. On the other hand, storing a computer file with double the bit depth takes twice the space. To give you an idea, CD audio is recorded with 16-bit depth, meaning that each sample can have one of 65536 possible values.

Basically, sample rate is how many samples are taken per second, and bit depth is how accurate those samples are. Increasing either gives higher quality sound, and also larger files.

One more thing to consider: modern audio players are almost all stereo, meaning that different sounds come out of different speakers. The most basic version has different sound for the right and left speakers, but higher-end systems can have 6, 7, or even 8 speakers to produce surround sound. In a standard PCM audio recording, each audio channel will be stored separately, so a two-speaker recording would take up twice as much space as a mono recording. However, as we'll see later, there are compression techniques that greatly reduce this requirement for extra space.

This uncompressed PCM sound is used on standard audio CDs. They use a sampling rate of 44.1 kHz (44,100 samples per second) and a bit depth of 16, for historical reasons. (Digital audio was originally stored on videotape, and 44.1 kHz was two stereo channels, with 3 16-bit samples, per horizontal line of the video.) Because CD audio has become the standard, 44.1 kHz and 16-bit depth have become the commonest settings when distributing music files.

Sony's Blu-Ray high-density DVDs also support uncompressed PCM sound, but allow up to 7.1 audio channels (7 speakers plus one subwoofer).

Audio Compression

Sound files can be very large. An uncompressed album (like what's on a CD) can take 600 megabytes or more. If audio is being distributed over the Internet (or another network), huge files translate into lots of bandwidth, which can make podcast downloads or web radio expensive and slow. For anyone who wants to carry music around on a portable player, or send music over a network, it's critical to reduce audio file sizes. The process of making files smaller is called compression.

*Desk**Share**®*
Software to help you excel !

Most people are familiar with general file compression programs like Zip, Rar, and 7-Zip.   For audio files, these systems only produce about a 10% reduction in size, so special methods have been invented specifically for sound.

We'll divide audio compression into two types: lossless and lossy. Lossless audio compression does not change the recorded sound. The original PCM information can be recovered exactly from the compressed file. An example of lossless compression is FLAC (Free Lossless Audio Codec). Lossy compression throws away some information. The various techniques involved are designed to throw away mostly the things that the human ear and brain can't perceive. The commonest form of lossy compression is MPEG 1 Layer 3 (Mp3).

Lossless compression methods are most often used either by people who want the very highest audio quality even at the expense of greater file sizes, or to create archive copies of sound, before creating smaller files like MP3s to play on portable devices. Sound editing is also frequently done using losslessly-compressed files, because the file must be saved many times. Each time a file is saved using lossy compression, some information is lost—which means the sound fidelity gets worse.

Why not always use lossless compression, which would give the best possible sound quality? Because the best lossless compression only reduces file sizes by 30-80%. Acceptable lossy compression can reduce a file by 90% or more. If storage space or bandwidth are major concerns (for instance, on a flash memory-based portable audio player), lossy compression may be the best choice. Even the most finicky of audio enthusiasts will usually be satisfied with some loss of fidelity for certain applications, for instance the podcast of a talk show.

Lossless compression does not mean perfect sound fidelity. Some information is always lost during the original process of recording and digitizing the audio. What lossless compression does guarantee, is that no information, no quality, is lost during the compression process.

Lossy compression uses what's known as the psychoacoustic model. Essentially, the model describes what part of an audio signal the human ear/brain system notices the most. The compressor assumes that it can throw away information that the human listener wouldn't notice anyway. For instance, most adult humans can't hear sounds above a frequency of 22 kHz, so losing that part of the signal will be nearly indetectable. For another example, a very quiet sound may be masked by a louder one. In those cases, deleting the information about the quiet sound can save space without affecting the listening experience.

Lossy compression is extremely complex, and good compressors can produce a much better listening experience without producing larger files, even when using the same compression methods. Audio enthusiasts can hear a very large difference between the same song compressed to MP3 using two different programs, even though both files are the same size and can be played on the same players.

In lossy compression, you always balance file size against quality. By instructing the compressor to use lower bit rates (see the section on bit rates for video compression), you can make the file smaller but the quality worse. Generally, people choose a high bit rate for music, whereas for spoken word recordings (like podcasts or dictation) much lower bit rates are fine.

A constant bit rate file uses the same number of bits to describe every second of the music. For instance, if the rate is 64 kilobits per second (kbps), each second of music uses 64,000 bits. Newer sound encoders can use variable bit rates, in which less space is used when describing simpler segments of the file. For instance, with a constant bit rate, a full second rest (silence) would still take the same 64,000 bits to describe as a second containing four guitar notes, a drum beat, and the singer singing the word "Hello". Using variable bit rates, the silent second could be stored in a small fraction of the space. By using less space for the simpler parts of the audio, the complex parts can be stored with higher bit rates—that is, higher fidelity, resulting in better quality audio.

Compression Methods

There are many techniques for compressing audio. One that is used both in lossy and lossless compression is linear predictive coding (LPC). It relies on the fact that most of the time, a given sound wave is repeated many times with only slight variation:
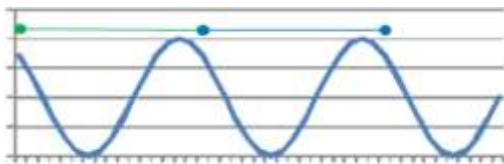

Figure 4: Pure tone, each cycle identical

Rather than store many thousands of sound samples to indicate each point of this curve, linear prediction stores only the differences between the cycles:
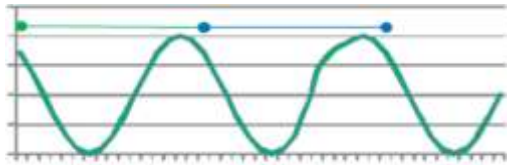

Figure 5: Varying tone. The second cycle rises faster and holds the high point longer
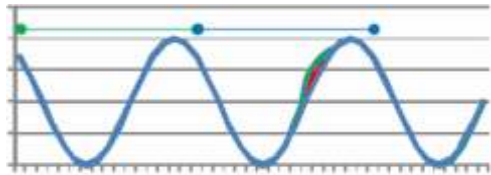

Figure 6: The red area is the difference between the two curves

Figure 4 shows what's known as a pure tone, in which each cycle is identical to every other. These are almost always made electronically, for instance by a synthesizer. Figure 5 shows a varying tone, in which the cycles vary. These are made by living things (like singers) or physical instruments.

Storing the pure tone takes very little space—essentially the compressor would say "do 5,000 of these." Storing each of the 5000 cycles using pulse modulation would take a lot of space. What linear predictive coding does is let a varying tone be stored, by only describing how much it varies from the pure tone.

In this illustration, all that would be stored for the second signal cycle is the representation of the "bulge" shown Figure 6. Actual LPC is very complicated, including methods for compressing information about skipped cycles, slightly shorter or longer cycles, and so forth.

(This is very comparable to interframe compression in video files, except that in sound compression the compressor may have to separate several different signals in the original and compress each separately. For instance, if a pianist plays a chord by striking three keys at once, a compressor would probably use LPC on each string's note individually.)
Linear prediction is used in lossless audio compression, and in some lossy compressors as well.

Stereo Signal Compression

For stereo signals, compression often involves some method of combining the data from the right and left channels and then sending the differences. Again, since the two channels often carry mostly the same audio, sending differences instead of a complete copy of both channels can save space/bandwidth. For example, MP3 files create a mid-channel that combines all sound information, and also a side channel that let the player determine what sound should play from each speaker. (Not all MP3 files use this method.) As a result, when the two channels are playing very similar sound, a file that includes both right and left channels can be only slightly larger than a mono file with a single audio channel.

There are many other compression methods used for audio, including things like Huffman encoding and quantization. This is an extremely technical and highly mathematical subject, and not especially useful to people who just want to make and play audio files.

Common Lossless File Formats

There are many compressed file formats available which lose no data. We'll only mention two here. Since no information is lost making one of these files, converting (transcompressing) from one to another will produce exactly the same sound. Remember when comparing these that by definition all lossless formats are of equal fidelity. One common use of these file types is to back up (non-copy protected) CDs in case of damage, since no quality is lost.

Free Lossless Audio Codec (FLAC)

FLAC files are popular among music traders. There are many compressors and many players available, and even some portable music players can play FLAC files directly (without converting into MP3 or AAC first).

Monkey's Audio (.APE)

APE files tend to be slightly smaller than FLAC, but encoding and decoding both take longer.

Common Lossy File Formats

MPEG-1 Layer 2 (Mp2)

MP2 is a compression system defined by the same standard as MP3 (which is MPEG-1 Layer 3). There's also a layer 1 (which is almost never seen). Why would one standards document have three different ways to compress audio? They are used for different purposes. MP2 requires much less computation for the system that creates them. That is, they can be created faster. When one is (for instance) recording audio, a system might not have the horsepower to record audio in MP3 format at real time, but be well able to handle the easier job of recording MP2 files.

Although not used much for consumer music recordings because it creates larger files than MP3, MP2 compression is still used in digital television, and on some DVDs. You will occasionally encounter MP2 files, especially when looking at older archives. Almost any player (hardware or software) that can play MP3 files can also play MP2s.

Mp3

MP3 files were the first widely-used means of music file sharing, and the first file type used on digital portable music players. MP3 is probably the most compatible audio file format—almost any audio-related computer program or digital music player can work with MP3s. Newer systems like AAC can often produce smaller files with equivalent or better sound quality.

Many audio enthusiasts say that MP3 compression dramatically harms sound quality. However, in listening tests, some songs are actually preferred in MP3 format over the original CD. The compression process does change the sound, but apparently it adds a "roundness" or "smoothness" that the listeners like.

Advanced Audio Codec (AAC)

Apple's iTunes music store, their iPod and iPhone devices, and other popular devices like the Sony PlayStation 3 all use audio compressed with the AAC. Interestingly, AAC is another MPEG standard. MP3 was defined by the MPEG-1 document. AAC is part of both the MPEG-2 and MPEG-4 documents. It is the standard audio compression system for MPEG-4 video.   Aside from generally offering better sound quality than its predecessor, AAC has built-in error correction features (in case a file is corrupted or a disc is damaged).

Although it offers technical advantages over MP3, AAC remains less popular. Partly, this is explained because MP3 is "good enough", so people have less motivation to change over. Newer MP3 encoders (like Digital Media Converter) produce much better audio quality than the ones used in the 1990s, diminishing the AAC advantage. Also, the single biggest source of AAC files, the Apple Music Store, until very recently only sold DRM-protected files, and consumers preferred music that they could use as they chose.

Ogg Vorbis (.ogg)

Ogg Vorbis is a sound file format specifically created to replace MP3. Aside from better sound quality, the motivation was avoiding the patent protection on all the MPEG standards. Some developers wanted to create a sound compression system that they could freely develop and use without paying patent fees.

Ogg files are rated by listeners as very high quality, but the format is supported by few portable digital players. However, Ogg files have been very popular with video game makers. The audio in games like Unreal Tournament and Grand Theft Auto: San Andreas is stored in Vorbis format. It is also used by some radio stations.

RealAudio (.ra)

RealNetworks' RealAudio format can produce acceptable sound quality at very low bit rates. At higher bit rates, it appears to offer little advantage over other compression methods. This format was originally designed for Internet streaming applications and is still widely used for that purpose. It isn't supported on many portable players. Note that RealAudio actually includes at least 10 different compression methods, including Real's version of AAC.

Windows Media Audio (WMA)

Microsoft has created its own audio compression system as part of its Windows Media family of products. The actual compression method seems to be similar to AAC. Windows Media is the second-most-supported audio file format. WMA files can be played on many portable players, DVD and CD players, etc. Microsoft has now released their Windows Media Professional format, which gives higher fidelity, and supports 5.1 and 7.1 channel audio. They've also released a lossless version of WMA. Neither is widely supported at this writing, but this can be expected to change in the future.

Special: Musical Instrument Digital Interface (MIDI, .midi or .mid)

MIDI files are not the standard type of audio file, created by recording actual sounds and converting them into a series of numbers. MIDI files are instructions for playing musical instruments. Instead of recording pitch or tone directly, MIDI files contain instructions like "Play the piano middle C key at this specific velocity, wait .34 second, then release the key." MIDI definitions exist for every real instrument and some that only exist as computer models.
MIDI files will often be smaller than a compressed audio file of the same music, because instructions by their nature take up less space than sound recordings. Interestingly, the same MIDI file can be used to play music on a computer, or to control instruments to physically play the same sounds, by moving keys, etc. MIDI files can only be created by specialized MIDI editors. Many computer audio players (but very few portable digital players) support MIDI.

Audio In Video Files

Audio in video files is generally compressed. Uncompressed audio can take up more space than the video, which can double or more the bandwidth required to stream a movie. Both the MP3 and AAC compressors are used in video files. Windows Media Video files, as you might expect, use Windows Media Audio compression. RealVideo uses RealAudio codecs.

Making Audio Choices

What's the right audio file format, or compression system to use? The answer depends on your needs.

In general, lossless compression is used for archiving only. Since all lossless formats are exactly equal in quality, the choice here probably depends on what software you like to use. FLAC is very widely supported in software. Also, FLAC is actually supported by some digital players, if you want to carry the very highest-quality recordings around with you.

Lossy compression is more complicated, if only because there are actual differences in fidelity to consider. In listening tests, there is only a very slight difference among the various modern lossy compressors (MP3, AAC, Ogg Vorbis, WMA, etc.) Generally the newer standards do better at low bandwidths (64 kpbs).

If you want your files to play on anything, MP3 is a good choice. If you are very concerned about space or bandwidth requirements, for instance if your portable player has only 64 megabytes of flash memory and it plays one of the newer formats, go with either AAC, Ogg Vorbis, or WMA, which lets you get the best quality from highly-compressed files.

If you're compressing the sound track of a video file, be sure to use an appropriate technique. For instance, technically it's possible to make a Windows Media file using MP2 audio compression, but there's no guarantee that this will play back—players expect WMV files to use WMA compression. (Digital Media Converter will not let you choose inappropriate compressors for WMV files.)

Similarly, MP4 files should use one of the MPEG-4 audio compressors, like AAC.

There are other settings you can select, when creating audio or video files. (Each audio type has different settings available—for WMA and WMV and MPEG-2 files, the audio settings are built into each profile.) Here's some advice:

$$\frac{\text{Bit rate in bits per second x length of the file in seconds}}{8}$$

(We divide by 8 because file sizes are normally expressed in bytes, which are 8 bits long.)
For example, if you create an MP3 file with a constant bit rate of 128 kpbs (128,000 bits per second), which is 60 seconds long, the size will be:

$$\frac{128000 \times 60}{8} = 937.5K$$

Actual files will be slightly larger because they contain some non-audio information, including the metadata.

The bit rate is also the amount of bandwidth that the sound will use if the file is streamed. Bit rates are set for WMA/WMV, MPEG 1 and 2, and MP3 files. Generally, the higher the bandwidth, the better the quality and the larger the file. You'll have to judge how important sound quality is for you, but many people follow a general rule that 64 kbps is enough for conversation, and 128 kbps for music.
Keep in mind that the quality of the actual listening experience also depends on the playback equipment. Even if you store the audio at 320 kbps, if it's played back on a standard TV set, it will probably be impossible to tell the difference from a 128 kbps file.

Sample Rate or Frequency

As we described above, the sample rate controls how many times per second the sound levels can be changed. For music, always pick at least 44.1 kHz (the CD standard). Lower sample rates probably don't hurt recordings of people talking. Generally, a sample rate setting will exist for all the formats that don't have a bit rate setting.

Sample Size or Bit Depth

This is the accuracy of each sample. For music, the 16-bit standard used on CDs is probably a good minimum. For spoken words, you can experiment with smaller samples.

Please remember that these are only suggestions. The sure way to get the sound quality you want is to experiment. Be sure to play back the files on the same equipment your audience will be using.

What We Did Not Cover:

Audio is a huge subject, and this is only an introduction. In this article, we haven't discussed any of the various specialized audio systems for speech recording and audio conferencing, like DCT, GSM, and Speex. We also haven't dealt with file formats that contain playlists, like .asx and .ram. And there are many, many audio codecs we haven't talked about, from A-law to WavPack.

Wrapping It Up

If you're interested in learning more, Hydrogen Audio (hydrogenaudio.org) is an excellent resource.

We hope that this guide to audio and video files has been helpful. If you have any questions, corrections, or suggestions for improving it, please send e-mail to: writer@deskshare.com

Thank you for reading.

DeskShare publishes a complete line of video editors and converters. You can find more information about them at our web site, www.deskshare.com.

DeskShare®
Software to help you excel !