A Brief Guide To Digital Audio Formats
By Laurence Fenn

Audio recording has come a long way from the wax cylinder, vinyl and cassette tape. With PCs and portable music players, digital recordings are very common. The format that they use can vary, and this paper will attempt to explain the differences.

Audio formats can be split into two categories, the type that record/playback audio, and the type that playback audio. The latter covers file types like midi, and mod files whilst the former covers wav, mp3 and others.

WAV

WAV (or WAVE) is short for Waveform audio format and is a Microsoft and IBM file format standard for storing audio. It is the default format for digital audio on Windows PCs. Though a WAV file can hold audio compressed with any codec, by far the most common format is PCM audio data.

Since PCM uses an uncompressed, lossless storage method which keeps all the samples of an audio track, it is the standard that offers the same quality as data on audio CDs and therefore offer the best quality. PCM is a straight representation of the binary digits (1s and 0s) of sample values. The Windows start up sound and system sounds are all wav files, which is fine for a short duration. As wav files can take 10Mb per minute, depending on whether they are recorded in stereo and what sampling rate they use, transferring files in this format is not suitable.

When you convert analogue sound to digital by recording music with your computer (or rather sampling the audio to be more accurate), the resulting file will be in wave (.wav) format. The best quality will be a 16 bit, 44KHz sampling rate (which is the same as CD audio, although you can go to 48KHz). The variety in number of bits (usually 8 or 16), number of channels (two for stereo, one for mono), sampling rate (44KHz for high quality music, 22KHz or lower for low quality or voice only) and codec used means that you can save the same recording in lots of different files sizes/quality. They may not sound the same, but it depends on what you want to do with the recording.

There are a few other codecs other than PCM which are commonly used with WAV files:

DPCM (Differential Pulse Code Modulation) is a simple form of lossy compression that stores only the difference between consecutive samples. DCPM uses 4 bits to store the difference, regardless of the resolution of the original file. With DCPM, an 8-bit file would be compressed 2=1, and a 16-bit file would be compressed 4=1.

ADPCM (Adaptive Differential Pulse Code Modulation) is similar to DCPM except that the number of bits used to store the difference between samples is varied depending on the complexity of the signal. ADPCM works by analyzing a succession of samples and predicting the value of the next sample. It then stores the difference between the calculated value and the actual value.

u-law Compression (pronounced “mew-law”) is a common lossy compression scheme, similar to ADPCM, which can be used on AU, AIFF and WAV files.

MP3 is part of the official MPEG-1 standard (see the MPEG standard in video compression) and can used as a codec, as well as a file format in it’s own right. As with the MPEG video, high compression rates can be achieved with little apparent lose of quality (see the description of the MP3 format for more information). An example of this is in the Excel sound quizzes that are available on my web site at http://www.lfenn.co.uk. If I had saved the music extracts in PCM format, the sound quality would be as good as CD, but the quizzes would be ten times as big.

AIFF and AU

AIFF is the default audio format for the Macintosh, and AU is the default format for SUN systems. Both of these formats are supported on most other platforms and by most audio applications. Each of these formats can be compressed, but compression sometimes creates compatibility problems with other platforms.

MOV (Apple QuickTime)

QuickTime is a widely used multimedia format from Apple Computer that supports both streaming audio and streaming video. Much of the MPEG-4 standard is based on QuickTime, and it is widely used for streaming video on the Web. An MOV file doesn’t need to contain video, and QuickTime also comes with it’s own software synthesizer, which you can use to convert midi files to mov files.

WMA (Windows Media Audio)

Microsoft's Windows Media Audio (WMA) format is Microsoft's respond to MP3. WMA performs very good at lower bit-rates and is reported to produce quality indistinguishable from the original CD at 128 kbps. WMA is supported by most full-featured player programs and by many portable players. WMA is royalty-free when incorporated into software that runs on the Windows platform. Above all WMA offers the advantage that copyright-protected songs cannot be published any further (Digital Rights Management).

RA (RealAudio)

RealAudio was the first widely used system for streaming audio and video over the Internet. It is a proprietary format, but it is used by many online music stores for sample clips of songs. The RealPlayer also provides support for MP3. RealAudio (or ra) files are streamed using a ram file, which tells the browser to launch the Real Player and start playing the file when it has enough in it's buffer. This is same way that RealVideo files are streamed from web sites. Without this method, the entire ra file would be downloaded before playback began.

MP3 (MPEG Audio Layer III)

MP3 was introduced as a part of the official MPEG-1 standard in 1992 and until today it is the most successful audio-standard since WAV. The German Fraunhofer Gesellschaft (FhG), which has developed this audio-compression still holds the key patents the MP3-techology inherits. Using MP3-compression PC users were able to compress an ordinary music CD to one tenth of it's original size - thus 12 hours of music could be stored on a recordable CD that on the other hand could be played by an MP3 CD player or an ordinary PC. MP3s are created by taking wave audio data and processing it with a special algorithm. This algorithm removes parts of the audio that theoretically cannot be detected with the human ear; in actuality, there will be some degradation of quality, but this depends on the quality (bit rate) with which you choose to encode the file.

The net result is an MP3 file which is vastly smaller than the original wave file, but sounds very nearly as good. As an example of the huge size different between a wave file and an MP3, a three minute song will take up 30Mb as a wave file, but only between 2 and 7Mb as an MP3 (depending on the bit rate you choose). This explains why MP3 files are so popular for trading music on the internet.

The bit rates, i.e. number of binary digits streamed per second, is variable for MP3 files. The general rule is that the higher the bit rate, the more information is included from the original sound file, and thus the higher is the quality of played back audio. In the early days of MP3 encoding, a fixed bit rate was used for the entire file.

Bit rates available in MPEG-1 layer 3 are 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 and 320 kbit/s (bits per second), and the available sample frequencies are 32, 44.1 and 48 kHz. 44.1 kHz is almost always used as this is the audio CD frequency, and 128 Kbit is some sort of de facto "good enough" standard. MPEG-2 and (non-official) MPEG-2.5 adds more bit rates: 8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160 kbit/s.

However, audio in MP3 files are divided into chunks called frames, which all have a bit rate marker, so it is possible to change the bit rate dynamically as the file is played. This technique makes it possible to use more bits for parts of the sound with high dynamics (much "sound movement") and less bits for parts with low dynamics. Some encoders utilize this possibility to greater or lesser extent.

OGG

The development of the OGG standard began in 1993, then known as “Squish”. OGG was right from the start an open source project and hence is free of any patents. It was designed as a substitute for MP3 and WMA and by now it is almost as popular and well known as MP3. Above all, the algorithm is still being developed what is mainly due to its flexibility. Although the sound-quality gets better with every further development the files are backwards compatible and can be played with older players as well. Like MP3 OGG offers encoding at variable bit rates. Using this compression parts of the song are encoded with a higher compression than others what depends on the source. Most times, this compression goes along with squishy noises or even small interruptions. OGG is also one of the very few formats that support multi-channel compression. Surround-files could theoretically be compressed with more than two channels. OGG is, like it's predecessors, streamable and although the used player has to support this feature, it's one of many good reasons for OGG. Many games use ogg files for their background music, such as Unreal.

mp3PRO

mp3PRO is the next generation of MP3. A division of the Fraunhofer Institute is working on this together with Thomson multimedia. mp3PRO is said to offer the same quality of MP3 at half the file size. This is achieved by a further compression of a tone's high frequencies. This SBR (Spectral Band Replication) is believed to be almost loss-less and represents the PRO in the name. Like WMA mp3PRO is backwards compatible, that means mp3PRO-files can be played with common MP3-players. These files however sound very dull and rustled.

MID (Musical Instrument Device Interface)

Midi is an entirely different sort of file. Unlike the other formats, it is not compressed audio. MIDI is a kind of 'language' that allows computers and certain musical instruments to communicate. This language consists of instructions telling the instrument (or the MIDI synthesizer in your sound card) which notes to play, with what instrument, and when. MIDI can be used entirely within a computer, with no external instruments, although you can control keyboards and drum machines with the appropriate interface. MIDI files have no sound of their own (that depends on the device you play the file back on) and are quite small.

RMID, RMI - is a standard RIFF file format. The method of saving data in chunks is the basis for Interchange File Format. A MIDI File format is a "broken" IFF. It lacks a file header at the start of the file. In order to fix the MIDI File format so that it strictly adheres to IFF, Microsoft simply made up a 12-byte header that is attached to the beginning of the MIDI file, and thereby came up with the RMID format. If you look in the media directory of your Windows PC you will see an example midi file called canyon.mid. But users of later versions of Windows will find some RMI files like Dance of the Sugar Plum Fairy and Beethoven's Fur Elise, which are much better.

MOD

Modules are digital music files, made up of a set of samples (the instruments) and sequencing information, telling a mod player when to play which sample on which track at what pitch, optionally performing an effect like vibrato, for example. Thus mods are different from pure sample files such as WAV or AU, which contain no sequencing information, and MIDI files, which do not include any custom samples/instruments. Mods are extremely popular in the demo world and offer a way of making music of an acceptable level of quality rather cheaply.

Mods' sequencing information is based on patterns and tracks. A pattern is a group of tracks with a certain length, usually 64 rows. The tracks are independent of each other, meaning that a four track mod can play four voices or notes simultaneously. The patterns can be sequenced in a play list, so that repeating the same sequence of patterns doesn't require rewriting of them.

There are a variety of MOD file types, from MOD (4 tracks, 31 samples), to 669 (8 tracks, 64 samples), S3M, XM and IT. With so many formats it can be confusing, especially finding software to play the files (let alone create them). Fortunately Winamp has built in support for many of the formats.

SBK, SF2 (Soundbanks and Soundfonts)

Creative Labs wanted a new sound format to help with the launch of its then new AWE soundcard range. Soundbanks (SBKs) help make midi files sound realistic by storing instruments using samples in a format that midi files could access. The SBK format became the improved SF2 or soundfont, and by using multiple samples and looping, users had the chance to effectively have their own polyphonic sampler on their PC. A midi file playing a piano piece can use a soundfont with samples from a Steinway piano, instead of using a synthesized piano sound. Vocals and effects can be added, and many examples can be found on my web site at http://www.lfenn.co.uk. Other sound cards are now compatible with this format.

How do I play/create these file formats?

Windows Media Player can handle WAV, RMI, MID, MP3 and WMA files of course. MOV files and RA files are played with QuickTime Player and Real Player respectively. The other files can usually be played with WinAmp, which can also handle WAV, MID and MP3 files. This player can use plug-ins to handle other types of files, and can play most of the MOD file types as well. You'll need a Creative Labs soundcard or a compatible type to load SBK or SF2 files into memory (the AWE32/64 cards had to use memory on the card itself, whereas the newer cards use your system memory).

If you want to record audio, then you can use the Sound Recorder included in Windows, but this has a limit of 60 seconds for files. There are ways to get around this, but you would be better off using a dedicated audio program like CoolEdit or SoundForge.