FILE SIZES, DATA RATES AND ALL THAT

We live in a digital era. Audio is now becoming digital data handled by computer processing, with analogue conversion occurring as the interface to this world.

All digital audio has to be sampled. The most common technique for this uses analogue to digital converters with linear PCM encoding where the audio is sampled then a digital word is assigned representing the value at the instant of sampling. Examples include PCM encoding and .WAV files.

The reverse or decoding procedure uses digital to analogue converters generating instantaneous voltages corresponding to the digital value.

One problem with the decoding step is that the output contains undesired switching transients imposed by the steps in the voltage occurring between decoded words. This shows up as harmonics at multiples of the sampling frequency and needs to be removed from the audio by filtering. The introduction of these harmonics is called digital aliasing and the filters required for removal are thus anti-aliasing filters.

The sample rate dictates the ultimate bandwidth for the resultant audio. To achieve a bandwidth of 20 kHz, a sample rate of at least 40ks/sec is required. We say "at least" because any effect the anti-aliasing filter tail has on the highest audio frequencies will also need to be allowed for[1].

The bit budget

As with all media there are constraints. In audio it is the total data volume - the bit budget.

There are three component parts to the budget – channel count, sample rate and word size. Decisions made with these three parameters dictate the amount of data created, shipped, stored and ultimately decoded for re-play.

Basic channel count scales directly. Double the channels requires double the data and double the data rate. Ten channels need ten times the volume and rate.

One way to reduce this is to derive virtual channels. Virtual channels require either analogue or digital signal processing capability but do not add to channel count bandwidth directly. Strategies include transmitting one channel and the difference signal for one or more other channels. This saves on data when the channels share significant amounts of common content.

Do we really need 24 bit, 96 ks/sec?

Word size was a critical factor when computers were limited to 8 bit words. 16 bit audio required double the processing steps. In addition to being faster, most modern computers use 32 bit operation and so this is not as much of an issue.

Word size dictates the ultimate resolution of the sampled audio. Smaller words mean less resolution and so smaller dynamic range or higher errors (distortion).

Present day music CD's use 16 bit words and so can have a dynamic range of 96 dB (alternately can resolve to .0015%). This is probably acceptable.

Professional grade equipment using true 24 bit sampling has a dynamic range of 144 dB (resolve s to 0.00000596%). This could be considered excessive for audio applications as

---

[1] The Nyquist sampling rate is double the highest desired frequency.

humans can only cope with acoustic dynamic ranges of around 120 dB without risk of damage.

Sample rates of 96ks/sec or more are easily achieved with modern systems. These systems support frequencies up to a maximum bandwidth of 48 kHz. This would be great for recording bats and ultrasonic alarms, but do humans need this?

Our experience has been that young adult humans can hear sounds above 20 kHz - but not much above. A response to 22 kHz could be justified. This implies a sample rate of around 48 ks/sec. Whilst it is true that higher sample rates could be required for extensive digital processing and re-sampling associated with professional audio production, a factor of 256 is possibly excessive and anything higher most certainly is.

The net result of increased word size and higher sample rates is an increase in overall data volumes and necessary transfer rates. As an example, a 24 bit, 96ks/sec 5 channel 1 hour audio DVD (no pictures) would require 1.2 GB of data on a system capable of handling 24 bit words as single bytes. If the specification were reduced to 48 ks/sec the data requirement would halve.

Digital file compression

It is possible to reduce the amount of storage space and transmission bandwidth by removing redundant data during encoding. This process is called compression. There are two methods of doing this - lossless compression and lossy compression.

Lossless compression encoders just consider the audio as data words and use data redundancy techniques to reduce the file size. With lossless compression, the original data is fully recoverable after encoding/decoding. Examples include MLP (Meridian Lossless Packing), ALE/ALAC (Apple lossless encoder) LPAC, TTA, WAVPack, Dolby TruHD and WMA 9 series. Compression ratios of 60% are typically achieved with lossless compression systems.

Lossy compression irretrievably removes data according to some strategy. Key amongst these strategies are the perceptual coders that remove data according to whether it relates to audio that can be perceived (heard). This includes noise floor assumptions and the relative loudness of the sound in frequency bands and hence the masking of these sounds by the louder bands. Because it is assumed to be masked, it can be removed by the encoder. Examples include Dolby Digital, Dolby AC3, MPEG1 layer 3 (MP3), AAC (Advanced Audio Coding). Compression ratios of between 80 and 95% are achieved by lossy perceptual encoders.

To add to the format complexity, the more recent formats including MPEG also support data file encapsulation and delivery. This means it is possible, for example, to use MPEG data streams for delivery of lossless compressed files or even basic linear PCM and still fully recover the data with the relevant decoder at the other end of the process. All this is very confusing for the end customer.

Some lossy encoders allow specified data rates. For example it is possible to encode MP3 at 192 kB. Use of over-sampled lossy encoders is of dubious value and can actually require more data space than the original or lossless compressed file.

There is continued high interest in lossy audio compression software development and proprietary solutions abound, but it is preferable to use original source or lossless compressed files where circumstances permit. And the promoters keep promising cheap bandwidth "by the mile" and mass storage "by the ton" so why bother compressing anyway? And if you do not have the necessary decoder format your favourite digital recording will be useless.

Graeme Huon HuonLabs 2008