Compression

Compression refers to reducing the quantity of data used to represent a digitised object.

There are two distinct reasons to compress data:

- The uncompressed version is **too big** (whatever “too big” means).
- The encoding is not optimal.

The second reason is not serious enough to compress in practical applications (unless coupled with the first), so, in the end, people compress data to reduce the quantity.

Depending on the impact of compression on the quality of the resulting encoding, we distinguish between **lossless** and lossy compression.

Audio and video

It is absurd to treat audio and video applications as inherently requiring more compression than other applications. It is the user's demand (backed with dollars) to receive audio and video in real–time that makes compression of audiovisual signals important, even when it comes at the cost of inferior quality.

Consider a police record of a person: it can easily be compressed by more than 50% without any special techniques (SIN: 30 bits instead of 72, DOB: 16 bits instead of 64, a 10–letter name: 48 bits instead of 80, etc.). In practical application the savings (an access time reduction from, say, 4ms to 2ms) does not warrant any effort.

Hence, data compression is particularly relevant in audio and video and it is because of the real–time delivery required.

Quality

In the absence of any better definition the following is used for audio signals:

CD quality: 2 audio channels, each carrying 44,100 16–bit samples per second = 1.411 Mb/s.

Cinema quality audio: 6 audio channels, each carrying 48,000 16–bit samples per second = 4.6 Mb/s in 5.1 Channel Surround.

Cinema quality audio: 8 audio channels, each carrying 48,000 16–bit samples per second = 6.14 Mb/s in 7.1 Channel Surround.

Note that the industry is moving to 96kHz sampling (instead of 48kHz) and the number of channels will increase to 16. Sadly, even 22.2 (24–channel) sound is around the corner.

Compression is not necessary: a standard CD may contain 60 minutes of high–quality orchestral music, even though 2 hours of lossless compressed Hi–Fi music or 7 hours of MP3–quality music can be stored on it.

Lossless compression is used heavily but lossy compression is required for good cost–effective results.

Lossless compression

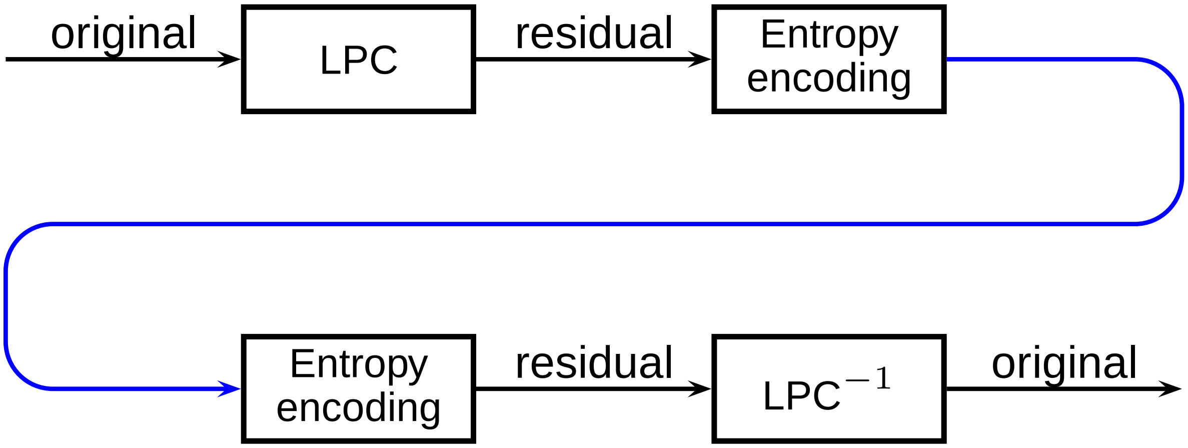
reduces the size of audio files by 40–50% without any loss of contents.

Standard techniques used for general data compression (**zip** type), such as Huffman codes, **Lempel–Ziv–Welch** are usually labelled together as **entropy encoding**.

They do not perform very well when applied to audio (or video) signals. The methods used in practice combine entropy encoding with an additional preprocessing called **predictive coding** (in its simplest form, Linear Predictive Coding–LPC).

The role of PC is to remove from the input its correlation with the previous sample (this correlation is very high in audio–visual signals). What is left by the preprocessor is an uncorrelated residual which is a good feed for entropy encoding.

At the receiver a similar postprocessor adds the correlation back.


original
LPC
residual
Entropy encoding
Entropy encoding
residual
LPC^{-1}
original

MPEG–4 ALS

The **MPEG**4 ALS standard is an example of lossless compression (March 2005, then March 2009).

MPEG–4 Audio is not restricted to audio data; it can be used for any kind of digital data that has strong temporal correlation (the standard names medical and seismic data, which is a bit dubious, given the low accuracy of the original measurements).

Lossy compression

MPEG-1 Audio was created as part of an effort to standardise lossy compression synchronising audio and video.

Three modes were proposed:

Layer 1 is computationally cheapest or poor quality and is not used much.

Layer 2 was used in Video CDs (**VCD**) at 128kb/s.

Layer 3 intended for streaming video uses various rates from 64 kb/s to 320 kb/s.

Sampling rates vary with 3 rates common: 32kHz, 44.1kHz and 48kHz (44.1kHz is king here). Sampling always uses PCM sampling

MP3

The **universal scapegoat** of lossy compression techniques is the **MPEG**–1 Layer 3 standard.

Known as MP3 it is so widely used that it became almost synonymous with lossy compression.

MP3 is standard only to a certain extent: the decoding part is quite precise but the encoding and the tricks used in compressing while retaining the essence of the original sound are left to the manufacturers.

MP3 essentials

While many tricks are used, the basic ones are:

Psychoacoustic model: removing the tones that are not going to be heard by a human anyway.

Polyphase filterbank combined with MDCT (similar to FFT) divides the signal into 32 sub–bands.

See: polyphase filterbank in action.

The PA model indicates how many bits to allocate to each sub–band created by the PFB.

PFB

The filterbank is a parallel device (512 parallel filters in MP3) that divides the audio signal into 32 equal–width bands and performs a Fourier Transform on signal in each band.

Once transformed, the original signal cannot be recovered without distortion, so PFB is lossy. A lossless cosine transform is used to compensate for some distortion introduced by PFB.

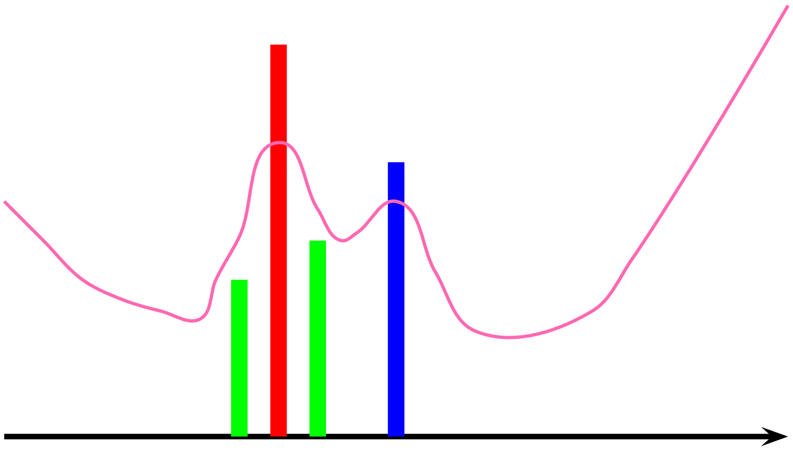
Psychoacoustic model

Masking

MP3 uses auditory masking a phenomenon well-known (and undisputed): the human brain ignores a tone when it coincides with another tone, of similar frequency and higher amplitude. Masking is cumulative: two high–volume signals raise the audibility threshold more than any single one of them.

Wikipedia's version

My own version:



Masking thresholds

The steps taken by MP3 are:

1. The 32 sub–bands are formed.
2. Masking is used: the masking threshold is computed for each sub–band (or portions of) based on the signal in adjacent bands.
3. The signal that is below the threshold is not encoded.
4. The signal that is encoded uses a lossy encoding with at most as many bits lost as the threshold indicates.

Example copied from Dave Marshall

See Dave Marshall's tutorial.

Frequency division yields the following distribution of the signal (only the sub–bands 4–10 are shown):

Band	4	5	6	7	8	9	10
Level (dB)	10	6	2	10	60	35	20

The 60dB level in band 8 results in a masking threshold of 12dB in band 7 and 15dB in band 9. Band 8 is below the threshold, so it is not encoded; band 9 is above, so it is encoded with fewer than 3 bits of quantisation error (2 bits $\approx$ 12dB, 3 bits $\approx$ 18dB).

In the case of most of popular music the reduction in bandwidth is significant without much loss of quality.

MP3

MP3 is not exactly a standard–it represents a collection of similar products based on the MPEG–1 Layer 3 standard which is very vague when it comes to encoding (it specifies decoding quite precisely).

In particular the bit rates used are left to the manufacturers to choose from a very wide range: from 32kb/s to 320 kb/s. The most common rates are 128, 192, and 224 kb/s. An MP3 frame consists of a 32–bit header followed by 256–bit side information and 1152 samples. The data part ("samples") is compressed using a Huffman code.

Side information: information needed to decode the main data Huffman table, parameters, etc.

The samples represent: 32 sub–bands of 3 consecutive groups of 12 consecutive samples in each group. Each sample is up to 15 bits long; the result is, or at least can be, a variable bit rate (**VBR**) stream.

Video compression

Video differs from audio mainly in one aspect: far more data. Uncompressed video can exceed 1 Gb/s (1080p **HDTV**) which is prohibitive for storage and digital transmission.

Hence compression is a necessity. Compression comes in two basic forms:

Spatial: in intraframe compression (**JPEG**).

Spatial and Temporal: in interframe compression (**MPEG** and H.261).

All the methods are **lossy**.

Frame prediction

The stream of frames is divided into groups, each group consisting of:

1. One **I–frame** which is a complete compressed frame independent of other frames (as in JPEG).
2. A sequence of **P–frame**s which are differential frames, showing the difference between what this frame should be and what the previous frame should have been.

There also are B–frames which are more compressed than P–frames.

Intraframe

A fairly complex treatment of a single colour frame. Main aspects:

- The image is divided into macroblocks of 8×8 (or 16×16 in MPEG–2) pixels.
- Each macroblock is subjected to a DCT and compressed using discrete quantisation levels.
- Each macroblock is represented using the YCbCr encoding (Y=luminance, then B–Y difference and R–Y difference) which is also called YUV. (See Barns.)
- The most common encoding has 4 Y–blocks (no loss), and single Cb and Cr blocks (75% loss) although provisions exist for lower loss levels.
- The final stage of compression is the zig–zag scanning and entropy encoding using a Huffman code.

JPEG claims a 27:1 compression ratio.

Video standards from MPEG

MPEG–1: audio and VCD video (up to 1.5 Mb/s).

MPEG–2: DVD compression, interlaced and progressive scan, multiple channels audio (up to 5.1).

MPEG–4: geared towards multimedia. MPEG–4 is made of many standards, some of them not fully developed.

MPEG–7: describes how information about the content should be stored (not a compression standard).

MPEG–21: wishful–thinking standard.