

MP3 TECHNIQUES SIMPLIFIED

Lt Col VD Singh

It sounds amazing as to how a wave file can be reduced in size as an mp3 file and yet we hear the music without being able to find the difference in aural perception. While the process is mathematically complex, the MP3 format uses a few techniques and tricks which are explained in simple language below:

- Minimal audition threshold
- Masking effect
- Bytes reservoir
- Joint Stereo coding
- Huffman coding

Minimal audition threshold

The minimal audition threshold of the ear is nonlinear. It is represented, according to the law of Fletcher and Munson, by a curve dug between 2Khz and 5Khz. It is not therefore necessary to code sounds situated under this threshold, because they will not be perceived.

Masking effect

This system is based on masking properties of the human ear: When you look at the sun and if an aircraft passes ahead, you do not see it because of the too predominant light of the sun. In audio, it is similar. During strong sounds, you do not hear the weakest sounds. Take as an example a piece of pipe organ. When the organist does not play, you hear the breath in the piping, and when he plays, you no longer hear it because it is masked.

It is therefore not necessary to code all the sounds. This is the first property used by the MP3 format to earn some space. For this the MP3 encoder uses a psychoacoustic model modeling the behavior of the human ear.

Bytes reservoir

Often, some passages of a musical piece can not be coded to a given rate without altering the musical quality. The MP3 then uses a short reservoir of bytes that acts as a buffer by using capacity from passages that can be coded to an inferior rate in the given flow.

Joint Stereo (JS) coding

In case of a stereophonic signal, the MP3 format can then use a few more tools, referred to as Joint Stereo (JS) coding, to further shrink the compressed file size.

In many mid-range Hi-fi sets , there is a unique subwoofer. However you usually do not have the feeling that the sound comes from this boomer, but rather from

satellite speakers. Indeed for very low and very high frequencies, the human ear is no longer able to locate the spacial origin of sounds with full accuracy. The mp3 format can therefore (optionally) revert to such a trick by using what is called Intensity Stereo (IS). Some frequencies are then recorded as a monophonic signal followed by a few additional information in order to restore a minimum of spatialisatation.

The second joint stereo tool is called Mid/Side (M/S) stereo. When the left and the right channels are quite similar, then a middle (L+R) and a side (L-R) channels are encoded instead of left and right. This allows to reduce the final file size by using less bits for the side channel. During playback, the MP3 decoder will reconstruct the left and right channels.

Huffman coding

The MP3 also uses the classic technique of the Huffman algorithm. It acts at the end of the compression to code information, and this is not therefore itself a compression algorithm but rather a coding method.

This coding creates variable length codes on a whole number of bits. Higher probability symbols have shorter codes. Huffman codes have the property to have a unique prefix, they can therefore be decoded correctly in spite of their variable length. The decoding step is very fast (via a correspondence table). This kind of coding allows to save on the average a bit less than 20% of space.

It is an ideal complement of the perceptual coding: During big polyphonies, the perceptual coding is very efficient because many sounds are masked or lessened, but little information is identical, so the Huffmann algorithm is very seldom efficient. During "pure" sounds there are few masking effects, but Huffman is very efficient because digitalized sound contains many repetitive bytes, that will then be replaced by shorter codes.