High Quality Audio Coding: MPEG-1/2 layer-3 (MP3)

MPEG is the acronym for **m**oving **p**icture **e**xperts **g**roup. The group was established in 1988 to develop generic open standards for development of coders for moving pictures and audio. Open standards are specifications that are available to developers interested in implementing the standard. Usually an implementation example is provided to avoid misinterpretation of the text of the standard.

The audio coding standard developed by the MPEG group is used in many applications including digital audio broadcasting, internet audio, portable audio, DVD and audio storage. A popular version of MPEG is MP3 developed at the Fraunhofer Institute, a German audio research laboratory. MP3 (short for MPEG-1/2, audio layer) is a subset of MPEG compression that can take a 20-megabyte, two-minute music track on a CD, and reduce it to about 1.4 megabytes, the size of a floppy disk.

There are a few versions of MPEG standard each version includes a higher level of quality, flexibility and application.

MPEG-1 audio consists of three operating modes, called layers, with increasing complexity and performance from layer-1 to layer-3. MPEG-1 defines audio compression at sampling rates of 32 kHz, 44.1 kHz and 48 kHz. It works with both mono and stereo signals, and a technique called joint stereo coding can be used for efficient coding of the left and right channels. MPEG-1 Layer-3 provides high quality audio at about 128 kbps for stereo signals.

MPEG-2 introduced new concepts for video coding and digital TV. It also extends MPEG-1 audio sampling rates to half rates to include 16 kHz, 22.05 kHz, and 24 kHz.

MPEG-3 was to define video coding for high definition television (HDTV) applications. However, as MPEG-2 contains all that is needed for HDTV, MPEG-3 was rolled into MPEG-2.

MPEG-4 is more concerned with new functionalities than better compression efficiency. The major applications of MPEG-4 are mobile and fixed terminals, database access, communications and interactive services. MPEG-4 audio consists of audio coders spanning the range from 2 kbps low bit rate speech, up to 64 kbps/channel high quality audio.

MPEG-7 is a content representation standard for multimedia information search engines, filtering, management and processing of data.

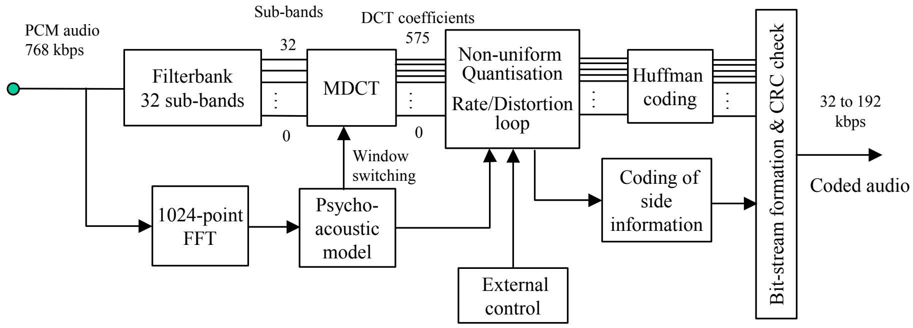

Figure 12.30 Illustration of MPEG1 layer-3 system.

MPEG Bit Rates

MPEG audio does not work just at a fixed compression ratio. Within the prescribed limits of the open standard, the selection of the bit rate is left to the implementer and/or the operator of the audio coder. The standard bit rate is defined in the range from 8 kbps to 320 kbps. The open standard also enables the use of variable bit-rate coding and fixed bit-rate coding within the prescribed limits.

MPEG Structure

Figure 12.30 illustrates a block diagram structure of an MP3 coder. It consists of the following signal processing subsystems.

The Filter bank consists of 32 poly-phase filters each followed by a modified discrete cosine transform (MDCT). The signal in each band can be encoded differently depending on its energy and contribution to the perceptual quality of the output audio.

The Perceptual model calculates the noise masking thresholds; that is the amount of noise in each frequency band that would be masked and made inaudible by the signal energy at and around that band. The frequency bands for calculation of the masking thresholds are based on the so-called critical bands of hearing. The auditory critical bandwidth at any frequency f are the band of frequencies Δf that affect the threshold of hearing of a tone at f. If the quantisation noise energy can be kept below the masking threshold then the compressed signal would have the same perceptual audio quality as the original signal.

Quantisation and coding is achieved through an iterative two-stage optimisation loop. A power law quantiser is used so that large values are coded with less accuracy, as a higher signal energy would mask more noise. The quantised values are then coded by Huffman coding. To adapt the coder to the local statistics of the input audio signal the best Huffman coding table is selected from a number of choices.

The Huffman coder is a probabilistic coding method that achieves coding efficiency through assigning shorter length codewords to more probable (i.e. more frequent) signal values and longer length codewords to less frequent values. Consequently, for audio signals smaller quantised values, which are more frequent, are assigned shorter length codewords and larger values, which are less frequent, are assigned longer length codewords.

The relative proportion of smaller and larger sample values are controlled using a global gain and set of subband scale-factors. If the number of available bits is not enough to encode a block of data then the global gain can be adjusted to result in a larger quantisation step size and smaller quantised values. This is repeated with different values of quantisation step size until bit requirements is no more than the available bit resources.

To shape the quantisation noise so that it is below the masking threshold scale-factors are applied to each band. If the quantisation noised in a band exceeds the masking threshold then the scale factor for that band is adjusted to reduce the noise below the masking threshold. However this adjustment and control of distortion noise can result in a higher bit rate than allowed. So the rate adjustment loop has to be repeated each time scale-factors are adjusted. Therefore the rate loop is nested in the distortion loop. The rate and distortion loops can fail to converge and may go on forever, however in a good design several conditions can be checked to stop the iterations in an early stage.