Chapter 14
MPEG Audio Compression

14.1 Psychoacoustics

- The range of human hearing is about 20 Hz to about 20 kHz

- The frequency range of the voice is typically only from about 500 Hz to 4 kHz

- The dynamic range, the ratio of the maximum sound amplitude to the quietest sound that humans can hear, is on the order of about 120 dB

Frequency Masking

- Lossy audio data compression methods, such as MPEG/Audio encoding, remove some sounds which are masked anyway
- The general situation in regard to masking is as follows:
  1. A lower tone can effectively mask (make us unable to hear) a higher tone
  2. The reverse is not true – a higher tone does not mask a lower tone well
  3. The greater the power in the masking tone, the wider is its influence – the broader the range of frequencies it can mask.
  4. As a consequence, if two tones are widely separated in frequency then little masking occurs

Threshold of Hearing

- A plot of the threshold of human hearing for a pure tone

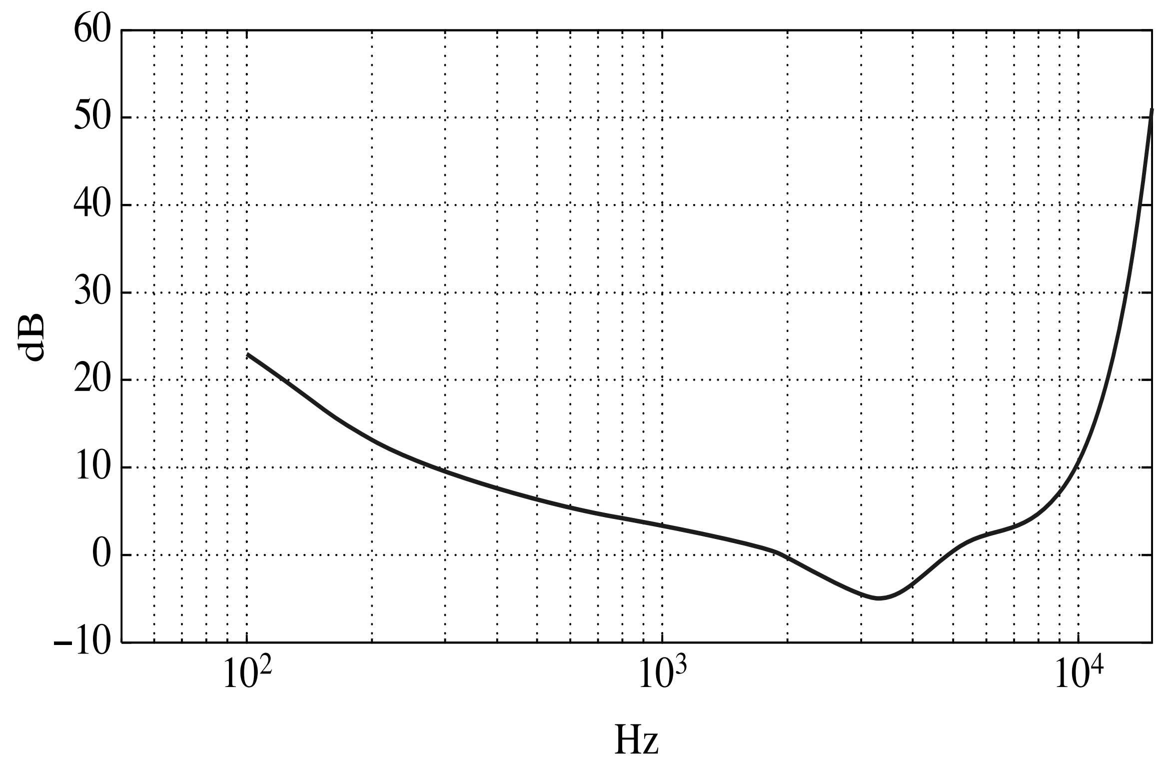

Fig. 14.2: Threshold of human hearing, for pure tones

Threshold of Hearing (cont'd)

- The threshold of hearing curve: if a sound is above the dB level shown then the sound is audible
- Turning up a tone so that it equals or surpasses the curve means that we can then distinguish the sound
- An approximate formula exists for this curve:

$$\text{Threshold}(f) = 3.64(f/1000)^{-0.8} - 6.5\,e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \tag{14.1}$$

  - The threshold units are dB; the frequency for the origin (0,0) in formula (14.1) is 2,000 Hz: $\text{Threshold}(f) = 0$ at $f = 2$ kHz

Frequency Masking Curves

- Frequency masking is studied by playing a particular pure tone, say 1 kHz again, at a loud volume, and determining how this tone affects our ability to hear tones nearby in frequency
  - one would generate a 1 kHz *masking tone*, at a fixed sound level of 60 dB, and then raise the level of a nearby tone, e.g., 1.1 kHz, until it is just audible

- The threshold in Fig. 14.3 plots the audible level for a single masking tone (1 kHz)

- Fig. 14.4 shows how the plot changes if other masking tones are used

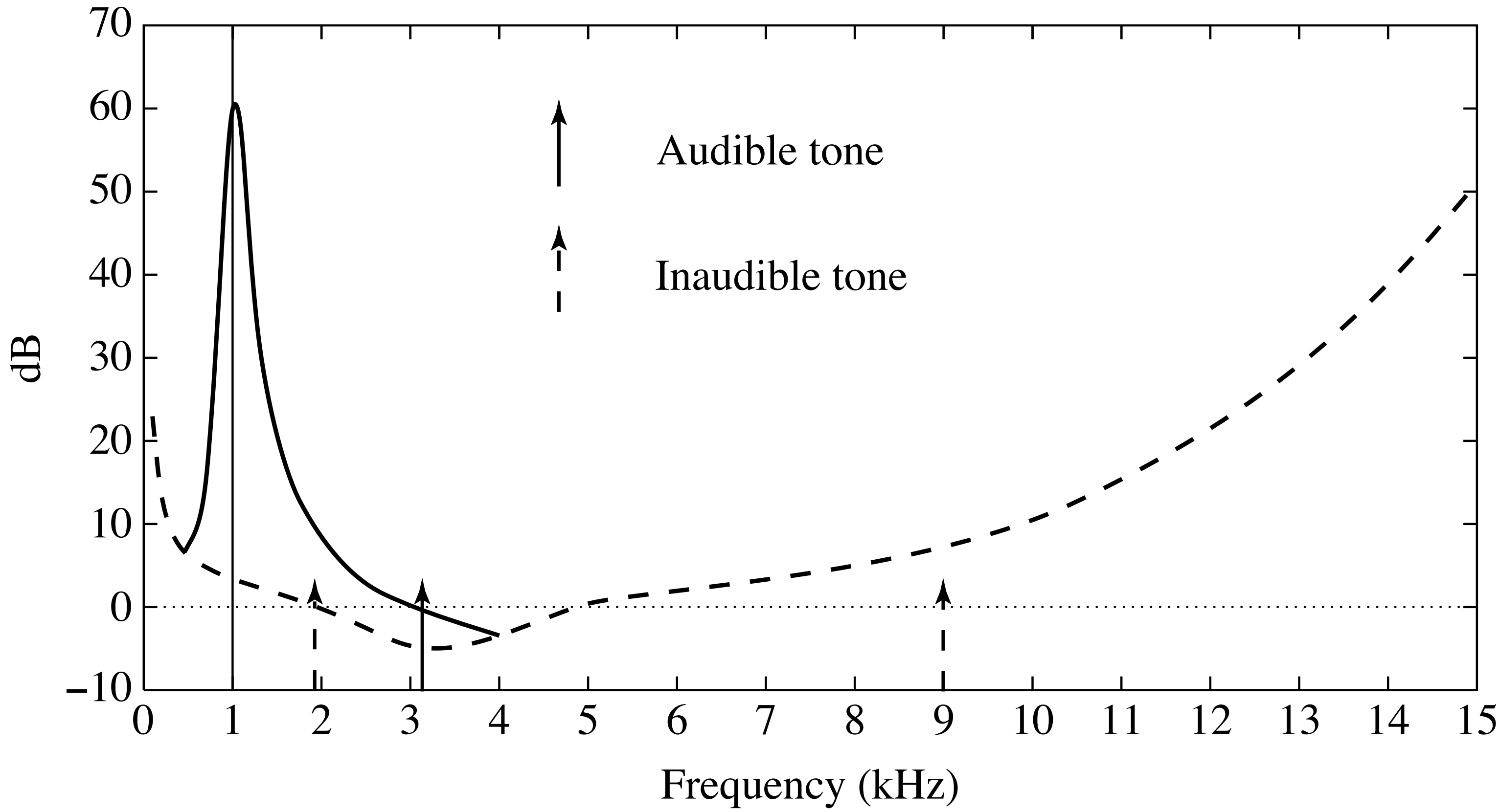

Fig. 14.3: Effect on threshold for 1 kHz masking tone

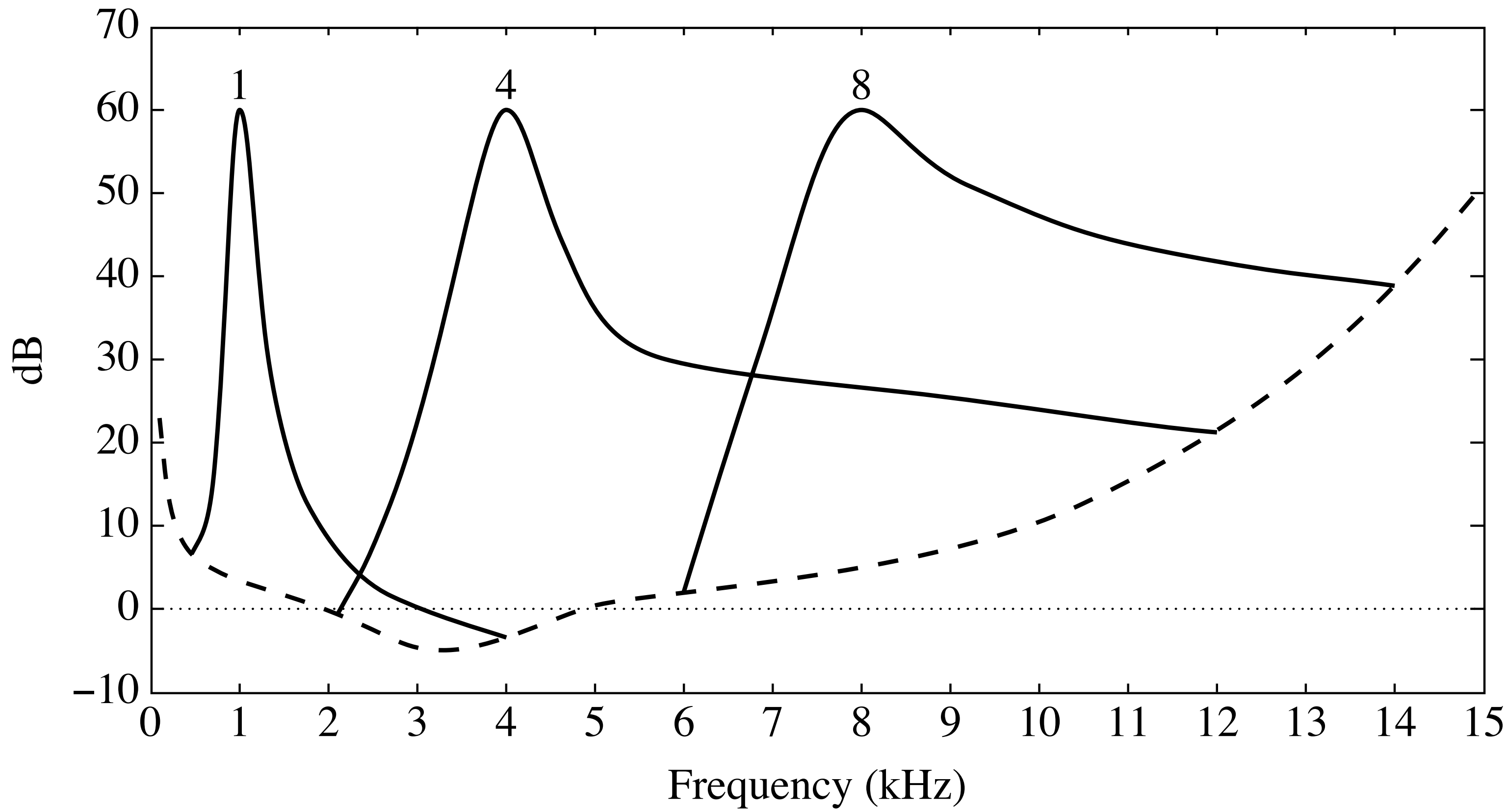

Fig. 14.4: Effect of masking tone at three different frequencies

Temporal Masking

- **Phenomenon**: any loud tone will cause the hearing receptors in the inner ear to become *saturated* and require time to recover

- The following figures show the results of Masking experiments:

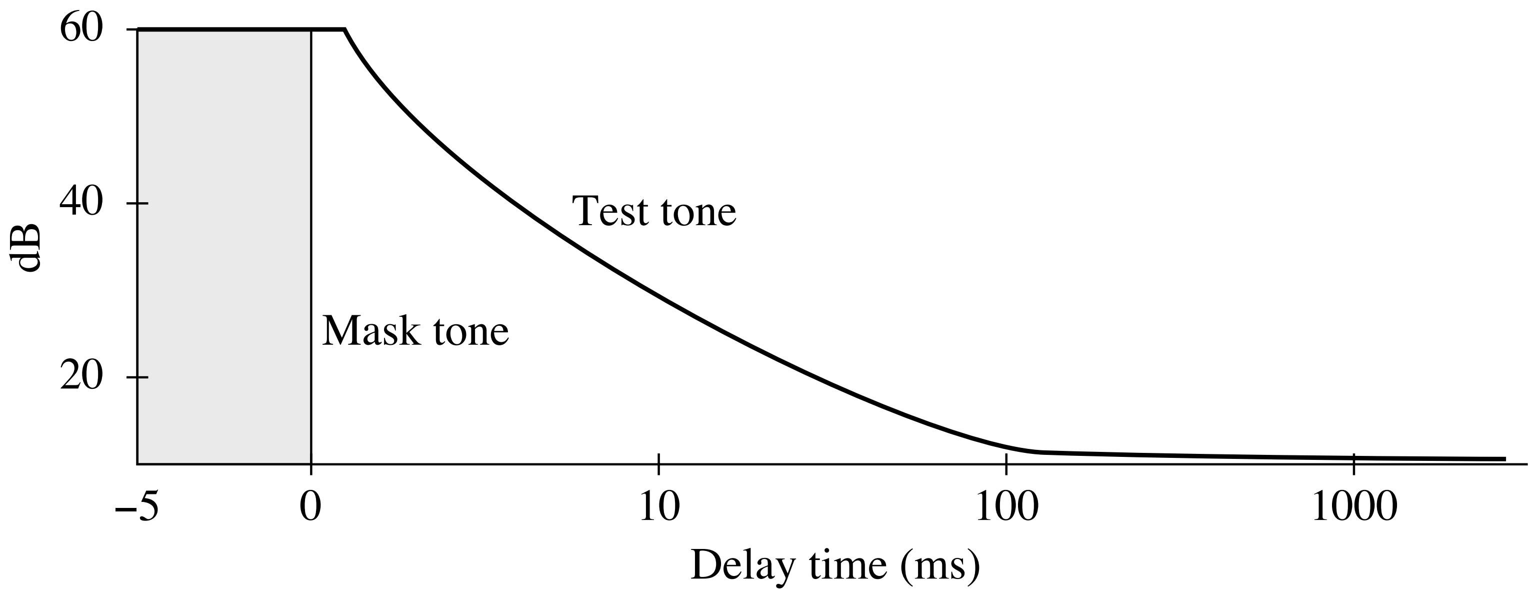

Fig. 14.6: The louder is the test tone, the shorter it takes for our hearing to get over hearing the masking.

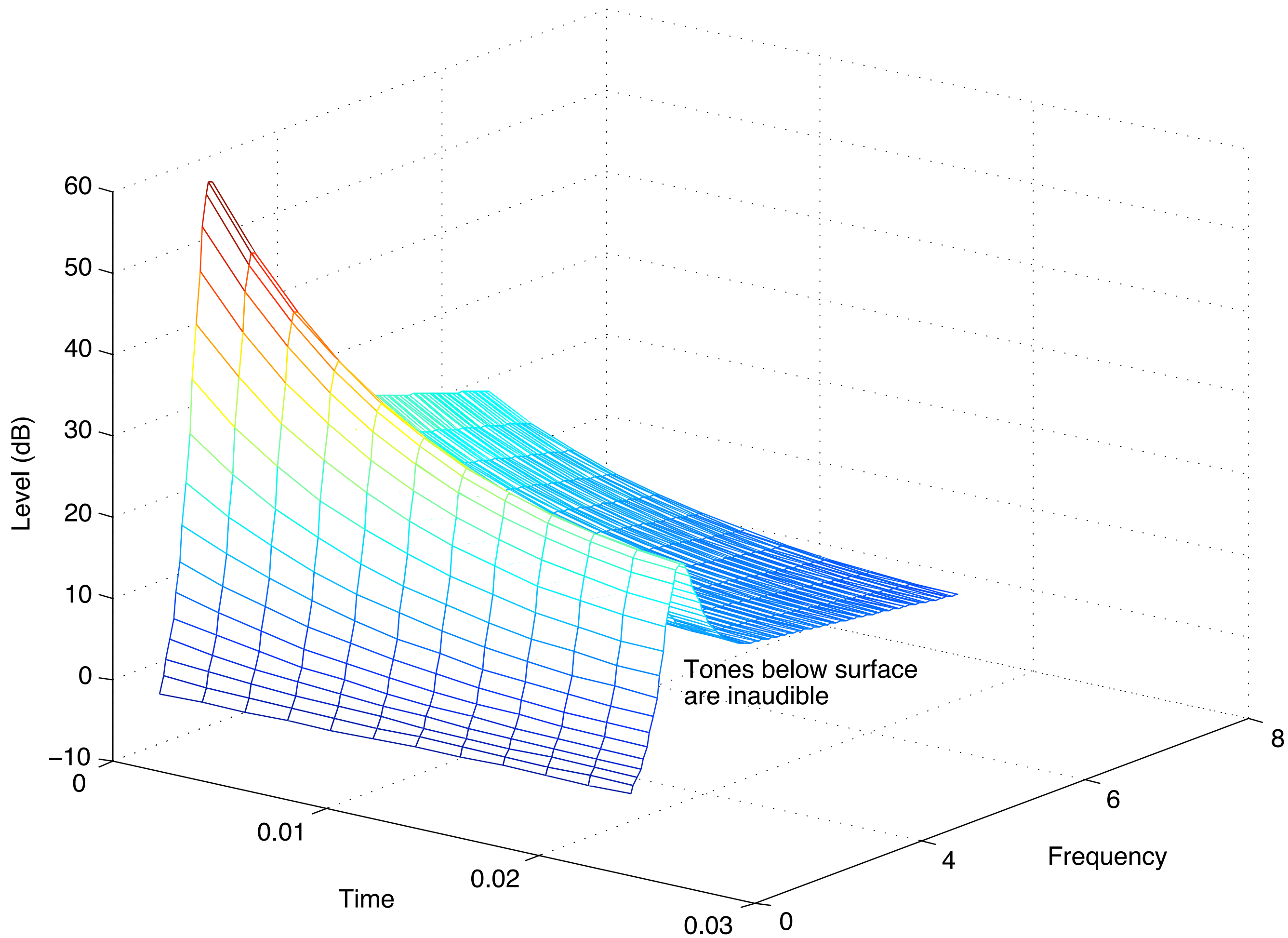

Fig. 14.7: Effect of temporal and frequency maskings depending on both time and closeness in frequency.

14.2 MPEG Audio

- **MPEG audio compression** takes advantage of psychoacoustic models, constructing a large multi-dimensional lookup table to transmit masked frequency components using fewer bits

- **MPEG Audio Overview**

  1. Applies a filter bank to the input to break it into its frequency components

  2. In parallel, a psychoacoustic model is applied to the data for bit allocation block

  3. The number of bits allocated are used to quantize the info from the filter bank – providing the compression

MPEG Layers

- MPEG audio offers three compatible *layers* :
  - Each succeeding layer able to understand the lower layers
  - Each succeeding layer offering more complexity in the psychoacoustic model and better compression for a given level of audio quality
  - each succeeding layer, with increased compression effectiveness, accompanied by extra delay
- The objective of MPEG layers: a good tradeoff between quality and bit-rate

MPEG Layers (cont'd)

- Layer 1 quality can be quite good provided a comparatively high bit-rate is available
  - Digital Audio Tape typically uses Layer 1 at around 192 kbps
- Layer 2 has more complexity; was proposed for use in Digital Audio Broadcasting
- Layer 3 (MP3) is most complex, and was originally aimed at audio transmission over ISDN lines
- Most of the complexity increase is at the encoder, not the decoder – accounting for the popularity of MP3 players

MPEG Audio Strategy

- **MPEG approach to compression** relies on:
  - Quantization
  - Human auditory system is not accurate within the width of a critical band (perceived loudness and audibility of a frequency)

- **MPEG encoder** employs a bank of filters to:
  - Analyze the frequency ("spectral") components of the audio signal by calculating a frequency transform of a window of signal values
  - Decompose the signal into subbands by using a bank of filters (Layer 1 & 2: "quadrature-mirror"; Layer 3: adds a DCT; psychoacoustic model: Fourier transform)

MPEG Audio Strategy (cont'd)

- **Frequency masking**: by using a psychoacoustic model to estimate the just noticeable noise level:
  - Encoder balances the masking behavior and the available number of bits by discarding inaudible frequencies
  - Scaling quantization according to the sound level that is left over, above masking levels

- May take into account the actual width of the critical bands:
  - For practical purposes, audible frequencies are divided into 25 main critical bands (Table 14.1)
  - To keep simplicity, adopts a *uniform width* for all frequency analysis filters, using 32 overlapping subbands

MPEG Audio Compression Algorithm

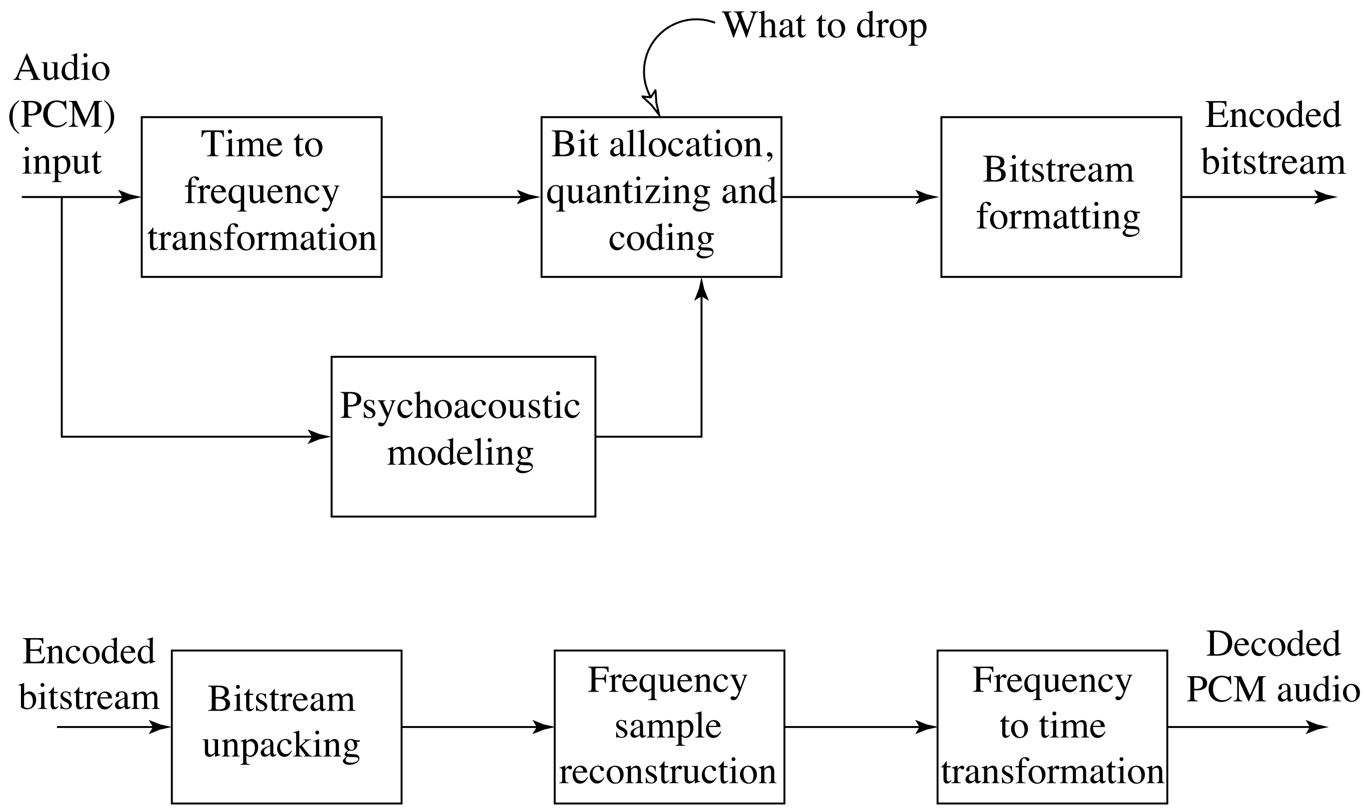

Fig. 14.9: Basic MPEG Audio encoder and decoder.

Basic Algorithm (cont'd)

- The algorithm proceeds by dividing the input into 32 frequency subbands, via a filter bank
  - A linear operation taking 32 PCM samples, sampled in time; output is 32 frequency coefficients

- In the Layer 1 encoder, the sets of 32 PCM values are first assembled into a set of 12 groups of 32s
  - an inherent time lag in the coder, equal to the time to accumulate 384 (i.e., 12×32) samples

- Fig.14.11 shows how samples are organized
  - A Layer 2 or Layer 3, frame actually accumulates more than 12 samples for each subband: a frame includes 1,152 samples

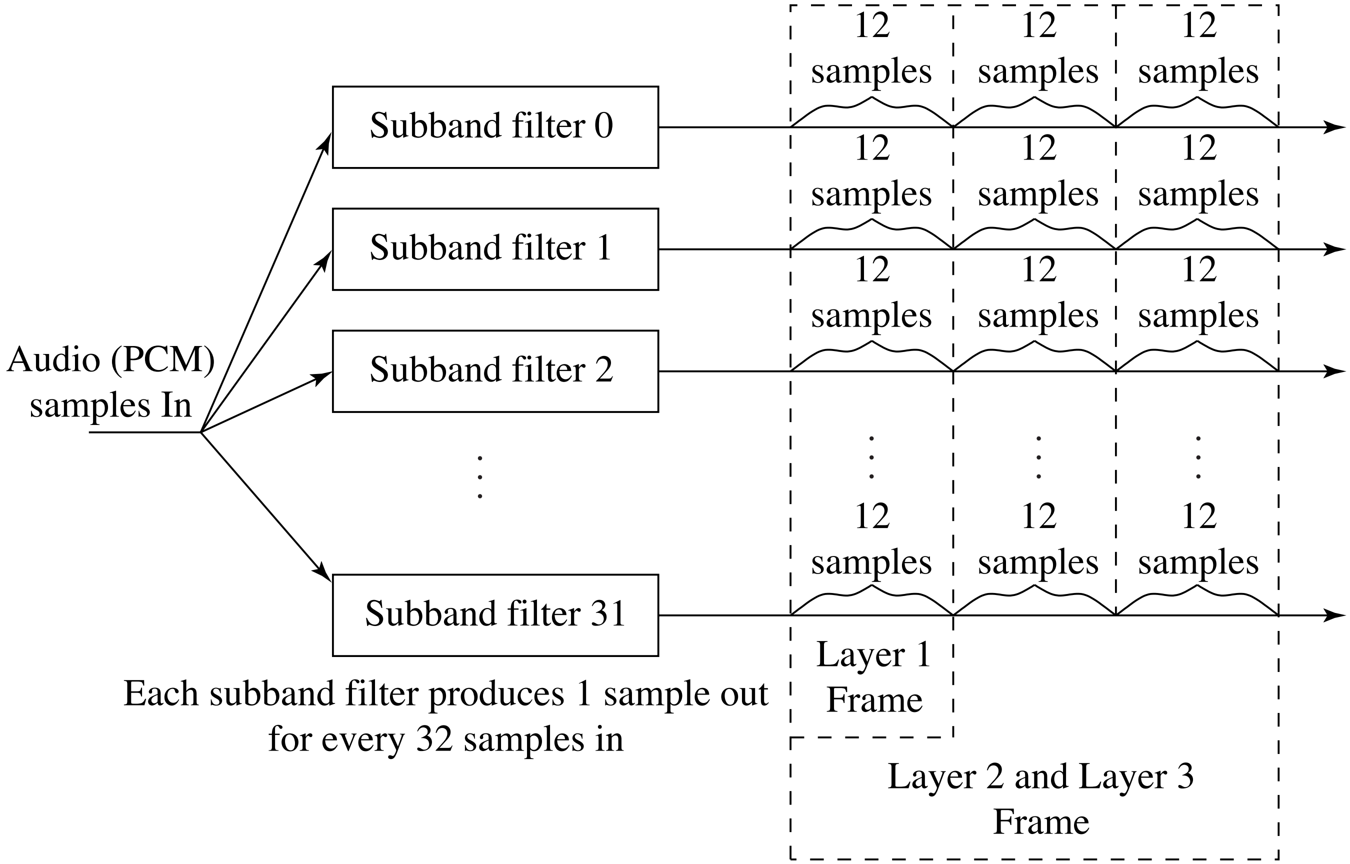

Fig. 14.11: MPEG Audio Frame Sizes

- Mask calculations are performed in parallel with subband filtering, as in Fig. 4.13:

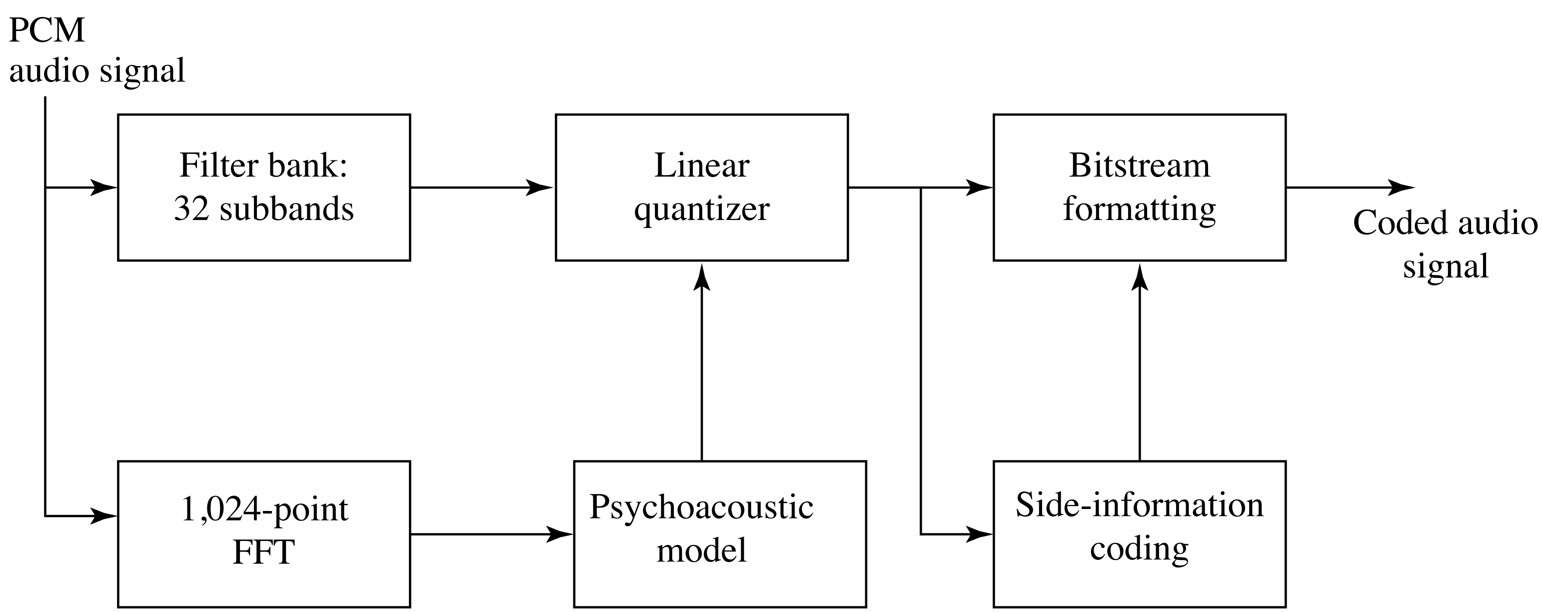

Fig. 14.13: MPEG-1 Audio Layers 1 and 2.

Layer 2 of MPEG-1 Audio

- **Main difference:**
  - Three groups of 12 samples are encoded in each frame and temporal masking is brought into play, as well as frequency masking
  - Bit allocation is applied to window lengths of 36 samples instead of 12
  - The resolution of the quantizers is increased from 15 bits to 16

- **Advantage:**
  - a single scaling factor can be used for all three groups

Layer 3 of MPEG-1 Audio

- **Main difference:**
  - Employs a similar filter bank to that used in Layer 2, except using a set of filters with non-equal frequencies
  - Takes into account stereo redundancy
  - Uses Modified Discrete Cosine Transform (MDCT) — addresses problems that the DCT has at boundaries of the window used by overlapping frames by 50%:

$$F(u) = 2\sum_{i=0}^{N-1} f(i)\cos\left[\frac{2\pi}{N}\left(i+\frac{N/2+1}{2}\right)(u+1/2)\right],\ u=0,..,N/2-1 \tag{14.7}$$

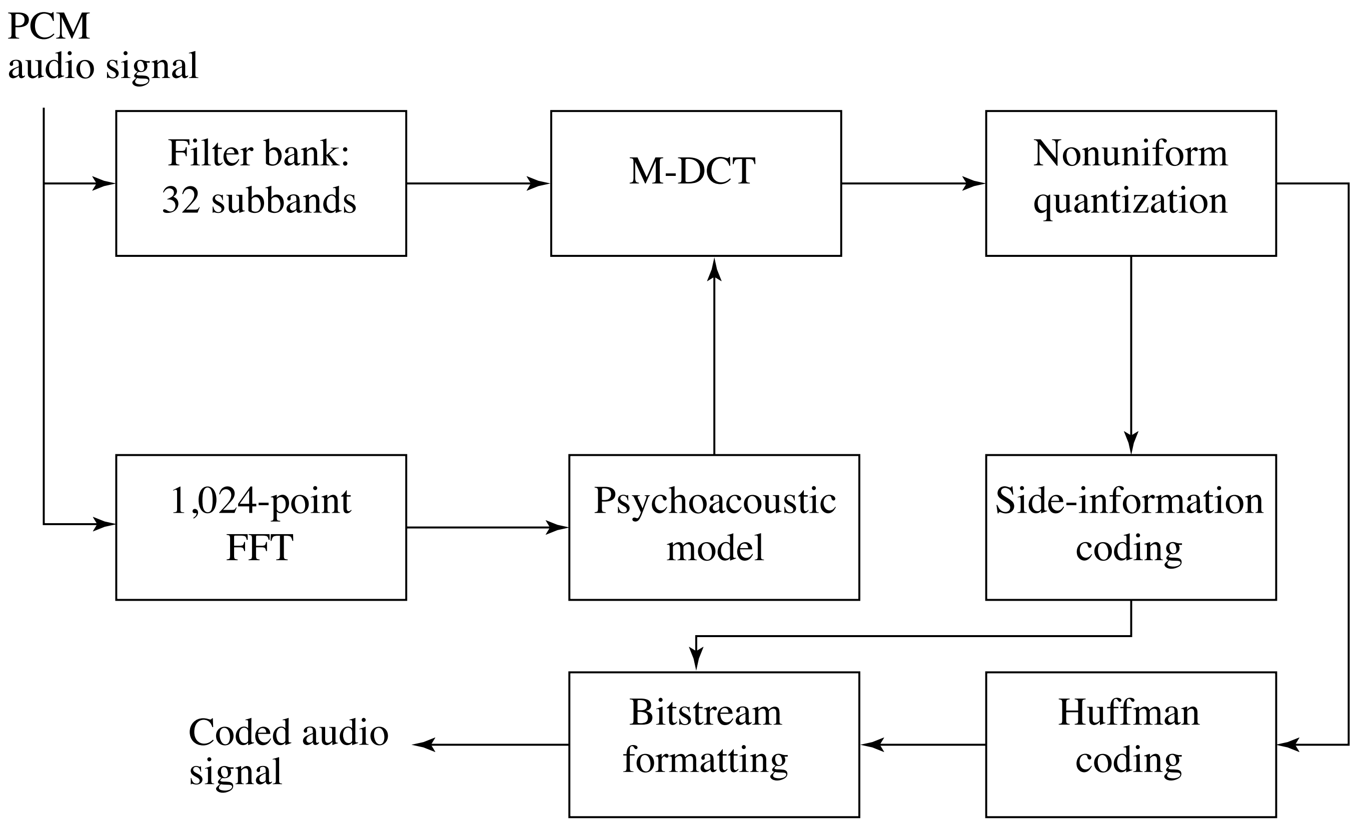

Fig 14.14: MPEG-Audio Layer 3 Coding.

- Table 14.2 shows various achievable MP3 compression ratios:

Table 14.2: MP3 compression performance

Sound Quality	Bandwidth	Mode	Compression Ratio
Telephony	3.0 kHz	Mono	96:1
Better than Short-wave	4.5 kHz	Mono	48:1
Better than AM radio	7.5 kHz	Mono	24:1
Similar to FM radio	11 kHz	Stereo	26 - 24:1
Near-CD	15 kHz	Stereo	16:1
CD	$>$ 15 kHz	Stereo	14 - 12:1

14.5 Further Exploration

⟶ Link to Further Exploration for Chapter 14.

In Chapter 14 the "Further Exploration" section of the text website, a number of useful links are given:

- Excellent collections of MPEG Audio and MP3 links.
- The "official" MPEG Audio FAQ
- MPEG-4 Audio implements "Tools for Large Step Scalability", An excellent reference is given by the Fraunhofer-Gesellschaft research institute, "MPEG 4 Audio Scalable Profile".