
2

MPEG/Audio Layer III

This chapter describes the algorithms in the MPEG/Audio Layer III compression standard. The chapter is an initial survey used as a foundation for a later definition of which algorithms within Layer III belong to the bitstream handling and which algorithms belong to the DSP part.

Although, only parts of the decoder will be implemented in this project, the following description also covers the audio encoding part in order to form a general view of the encoding/decoding process. The encoder description will be held at a level where only the functionality will be considered, whereas the description of the decoder will contain enough information to perform a classification of the algorithms involved.

2.1 MPEG/Audio Layer III Encoder Overview

In this section the MPEG/Audio Layer III encoder will briefly be described with emphasis put on the functionality. The description of the encoding process is based on the block diagram in Figure 2.1. Here the audio signal is assumed to be a single channel Pulse Code Modulated (PCM) signal sampled with a rate of 48 kHz and quantized to 16 bit. The MPEG/Audio encoder compresses the input signal to a coded bitstream and thereby reduce the bit rate from 768 kbit/s to some arbitrary bitrate kbit/s.

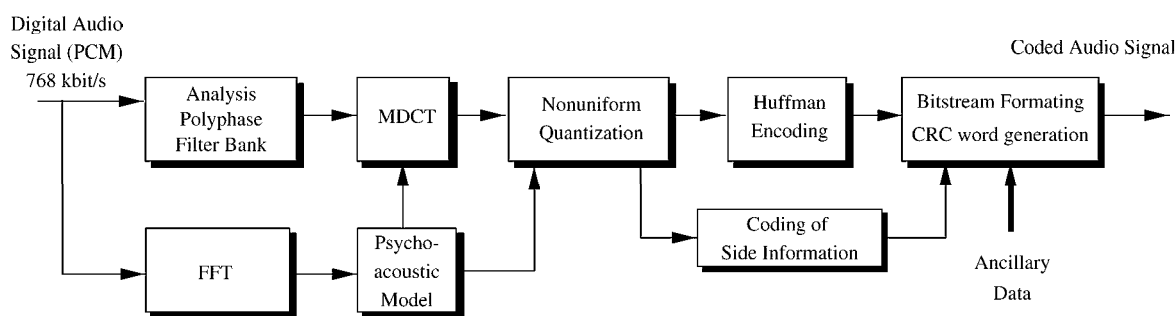


Figure 2.1: Block diagram of MPEG/Audio encoder.

Analysis Polyphase Filter Bank: The first step in the encoding process is the filtering of the audio signal through a filter bank. In this process a sequence of 1152 PCM audio samples are filtered by a parallel structure of bandpass filters into 32 equally spaced subbands and decimated by a factor 32. Each subband will thereby contain 36 subband samples. Assuming perfectly square bandpass filters, the Nyquist theorem guarantees that the original 1152 PCM samples can be reconstructed, by interpolating the subbands to their original sampling frequency followed by a summation of the 36 subbands. Since it is not possible to construct bandpass filters with a perfectly square frequency response, some aliasing will be introduced by the decimation. The alias contribution will be taken care of at a later point in the decoder.

The samples in each subband are still in the time domain, but calculating the total energy within each frequency band results in a energy distribution in the frequency domain. The output from the analysis polyphase filter bank can therefore be considered as a coarse spectral representation of the time samples, using only 36 spectral values.

MDCT: In this process the 32 subbands are mapped into a Modified Discrete Cosine Transform (MDCT) representation. Performing this transformation will enhance the spectral resolution to 36 frequency lines per subband. Prior to the transformation a windowing of the subband samples is performed. The windowing can be performed using either long windows or short windows depending on the dynamics within each subband. If the subband samples within a given subband shows a stationary behavior, a long windows is chosen in order to enhance the spectral resolution in the following MDCT. If the subband samples contains transients, three consecutive short windows are applied in order to enhance the time resolution in the following MDCT. In order to obtain better adaption when window transitions are required, two windows referred to as start windows and stop windows, are defined. In Figure 2.2 the four applicable window types are illustrated. 18 frequency lines output by the **MDCT** block after applying either a long window or three short windows are referred to as a “block”, that is a start block, an end block and a short block, corresponding to the start, short and stop windows respectively.

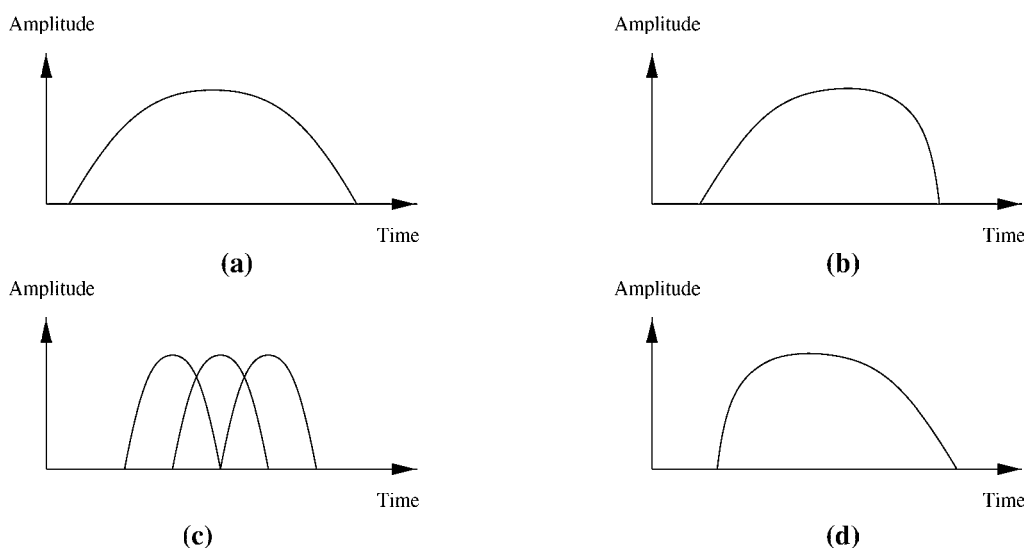


Figure 2.2: Illustration of the four applicable window types. (a) corresponds to normal window type, (b) corresponds to start window type, (c) correspond to short windows type and (d) correspond to stop window type.

The decision on which window types to apply is controlled by the Psycho Acoustic Model introduced later. Figure 2.3 shows an example of a sequence of windows applied to a subband.

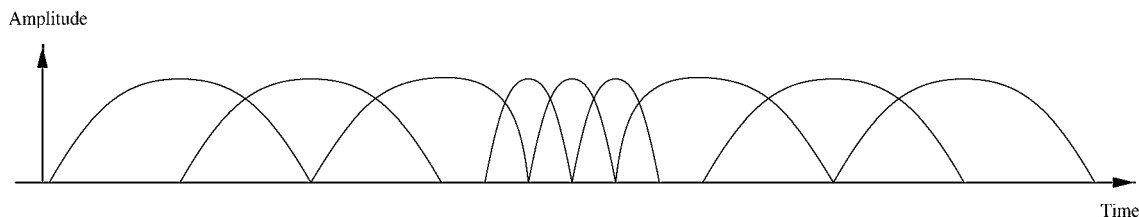


Figure 2.3: Illustration of a typical sequence of windows to be applied to a subband.

Performing an MDCT on a subband windowed with either of the long windows will produce 18 frequency lines due to 50 percent overlap. Using 3 short windows with 50 percent overlap will

produce 3 groups of 6 frequency lines, each group belonging to different time intervals. Having performed one MDCT transformation will therefore produce 576 frequency lines referred to as a granule. The MDCT block produces 2 granules when transforming 32 subbands.

Before passing on the frequency lines a reduction of the aliasing introduced in the Analysis Polyphase Filter Bank is removed. The aliasing is removed at this early stage in order to reduce the amount of information for transmission. The reduction is obtained by means of a series of butterfly computations, see Figure 2.11.

In Figure 2.4 all the processing applied to the 1152 PCM samples are illustrated in one figure. At this point the signal processing model has transformed the initial PCM samples into 32 equally spaced subbands within the audio signal.

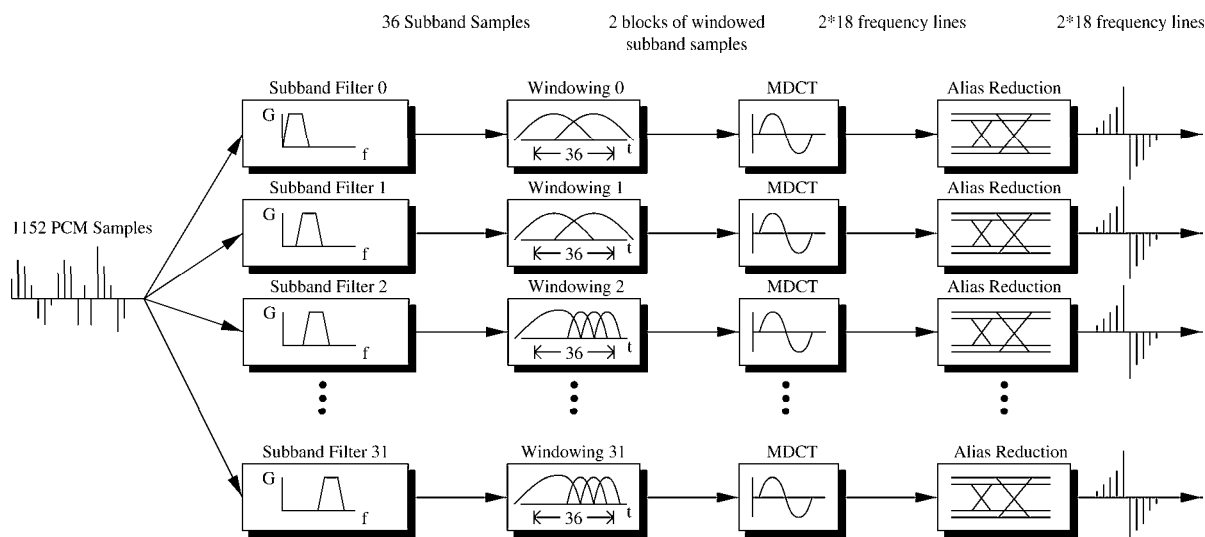


Figure 2.4: Summary of the signal processing applied to the 1152 PCM samples.

FFT: Concurrent to the Analysis Polyphase Filter bank calculations, two Fast Fourier Transformations (FFT) of the 1152 PCM samples are performed. Both a 1024 and a 256 point FFT are performed in order to provide a fairly high spectral resolution and information on the spectral changes over time. Obviously 1024 point will not give an optimal frequency domain representation of the 1152 samples in consideration. The number is chosen because an FFT, demands a power of two samples. Compared to an ordinary 1152 point Discrete Fourier Transform less than one percent computation power is required [Oppenheim and Schaffer, 1989].

Psycho Acoustic Model: This block contains a set of algorithms that constitutes the Psycho Acoustic Model. In short, the Psycho Acoustic Model is a model of the human sound perception. The model is used in the encoder only, in order to decide which parts of the audio signal have perceptual relevance and which parts have not. The results of the psycho acoustic evaluations are utilized in the MDCT block and in the Nonuniform Quantization block.

The Psycho Acoustic Model supplies the MDCT block with information about which window type to apply. The decision is based on a measure of the differences between the two presently calculated FFT spectra and the two previous spectra. If a certain difference is present the Psycho Acoustic Model will call for a transition to short windows. When the differences fade away, a transition back to long windows will be called for.

The Psycho Acoustic Model also supplies the Nonuniform Quantization block with information on how to quantize the frequency lines. The quantization of the frequency lines is adapted to the limitations of the human ears perception of audio. Heuristic experiments have shown

that the human ear has 24 frequency bands, called critical bands [Furui, 1989], in which the human ear is less frequency selective. When a dominant tonal component is present in an audio signal, frequencies in the associated critical band are not perceived very well. The dominant tonal component introduces a masking threshold below which frequencies in the same critical band are masked out. This effect allows for a more coarse quantization of the nearby masked frequency components, without introducing audible degradation. In the MPEG/Audio standard critical bands are approximated by scalefactor bands. The Psycho Acoustic Model analyzes the FFT spectrum to detect whether dominant tonal components are present. When dominant tones are present a masking threshold is calculated. Based on this threshold an upper limit for the quantization level required in the individual scalefactor bands is determined. Two psycho acoustic models applicable to the Layer III encoder are described in [ISO/IEC 11172-3, 1993], but other psycho acoustic models optimized toward certain properties can be implemented as well.

Nonuniform Quantization: In this block a nonlinear quantization of the frequency lines is performed. The nonlinearity is introduced by first raising each sample to the power of $3/4$. In order to further reduce the quantization noise, a scaling of the frequency lines in each scalefactor band is performed prior to the nonuniform quantization. Hence the output from this block is quantized frequency lines and scalefactors applying for each of the scalefactor bands. The scaling and a grouping of the frequency lines into scalefactor bands is performed in accordance with the psycho acoustic model.

Huffman Encoding: In this block a coding of the scaled frequency lines is performed using the Huffman coding algorithm based on 32 static Huffman tables. The Huffman coding scheme is one of the major reasons, why the MPEG/Audio Layer III algorithm can retain a high audio quality at low transmission bit rates. The coding scheme provides lossless compression and thereby reduces the amount of data to be transmitted without degrading the quality.

The reduction of data is obtained through a coding of the entropy, e.g. the likeness, inherently present in most data. In Huffman coding the entropy is based on a statistic distribution of the group of data values considered. From the data statistics a substitution table covering all data values is established. In this table data values with a high probability of being present in the data are associated with short code words and data rarely present are associated with longer code words. Encoding a block of data will produce a new representation of the exact same data, but using fewer bits.

In order to enhance the compression rate both granules are subdivided into smaller partitions and encoded in pairs or quadruples. The partitioning enables the use of different Huffman tables to compress the quantized frequency lines. How to perform the partitioning is further explained in Section 4.2.

Coding of Side Information: In order to enable the decoder to reproduce the audio signal all parameters generated in the encoder must be provided. The Coding of Side Information block orders all the parameters used through the coding process. e.g. boundaries for certain data blocks, quantizer step sizes, which windows are used for MDCT etc.

Bitstream Formatting CRC Word Generation: In this block the Huffman coded frequency lines, the side-information and a frame header are assembled to form the bitstream. The bitstream is partitioned into frames each representing 1152 PCM samples. A Cyclic Redundancy Code (CRC) can optionally be included for data validation in the decoder. In Subsection 2.2.1 the frame format is further explained.

Ancillary Data: This data is typically used for features like informing on artist name or music category. The Ancillary Data area is user defined and optional for the designer to implement. The feature often appears in digital radio broadcasting.

2.1.1 Stereo Encoding

The introduction of the Layer III encoder presented so far only applies for encoding of single channel audio. The MPEG/Audio standard also defines a method for encoding a dual channel audio signal and two methods for stereo redundancy encoding. In the following only a short introduction to the three methods will be conducted in order to describe how the signal processing model in Figure 2.1 can be applied for dual channel and stereo encoding. More information on the stereo modes will be presented in the next section describing the Layer III decoder.

Encoding of a dual channel signal, i.e. a bilingual or non processed stereo audio signals, is accomplished by time sharing the processing model in Figure 2.1. Encoding a dual channel signal does not introduce extra complexity, because the two channels are encoded independently of each other.

The two stereo redundancy encoding methods supported in MPEG/Audio are Middle Side stereo (MS) and Intensity stereo. The first method transmits the stereo signal as the sum and the difference of the two channels. The two new channels are both transmitted as described for the dual channel model.

Intensity stereo requires only one channel for transmission. In this channel the sum of the two audio signal channels is transmitted.

2.2 MPEG/Audio Layer III Decoder Overview

Next the Layer III decoder will be described at functionality level. The description will be based on Figure 2.5.

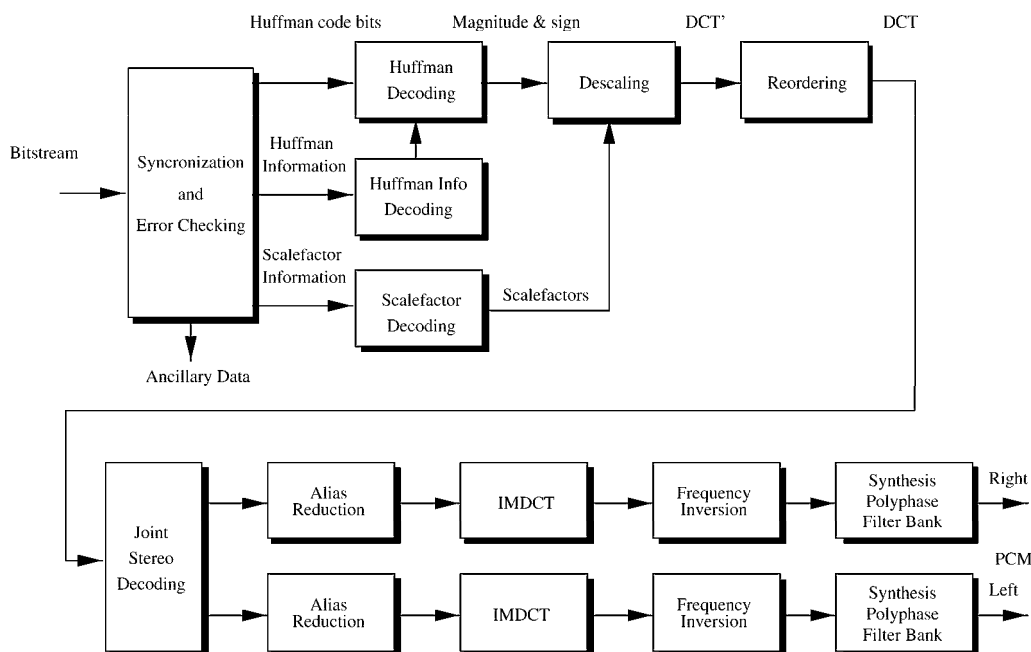


Figure 2.5: Block diagram of MPEG/Audio decoder.