# MPEG-1 Layer III Standard A Simplified Theoretical Review

**Article** · October 2019

CITATIONS
0

READS
10

**1 author:**

Joash Bii
Maasai Mara University
**3** PUBLICATIONS **0** CITATIONS

SEE PROFILE

ISSN: 2277 – 9043
*International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*
*Volume 8, Issue 10, Oct 2019*

# MPEG-1 Layer III Standard: A Simplified Theoretical Review

**Joash Kiprotich Bii**[1]

***Abstract*— MP3 is a standard that is used for encoding/decoding audio data. The standard can lower audio bit rate significantly without any loss. For this reason, it is key to understand how it does so, and secondly, find out if it is doable. Raw audio signals carry large data quantities and are neither suitable for transmission nor storage [2]. Therefore it is necessary to compress audio and at the same time maintain its quality as required by the International Organization for Standardization (ISO). MP3 was developed by the Motion Pictures Experts Group for audio and video compression. It is composed of three modes; the third referred to as Layer III. It is this layer that lowers raw-audio data bit rates from i.e. 1.4 Megabits per second to just 128 kilobits per second and can still reconstruct the signals to a level comparable to the original [3]. The objective of this paper is to review and provide a simple idea of what Layer III does in relations to audio compression and decompression.**

***Index Terms*— Audio, Compression, MP3, Encoding, Decoding**

## I. INTRODUCTION

It has been shown that there is a limit to how much data can be compressed without losing any information [4]. If compressed data is decompressed and the bitstream is identical to the original, it is referred to as lossless. The limit of compression depends on the probabilities of the bit sequences. Entropy coding is used in lossless compression. This kind of compression is essential when no data loss is tolerable. If compressed data is decompressed and the bitstream is not as identical as the original, it is referred to as lossy. Some amount of distortion is tolerated. Speech and images can be compressed by this lossy method as all details may not be of interest. Three basic lossless compression methods include: Run-Length Encoding (RLE), Move-To-Front Encoding (MTF), and Huffman Coding (HC).

- In Run-Length Encoding, a run is made for repeated bits and coded in lesser bits by only stating how many bits were there [5].
- In Move-To-Front encoding, it holds with the property that the occurrence of a character indicates that it is more likely to occur immediately afterwards [6].The positions of the symbols to be compressed are used to build the first table.
- In Huffman Coding, the characters in a data file are converted to a binary code [7], with the most occurring-higher probability characters being given shorter binary codes and the least occurring-lower probability characters being given longer binary codes and then for each symbol, a binary tree is generated.

## II. THE MPEG-1 LAYER III FORMAT

MPEG developed a standard with three different parts: Audio, video and system parts. The system part explains how to transmit audio and video signals on a single distribution media [1]. Video and audio was possible to transmit (1-2 Mbit/s) using MPEG-1. The audio has three layers or levels of compression: I, II and III, with III being very efficient, as it is capable of compressing audio by a factor of 12 without much data loss. The other two are 4 and 8 factors respectively. MP3 utilizes perpetual coding, a lossy process that tilters inaudible sounds [8]. Every human band is approximated by scale factor bands and a masking threshold worked out for each band, then scaled to reduce quantization noise of the frequency lines [9]. Huffman Coding is finally applied at layer III to further improve compression. Bitrate allows us to choose the quality of the encoding. Layer III standard defines 8 kbit/s - 320 kbit/s bitrates, with a default of 128 kbit/s. Two types of bitrates are specified here: Constant Bitrate (CBR) and Variable Bitrate (VBR). CBR encodes audio with same amount of bits while VBR handles the complex audio songs that cannot be encoded with CBR by allowing the bitrate to vary depending on the dynamics of the signal [1]. To set the quality, a threshold is specified to inform the encoder of the maximum allowed bitrate. The disadvantages of VBR are: Firstly, decoder timing is challenging and secondly broadcasting is needed by CBR. Resolution for the audio depends on the frequency of sampling. High bitrate produces better resolution while high sampling frequency enables the storage of more values, and in turn a bigger spectrum. MPEG-1 defines audio compression at 32 kHz, 44.1 kHz and 48 kHz [10]. [1] gives the anatomy of MP3 file and shows that because bitrate determines sample size, increasing it will cause an increase in frame size which depends on the frequency of sampling as follows:

$$\frac{144*bitrate}{Frequencysample} + Padding[bytes]$$

A frame has of five parts; header, CRC, side information,

[1] ***Bii Joash*** *is with School of Science, Department of Computer Science and Information Technology at Maasai Mara University, 861 - 20500, Narok, Nairobi.*

main data and ancillary data as shown in fig 1.

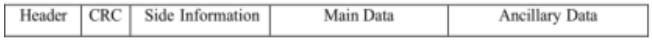


Figure 1: Parts of MP3 frame [1]

- Frame header
  Header is composed of 32 bits made up of a synch-word plus a description. This enables receivers to hook to the carrier, making it possible to broadcast a file. MPEG version is specified within the synch bit. When the protection bit is 1, CRC is used. A 4 bits bitrate informs the decoder what bitrate the frame is encoded. 1 bit private bit is used for application of triggers. 2 bits mode shows what channel is used The Copyright bit, if set it means that it is illegal to copy the contents. If home and 1 bit is set it shows that the frame is located in its original media. 2 bits emphasis bit informs the decoder that the file must be equalized again.


Figure 2: MP3 Header composition

- Side Information:
  This section contains what is important to decompress the main data. Its size relies on the mode compressed by the channel. When a single length value is written, the size of the field is constant. When two, then the first is used in mono mode and the second in all other modes. The bit reservoir technique in MPEG Layer III enables the left over free space in the main data area of a frame to be used by consecutive frames [1]. In order to locate where the main data of a certain frame starts, the decoder reads the main data begin value. Static frame parts are not included in the offset. Information for the Scale Factor Selection shows weather the same scale factors are transferred or not and transmitted four bits/channel. When a bit belonging to a scale factor band is zero the scale factors are transmitted for each granule [1]. Windows are utilized in channels. Twelve bits channels of single mode and double for stereo mode. Frequency lines of each channel are not coded with using similar code table of Huffman. Ranging from zero to the nyquist frequency, the frequencies are categorized into five regions to allow different Huffman tables of the spectrum and to enhance performance. Using maximum values that have been quantized, partitioning is created with the supposition that values at high frequencies are expected to have lower amplitudes. The big values field has the size of the partition. Scale factors compression tells the number of bits to be used for transmission. Twelve (for block type 2, short windows) or twenty one (for block type 0, 1, 3 long windows) scale factor bands could be generated from a channel. The windows switching flag shows whether another window than the normal is in use. When the flag of the block is mixed and set to two sub bands lower, they are transformed using a window and the remaining bands are trans-formed using the window in the block variable type. The MP3 standard is composed of Huffman codes in 32 tables, with entries of the fields informing which set to use during decompressing. When pre flag is active, the entries are added to the scale factors. Logarithmic quantization is then performed on scale factors with a step size of two.

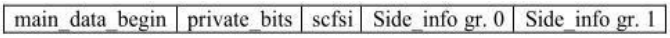


Figure 3: MPEG-1 Layer III side information

The regions of the frequency spectrum are shown in figure 4.

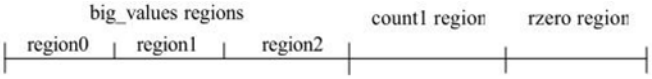


Figure 4 : Regions of the frequency spectrum

- Data:
  It consists of three parts namely the bits in Huffman, the scale factors, and the ancillary data. The work of scale factors is to lower the quantization noise. Scfsi field indicates whether scale factors are shared or not [3]. Band division involving the spectrum to scale factors is redone for each window size. The Huffman code bits stores information on how to decode. Hlen indicates the length of the Huffman code for x and y respectively. Hcode represents x and y codes. Rzero region is run-length coded because of all of the zero values. Since ancillary data is not explicitly provided, the ancillary is optional. The organization in granules and the scale factor channels are shown in figure 5 below.

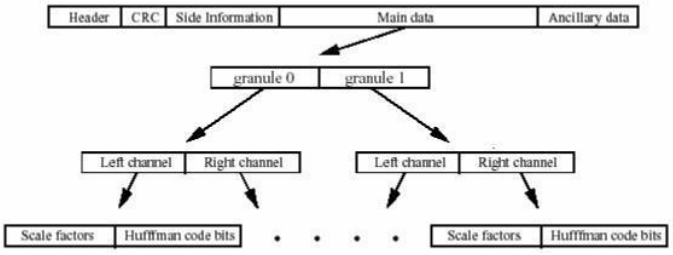


Figure 5: Organization of scale factors in granules and channels

## IV. ENCODING

MPEG-1 Layer III encoding scheme involves the analysis of the polyphase filter bank, transformation by Discrete Co-sine (modified), fast fourier transformation, phychoacoustic modeling, quantization (non-uniform), Huffman coding, and formatting CRC bitstream for word generation as depicted in Figure 6 below.

### *A. Analysis Polyphase Filterbank*

Through a polyphase filterbank analysis, a sequence of 1152 PCM samples are filtered into 32 equally spaced frequency sub bands depending [14]. Samples with carrier components are filtered into bands of bands, i.e. sub bands making the number of samples increase since every sub band

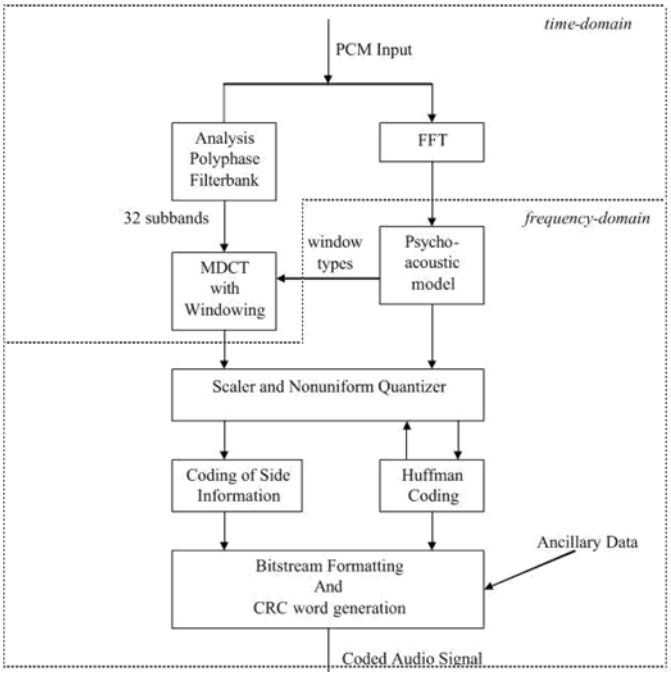


Figure 6: MPEG-1 Layer III encoding process

stores sub spectra of the sample. An example by raissi (2002) shows that filtering 100 samples increases the number of samples to 3200, which then are decimated by a factor 32 to lower the samples back to 100 [1]. It is impossible to construct band pass filters with a perfect frequency response, as some aliasing will be introduced by decimation.

*B. Modified-DCT*

After windowing (to reduce artifacts imposed by the signal segment edges), to split the 32 sub bands, (modified) DCT is applied and it splits into 18 sub bands making a granule of 576 lines [11]. MPEG defines four windows: normal, start, short and stop. Psychoacoustic model indicates what window type to apply based on the stationarity degree, then it forwards the information. If sub band signal frame presently shows small difference from the previous time frame, it applies the long window to enhance spectral resolution provided by the Modified DCT, otherwise the short windows is applied. To control artifacts, a high time resolution is needed. To achieve good adaptation during transitions, start and stop windows are utilized. Window long gets short, and vice versa, and then the aliasing by the polyphase filter bank is withdrawn to lower the amount of information needed for transmission.

*C. Fast Fourier Transform - FFT*

As the polyphase processes the signal, it also converts it to the frequency in the domain by Fast Fourier Transform method. The Fast Fourier Transform points are created on the samples of PCM to achieve a high frequency resolution on the spectral over time.

*D. Psychoacoustic Model*

The model retrieves data from Fast Fourier Transform, which are then applied to a set of algorithms to model the human sound perception and provide information about which sections of the audio signals are audible and which are not. Two present FFT and two previous spectra are matched in order to know which window to send to Modified DCT. If not similar, the short window is applied, and if similar, then Modified DCT is made aware to change to long windows.

To calculate band masking thresholds, the spectra is analyzed to find tonal components, and those frequencies below it are masked out.

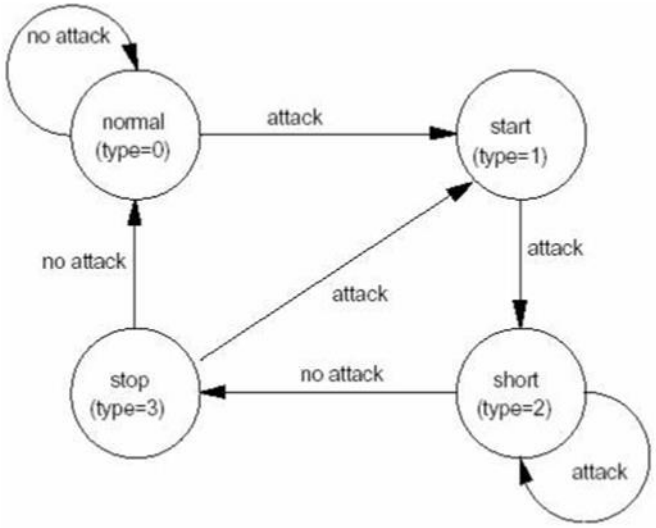


Figure 7: states representing Window switching decision

*E. Non-uniform Quantization*

For 576 values, Huffman coding and quantization are applied to repeatedly, in two nested rounds, and an outer loop is created to control distortion while an inner loop to control the rate. Frequency domain samples are quantized by the rate control loop and then the samples are determined.

Firstly, samples are quantized at increasing step sizes and then coded by Huffman code tables.

Secondly, bigger step sizes lead to smaller values, and then overall codded bit sum is calculated and matched with avail-able bits. If sum is larger than the number of bits, the size is increased and process repeated until it is sufficient. Distortion control loop is used to reduce the noise in quantization under the masking threshold for each scale factor band. The result is stored after this, and then the noise value is further reduced until it is no longer there within the loop and then the process ends.

*F. Huffman Encoding*

Values already quantized are coded using different tables. Layer III keeps a very high quality at low bitrates because of Huffman coding. Compression values are collected to allow the decoder or decompressor to regenerate the audio carrier. The values are then placed at the side information section.

*G. Stream and CRC creation*

By 1152 encoding procedure, a stream of bits is created containing the header, cyclic redundancy checker, a side info frame, and coded huffman line frames.

## V. DECODING

Decompressing or decoding MP3 data involves the reverse of the process described above, including huffman decoding, error checking, scale factor decoding, re-ordering, inverse transformation DC and basically inverse frequency processes. Fig 8 describes this scheme for MPEG-1 Layer III decoding.

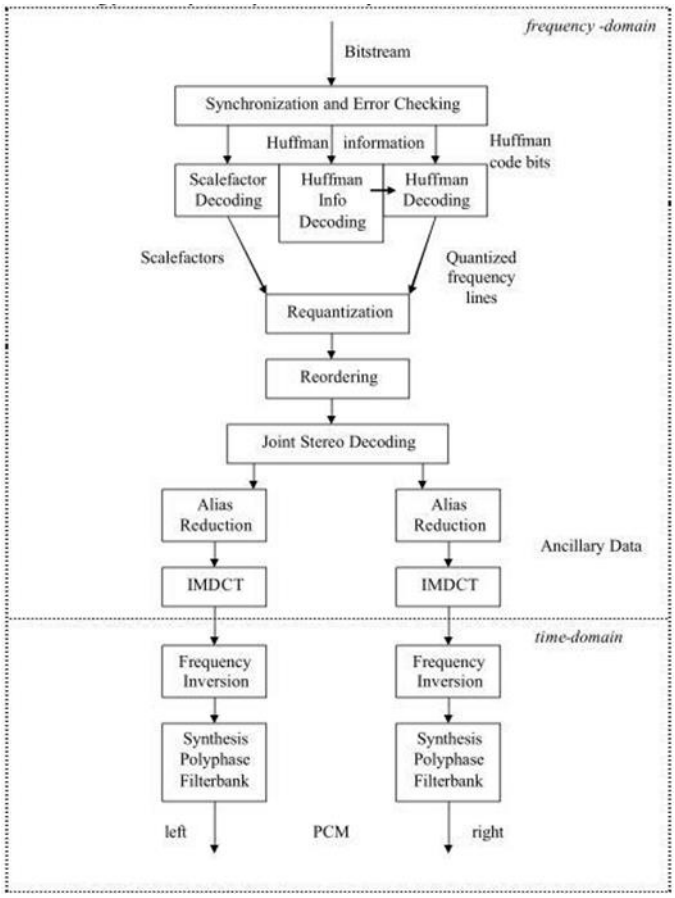


Figure 8: The decoding process of MP3 [1]

### *A. Sync and Error Checking*

It receives an incoming stream and then looks or searches to identify individual synch word in each incoming frame.

### *B. Huffman Decoding & Huffman info decoding*

Since a single code word at the center of the bits cannot be coined, as Huffman coding is a varing length algorithm, de-coding therefore starts where code word starts. The algorithm gives parameters to do appropriate decompression and the info unit insures that the frequency lines (576) are created. If lesser, the info decoding block initiates zero padding to recover the absence of data.

### *C. Scale factor decoding*

Decompression details are found inside the side info section, and are utilized for re quantization of the scale factors.

### *D. Re-quantizer*

To restore the lines of frequency, the scale factor values, gain, and flag columns are utilized, as they were created by the Modified DCT in the encoder for that purpose. The output is then re-quantized with the scale factor entries, the gain and flag columns and two equations as [1] describes, are applied based on the window that was used. The inverse power of 4/3 is then used to raise the two equations to the desired result then followed by re ordering sub frequency bands.

### *E. Re-ordering*

Frequency lines for short windows are re-ordered to sub bands to increase the general efficiency of the Huffman coding, followed by the other frequencies and then the last window. A search for windows that are shorter in the 36 bands is done, and when found, they are again re-ordered repeatedly until none is present in the bands.

### *F. Stereo Decoding*

In order to convert encoded stereo signal into different left or right signals, stereo decoding is necessary. The header of a frame contains the encoding method and is read from it.

### *G. Alias Reduction*

To generate a true audio signal, the artifacts of aliasing must be included to the signal once more. Generation involves a calculation that has 8 other calculations in form of a butterfly in every band. Using shorter blocks, aliasing is applied to the granules.If long block granules are present, i.e. those with more than value 2, the feed to synthesis filterbank is worked out using the pseudo code described in [15], for alias reduction followed with the Inverse Modified transformation, DCT.

### *H. Inverse Modified - DCT)*

At this stage, a mapping of the frequency lines is done to the 32 polyphase filter bands and IM-DCT outputs 18 samples for each. In [16], it delivers a clear IMDCT mapping and describes every bit in detail.

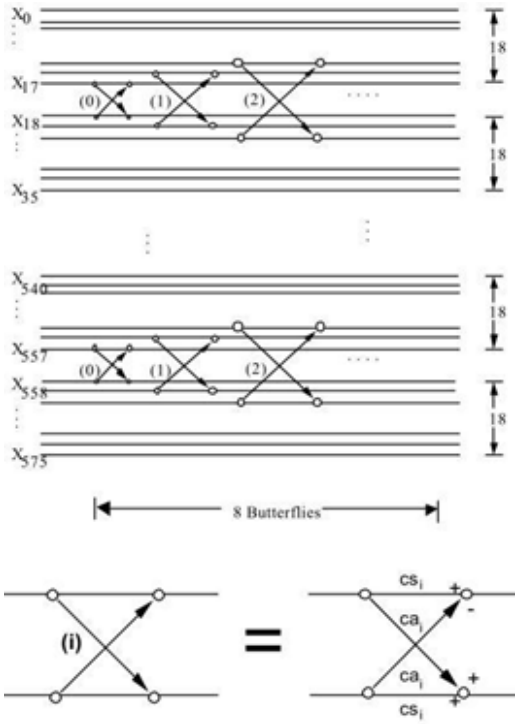


Figure 9: Alias reduction - coutesy of [1]

*I. Frequency Inversion*

Inversions in frequency require balancing uneven samples with uneven sub frequencies, as filtering phase may amend bits sequences. This process is undertaken for each uneven time sample of every uneven sub frequency.

*J. Synthesis Polyphase Filterbank*

Finally, a transformation of 32 sub frequency bands (18 time domain samples in every granule) is modulated to give 18 blocks of 32 samples of Pulse code through the synthesis filter bank.

## VI. CONCLUSION

Audio compression at layer III contains several sub pro-cedures but achieves maximum compression. The standard uses a psychoacoustic model to show non perceptible signals while the filter banks and DCTs effectively maps the frequency and the time-domains, then scales and quantizes the values and eventually applies Huffman coding algorithm. It combines loseless and lossy methods to perform compression as the two combined compresses the data to the required minimum or audible minimum. Because this layer only reconstructs the bits while not applying encoded data psychoacoustics, it therefore lessens the generation of MP3 decoder significantly.

## REFERENCES

[1] Rassol Raissi, "The theory behind MP3", Dec 2002. https://www.semanticscholar.org/paper/The-Theory-Behind-Mp3-Raissi/ed14d9a452c4a63883df6496b8d2285201a1808b

[2] Mahdi, O.A.; Mohammed, M.A.; Mohamed, A.J., 2012. "Implement-ing a Novel Approach an Convert Audio Compression to Text Coding via Hybrid Technique". International Journal of Computer Science Issues.

[3] E.KALPANA, VARADALA SRIDHAR and M.REJENDRA PRASAD, 2012 MPEG-1/2 audio layer-3(MP3) ON THE RISC based ARM PROCESSOR (ARM92SAM9263), International Journal of Computer Science Engineering (IJCSE) ISSN: 2319-7323 Vol. 1 No. 1.

[4] Shannon, C. E., and Weaver, W. 1949. The Mathematical Theory of Communication, University of Illinois Press, Urbana.

[5] https://en.wikipedia.org/wiki/Run-length encoding

[6] https://en.wikipedia.org/wiki/Move-to-front transform media at up to about 1,5 Mbit/s Part 3

[7] Harsimran Kaur1 and Balkrishan Jindal, 2015 Lossless text data compression using modified huffman coding a review ISSN 2319-5991 Vol. 4, No. 4

[8] Telecom ABC http://www.telecomabc.com/p/perceptual.html

[9] TDouglas L. Jones, 2012 Digital Filter Structures and Quantization Error Analysis. Online: http://cnx.org/content/col10259/1.1/

[10] Karlheinz Brandenburg, 1999 MP3 and AAC Explained, AES 17th International Conference on High Quality Audio Coding

[11] Qingzhong Liu, Andrew H. Sung and Mengyu Qiao, 2010 Detection of Double MP3 Compression ISSN: 1866-9964

[12] ISO/IEC 11172-3 Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s Part 3

[13] International Organization for Standardization webpage, http://www.iso.ch

[14] Qingzhong Liu, Andrew H. Sung and Mengyu Qiao, 2010 "Detection of Double MP3 Compression" Institute for Com-
plex Additive Systems Analysis New Mexico Tech, Socorro.
https://www.researchgate.net/publication/220340299

[15] Mariella Baldussi, 1995 "Decoding of ML Data for Layer III" ISO/IEC 13818-3 International Standard. Found at http://www.lim.di.unimi.it/IEEE/MPEG2/M225384.HTM

[16] Miroslav Galabov, DECEMBER 2004. "Implementation of IMDCT Block of an MP3 Decoder through Optimization on the DCT Matrix", DOAJ VOL. 13, NO. 4