



**Universidade Federal de Campina Grande**  
**Centro de Engenharia Elétrica e Informática**  
**Curso de Graduação em Engenharia Elétrica**

RUBENS BARRETO LEAL

**ESTUDO DO PADRÃO MPEG-1 AUDIO LAYER 3 PARA  
COMPRESSÃO DE ÁUDIO**

*Campina Grande – PB, junho de 2016*

RUBENS BARRETO LEAL

ESTUDO DO PADRÃO MPEG-1 AUDIO LAYER 3 PARA  
COMPRESSÃO DE ÁUDIO

*Trabalho de Conclusão de Curso submetido à  
Unidade Acadêmica de Engenharia Elétrica da  
Universidade Federal de Campina Grande  
como parte dos requisitos necessários para a  
obtenção do grau de Bacharel em Ciências no  
Domínio da Engenharia Elétrica.*

Professor Orientador:  
EDMAR CANDEIA GURJÃO

*Campina Grande – PB, junho de 2016*

RUBENS BARRETO LEAL

ESTUDO DO PADRÃO MPEG-1 AUDIO LAYER 3 PARA  
COMPRESSÃO DE ÁUDIO

*Trabalho de Conclusão de Curso submetido à  
Unidade Acadêmica de Engenharia Elétrica da  
Universidade Federal de Campina Grande  
como parte dos requisitos necessários para a  
obtenção do grau de Bacharel em Ciências no  
Domínio da Engenharia Elétrica.*

Aprovado em 15/06/2016

---

**Professora Luciana Ribeiro Veloso**  
Universidade Federal de Campina Grande  
Avaliadora, UFCG

---

**Professor Edmar Candeia Gurjão**  
Universidade Federal de Campina Grande  
Orientador, UFCG

## AGRADECIMENTOS

Primeiramente, gostaria de expressar aqui imensa gratidão a minha família, especialmente a meus pais por terem me proporcionado, com muito esforço, todas as condições e sustento para que eu tivesse uma educação de qualidade. Obrigado por terem sido as pessoas diretamente responsáveis por eu ter superado os momentos sombrios e de extrema dúvida que quase me fizeram desistir.

Agradeço a todos os professores que contribuíram para a minha formação, especialmente ao professor Edmar Candeia por ter me orientado neste trabalho.

Agradeço também à toda a coordenação administrativa do curso de Engenharia Elétrica, especialmente a Tchai, Adail e ao professor Damásio por toda a paciência, apoio e hospitalidade a mim demonstrados.

E por fim, gostaria de agradecer diretamente a Túlio, Thais, Nelson, Nicolau, Phelipe, Eduardo, Revson e Emanuelle por terem sido as pessoas com quem eu passei os meus momentos mais felizes durante minha vida universitária. Sei que nossos caminhos daqui para frente serão distintos, mas garanto que a amizade de vocês ficará para sempre guardada na minha memória.

*"Se você sente solidão quando está só, está em má companhia"*

Jean-Paul Sartre

## RESUMO

Este trabalho de conclusão de curso tem como objetivo principal efetuar um estudo da parte de áudio do padrão MPEG-1 de compressão e transmissão de áudio e vídeo, mais especificamente o padrão MPEG-1 Layer 3, mais conhecido como o formato de codificação MP3. São estudadas as principais características deste padrão, bem como os conceitos essenciais necessários para o seu entendimento. São também estudados os procedimentos de codificação e decodificação utilizados, buscando entender o seu funcionamento. Busca-se também compreender como é possível obter, com este formato de codificação, resultados de compressão tão significativos com o mínimo de degradação no sinal de áudio original.

## ABSTRACT

The main purpose of this work is to study the audio part of the MPEG-1 audio/video compression and transmission standard, more specifically the MPEG-1 Layer 3 standard, better known as the MP3 coding format. The study aims the main features of this format, as well as the essential concepts required for a better understanding. The study also aims to understand the coding and decoding procedures associated with the format. And finally, we also try to understand how it is possible to obtain, with this coding format, such significant results in the compression rate as well as minimum impact in the original audio signal.

# SUMÁRIO

1	Introdução .....	9
2	Noções de compressão de dados .....	10
3	Base teórica e conceitos essenciais .....	11
3.1	Psicoacústica e mascaramento de frequências .....	11
3.2	Modulação por codificação de pulsos .....	13
4	Características do padrão MPEG-1 Audio Layer 3 .....	14
5	Codificação MP3.....	15
6	Decodificação MP3.....	19
7	Estrutura de um arquivo MP3 .....	21
7.1	Estrutura de um frame.....	21
7.2	ID3 .....	24
8	Conclusões .....	26
9	Referências Bibliográficas .....	27

# 1 INTRODUÇÃO

A tecnologia de compressão do padrão MPEG-1 Audio Layer 3, mais comumente conhecido como MP3, é largamente utilizada, valendo-se destacar a indústria fonográfica e de entretenimento. Esta poderosa tecnologia baseia-se em um esquema de codificação/decodificação que possui o intuito de comprimir os sinais de áudio, de forma a produzir o mínimo de degradação no sinal audível.

Sinais digitais de áudio não-comprimidos com qualidade elevada consomem um grande volume de dados, inviabilizando o processo de armazenamento e transmissão destes sinais. Desta forma, é necessário que este grande volume de dados seja reduzido, ao mesmo tempo em que deseja-se o mínimo ou praticamente nenhuma perda da qualidade perceptível do sinal. Sendo assim, a International Organization of Standardization (ISO) estabeleceu que seria necessário um padrão que possibilitasse este nível de compressão de dados. Um dos grupos internos a ISO, que é a Moving Pictures Experts Group (MPEG) desenvolveu este padrão desejado, que contém um conjunto de técnicas de compressão não só para sinais de áudio, mas também para sinais de vídeo. Sendo assim, a parte do padrão referente a sinais de áudio inclui três camadas principais [12], das quais temos como objeto de estudo a terceira (Layer 3), que é o MP3. O MP3 permite efetuar excelentes níveis de compressão para sinais de áudio em alta qualidade, como os que estão presentes em CDs de conteúdo musical.

É importante enfatizar que o MP3 é um método complexo, mas bastante eficiente de compressão de áudio. Desta forma, sabendo da importância e do impacto que esta tecnologia possui, temos como propósito efetuar o estudo de toda a fundamentação teórica e das técnicas envolvidas neste padrão de compressão.

## 2 NOÇÕES DE COMPRESSÃO DE DADOS

No ano de 1949, o matemático americano Claude Elwood Shannon formulou aquela que seria a primeira teoria sobre compressão de dados. Nesta teoria, ele foi capaz de provar que existe um limite para efetuar uma compressão de dados sem perda de informação. Desta forma, quando fosse efetuada a descompressão destes dados comprimidos, a informação contida poderia ser entendida de forma praticamente idêntica ao que era antes da compressão ser efetuada.

Este tipo de compressão, em que não há perda de informação, é denominada de *compressão sem perdas*. O limite para que haja uma compressão de dados sem perdas de informação depende das probabilidades de ocorrência de determinadas sequências de bits associadas no conjunto de dados. Desta forma, é possível, por exemplo, detectar sequências de bits que são repetitivas e substituí-las pelo seu número de ocorrências, ocasionando a compressão desejada pela eliminação de redundâncias na informação.

Além da compressão sem perdas, existe também a *compressão com perdas*. Neste segundo tipo, a informação lida após a descompressão dos dados não necessita ser exatamente a mesma que era antes da compressão. Neste caso, existe uma tolerância a certos níveis de perdas na informação, que não necessariamente impactam na sua leitura. Um bom exemplo de compressão com perdas seria a sua aplicação na compressão de imagens, em que não se faz necessário que sejam identificados todos os detalhes possíveis para que se possa entender a informação ali contida. Sendo assim, a compressão com perdas é capaz de fornecer um resultado melhor em termos de diminuição do volume de dados, fazendo com que, dependendo da aplicação, este tipo de compressão seja mais interessante de ser utilizada. Quando não há tolerância alguma para perda de informação, se faz necessária a utilização de uma compressão sem perdas. Neste caso, a preservação total da informação original é garantida, mas há uma menor taxa de compressão.

## 3 BASE TEÓRICA E CONCEITOS ESSENCIAIS

### 3.1 PSICOACÚSTICA E MASCARAMENTO DE FREQUÊNCIAS

Entende-se por Psicoacústica como sendo o estudo de como o ouvido humano e o sistema nervoso interagem para que haja a percepção do som. Isto inclui como somos capazes de ouvir e quais são as respostas fisiológicas e psicológicas aos estímulos sonoros [7].

No que diz respeito a nossa percepção sonora, a maioria dos seres humanos não consegue perceber frequências que estejam abaixo de 20 Hz ou que estejam acima de 20 kHz. Esta largura de banda tende a ficar mais estreita a medida em que envelhecemos. No caso de um homem de meia idade, por exemplo, ele provavelmente não conseguirá perceber sons com frequências acima de 16 kHz. Frequências que variam entre 2 kHz a 4 kHz são as mais fáceis de serem percebidas pelo ouvido humano, sendo assim elas podem ser detectadas com uma intensidade (ou volume) sonora relativamente baixa. A medida em que as frequências são aumentadas até o final da largura de banda audível, a intensidade sonora também precisa aumentar para que o nosso ouvido consiga detectar estas frequências. Este tipo de comportamento pode ser visualizado no gráfico da Figura 1, em que descreve a variação do *Nível de pressão sonora* (*Sound Pressure Level*), abreviado como *SPL*, de acordo com a frequência em escala logarítmica.

O Nível de pressão sonora descreve a medida para determinar o grau de potência de uma onda sonora [13], e pode ser calculado pela seguinte expressão, dada em dB:

$$\beta = 10 \cdot \log\left(\frac{I}{I_0}\right)$$

Em que  $I$  é a intensidade sonora medida e  $I_0$  é a intensidade sonora de referência dada por  $I_0 = 10^{-12}W/m^2$ .

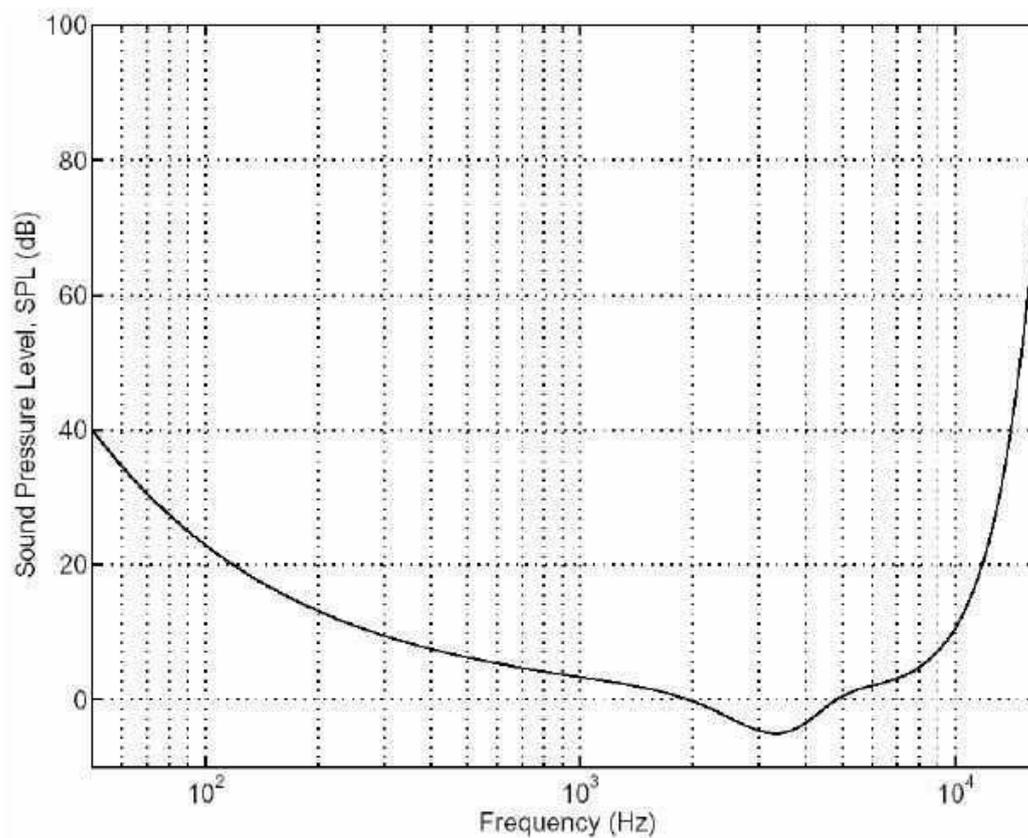


Figura 1 - Variação do Nível de pressão sonora de acordo com a frequência (fonte: [https://www.researchgate.net/figure/209435973\\_fig3\\_Figure-23-Absolute-Threshold-of-Hearing](https://www.researchgate.net/figure/209435973_fig3_Figure-23-Absolute-Threshold-of-Hearing))

Sabe-se então que o ser humano não consegue perceber todas as frequências presentes em um sinal sonoro. Desta forma, é possível implementar um dispositivo de codificação e decodificação de sinais, denominado de *codec*, acrônimo para *coder/decoder* (codificador/decodificador), que tira vantagem desta característica humana. Este processo de codificação de sinais baseado na limitação da percepção sonora do ser humano pode ser chamada de *codificação perceptiva*.

Quando ouvimos uma música reproduzida a partir de um CD, por exemplo, é impossível perceber toda a informação que chega aos nossos ouvidos. Desta forma, não há necessidade de se armazenar o sinal por completo, incluindo as partes inaudíveis. O processo que faz com que certas frequências menos significativas se tornem inaudíveis é denominado de *maskamento*. O efeito de *maskamento* descreve o comportamento do ouvido quando dois ou mais sons diferentes o estimulam simultaneamente, num curto intervalo de tempo e, isso consiste em apagar parcial ou totalmente algumas frequências do sinal de áudio,

devido à existência de uma componente tonal de frequência que sobrepõe as componentes menos significativas.

Experimentos mostram que o ouvido humano possui 24 bandas de frequência. Estas bandas de frequência são denominadas de *bandas críticas*. As frequências que se encontram dentro destas bandas críticas são mais difíceis de serem distinguidas entre si. Por exemplo, quando ouvimos uma música que possui uma componente de frequência tonal predominante, esta componente tenderá a produzir um mascaramento que será capaz de sobrepor as outras frequências que se encontram na mesma banda crítica.

### 3.2 MODULAÇÃO POR CODIFICAÇÃO DE PULSOS

A Modulação por codificação de pulsos, abreviada como PCM (sigla para *Pulse Code Modulation*), é basicamente a tecnologia mais antiga de digitalização de informação, incluindo a digitalização sonora [14]. Este tipo de digitalização sonora é um padrão para armazenamento e transmissão de sinais de áudio digital não comprimidos. CDs de conteúdo musical, por exemplo, costumam fazer o uso do PCM. O PCM fará uso de duas variáveis: a *taxa de amostragem* e o *bitrate*. A taxa de amostragem descreve quantas amostras por segundo a digitalização vai utilizar para medir a amplitude do sinal, ou seja, a quantidade de vezes que a amplitude do sinal é medida. O bitrate descreve o tamanho da amostragem digital, ou seja, a quantidade de bits em cada amostra.

Uma taxa de amostragem grande implica que mais frequências do sinal analisado serão incluídas na sua representação digital. Um bitrate maior acarreta numa representação mais fiel do valor do sinal nas amostras, uma vez que o número de bits é maior. Sendo assim, estes dois parâmetros são determinantes para a fidelidade de representação do sinal desejado. Uma maneira bastante simples e direta de compressão seria diminuir os valores da taxa de amostragem ou do bitrate, o que representaria uma diminuição imediata do volume dados utilizado. Porém, esta é uma alternativa que compromete bruscamente a fidelidade ao sinal original, o que para algumas aplicações se torna inviável.

## 4 CARACTERÍSTICAS DO PADRÃO MPEG-1 AUDIO LAYER 3

Com a elaboração do padrão MPEG-1, que começou no ano de 1988 e foi finalizado em 1992, foi estabelecida a divisão em partes deste padrão, que consistia em propor soluções de compressão tanto para sinais de vídeo como de áudio, ou até mesmo a combinação dos dois.

Para a parte de áudio foram estabelecidos três níveis de compressão, cada um com seu nível de complexidade. Estes níveis foram denominados de *Layers* (Camadas), os quais foram respectivamente numerados [12]. Existe então: MPEG-1 *Layer 1*, *Layer 2* e *Layer 3*. A *Layer 3* representa o padrão MP3 que hoje conhecemos e é, sem dúvidas, o mais eficiente dos três níveis citados em termos de compressão, porém é também o mais complexo.

Tomando como referência um sinal digitalizado em PCM, o MP3 é capaz de efetuar uma compressão altamente eficiente, sem praticamente nenhuma perda de qualidade perceptível. Para atingir tal nível de compressão, o MP3 faz uso da codificação perceptiva, filtrando frequências que o ouvido humano não é capaz de perceber. A codificação perceptiva é então um processo com perdas de informação, fazendo com que não seja possível recuperar a informação perdida após a compressão. Esta, porém, é uma consequência completamente aceitável, pois somente a parcela da informação que não é perceptível para nós é descartada.

Dois fatores determinantes para uma boa qualidade da codificação/compressão MP3 são o bitrate e a taxa de amostragem. O bitrate padrão normalmente utilizado no MP3 é o de 128kbps, mas este bitrate pode variar entre valores que vão de 8kbps a 320kbps. Um valor maior do bitrate acarreta na medição mais precisa das amostras de áudio, provendo assim uma melhor qualidade sonora. Porém, a qualidade sonora vai depender também principalmente da taxa de amostragem, pois enquanto que um maior bitrate proporciona mais precisão no valor das amostras, uma maior taxa de amostragem possibilita armazenar uma quantidade maior de valores. Isto faz com que haja um espectro de frequências mais amplo. O padrão MPEG-1 define 32 kHz, 44,1 kHz e 48 kHz como as taxas de amostragem para compressão de áudio.

É importante citar também quais são os modos de operação possíveis do MP3, que podem ser definidos da seguinte forma [16]:

- *Single Channel* ou *Mono* - neste modo, todo o áudio é codificado em um único canal;
- *Dual Channel* - aqui existem dois canais, codificados de forma totalmente independente entre si;
- *Stereo* - codificação em dois canais, porém correlacionados entre si;
- *Joint Stereo* - um tipo mais aprimorado de codificação Stereo, em que as redundâncias presentes entre os dois canais de áudio são detectadas e reduzidas para otimizar o processo de codificação.

## 5 CODIFICAÇÃO MP3

De uma forma simplificada, temos que o processo de codificação/compressão MP3 pode ser dividido em alguns passos. Primeiramente, com o sinal de áudio devidamente adquirido e digitalizado (normalmente em PCM), o primeiro passo consiste em passar o áudio digitalizado por um banco de filtros que dividirá o áudio em várias sub-bandas de frequência. Em paralelo a esta operação de filtragem, o áudio digitalizado de entrada passa pelo chamado *Modelo psicoacústico*, que utilizará o conceito de mascaramento para determinar quais frequências estarão ou não presentes em cada sub-banda estabelecida.

Com o áudio em formato de sequência de bits (também chamada de *bitstream*) dividido nas sub-bandas, é necessário que ele passe por um processo de alocação, quantização e codificação dos bits. Em seguida, é preciso formatar e organizar os bits alocados no passo anterior em uma sequência codificada de bits, resultando assim no arquivo MP3 comprimido.

O processo descrito pode ser ilustrado no diagrama de blocos simplificado da Figura 2.

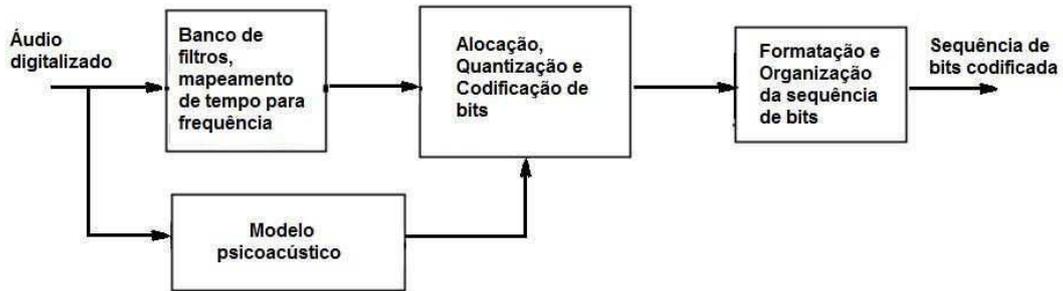


Figura 2 - Esquemático simplificado da codificação MP3 (adaptada de: <http://slideplayer.com/slide/7513652/>)

Como já foi mencionado anteriormente, por usar modelos psicoacústicos em sua implementação, o MP3 caracteriza-se como uma codificação perceptiva. Desta forma, o modelo psicoacústico vai tirar vantagem das limitações da nossa audição para efetuar o mascaramento de frequências que não são audíveis devido a presença de uma frequência tonal predominante. Assim, em cada sub-banda de frequências será possível mascarar as frequências menos significantes. Este processo pode ser visualizado na Figura 3.

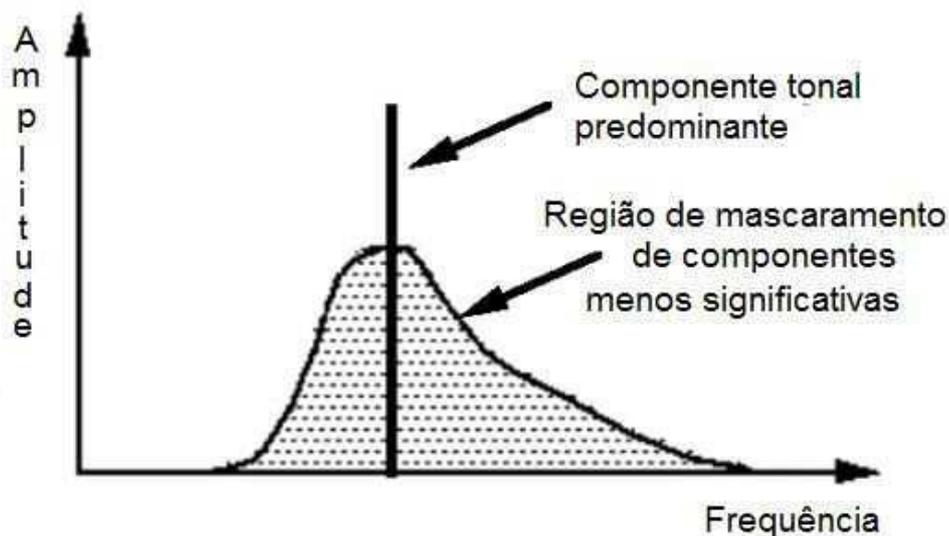


Figura 3 - Mascaramento de frequências (adaptado de: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/data-compression/lossy/mp3/psychoacoustics.htm>)

Se for feita uma análise um pouco mais detalhada do processo de codificação MP3, será necessário utilizar um diagrama de blocos um pouco mais sofisticado que o da Figura 2. Um diagrama mais completo pode ser visto na Figura 4.

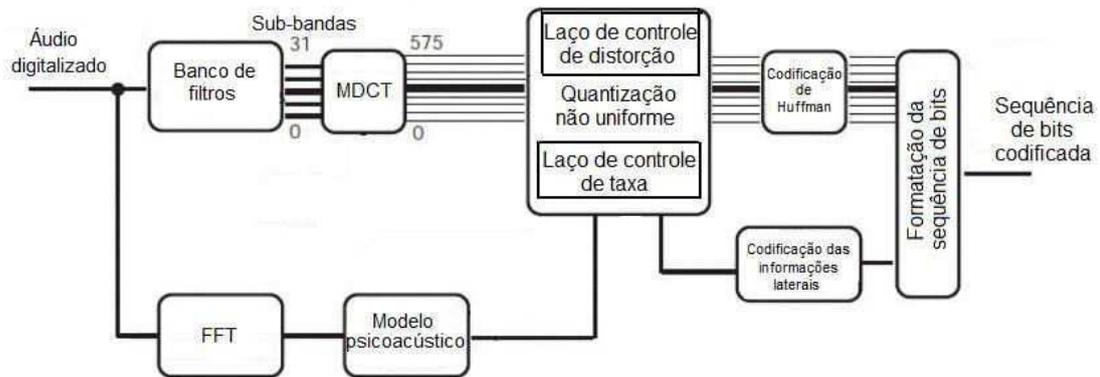


Figura 4 - Diagrama do algoritmo de codificação MP3 (adaptado de: [http://www.linkwitzlab.com/Brandenburg\\_mp3\\_aac.pdf](http://www.linkwitzlab.com/Brandenburg_mp3_aac.pdf))

Temos aqui, como já foi mostrado anteriormente de forma simplificada, o áudio digitalizado na entrada, que em seguida segue para o *Banco de filtros*. Ao banco de filtros chegam blocos de dados com 1152 amostras. Ao passar pelo banco de filtros, as amostras vão ser divididas em 32 sub-bandas que estão separadas em intervalos de frequências iguais. Esta técnica permite isolar diferentes componentes de frequência do sinal, o que é bastante útil para o processo de compressão.

Como o ouvido humano é mais sensível a algumas frequências do que outras, essas tem que ser exatamente preservadas na codificação, pois pequenas diferenças são significativas.

[17] Após passar pelo banco de filtros, as 32 sub-bandas seguem para o bloco *MDCT* (*Modified Discrete Cosine Transform*), que vai aplicar a Transformada discreta do cosseno modificada em cada uma das amostras das 32 sub-bandas.

Desta forma, há novamente uma subdivisão de cada sub-banda em 18 mais refinadas, originando agora 576 sub-bandas, que transportam os coeficientes de frequência gerados pelo bloco MDCT, o que permite uma eliminação de redundância mais eficiente.

Em paralelo a todo o processo de filtragem, temos a análise psicoacústica. O sinal de entrada passa pelo bloco *FFT (Fast Fourier Transform)*, que aplica a Transformada rápida de Fourier nas suas amostras, passando estas do domínio do tempo para o domínio da frequência, para que possam passar para o bloco *Modelo psicoacústico*.

Como o modelo psicoacústico recebe as amostras no domínio da frequência, é possível aplicar algoritmos que modelam a percepção sonora humana, determinando que partes do sinal são, ou não, audíveis. Desta forma, é oferecida ao bloco de quantização a informação de como ele deve quantizar os coeficientes de frequência que ele recebe do bloco *MDCT*. O modelo psicoacústico detecta as frequências dominantes calculando para cada banda crítica um *limite de mascaramento*. Estes limites são utilizados pelo bloco de quantização para que este mantenha o ruído de quantização abaixo destes limites.

No bloco *Quantização não uniforme*, que é o bloco de quantização, temos como entrada os coeficientes de frequência resultantes do bloco MDCT. Neste bloco, são feitas as operações de quantização e escalonamento dos valores vindos do bloco MDCT, de forma iterativa, em dois laços diferentes:

- *Laço de controle de taxa* - Este ciclo realiza a quantização das amostras no domínio da frequência e determina o tamanho do passo de quantização. As amostras são quantizadas com um tamanho de passo crescente, até que os valores quantizados possam ser codificados. Um passo maior leva a menores valores de quantização. É contabilizado o número de bits codificados e comparado com o número de bits disponíveis. Se esse valor é ultrapassado, o tamanho do passo da quantização é aumentado e repetido, até que o número de bits disponível seja suficiente;
- *Laço de controle de distorção* - Este ciclo controla o ruído de quantização causado pelo Laço de controle de taxa. O objetivo é manter o ruído de quantização abaixo do limite de mascaramento imposto pelo modelo psicoacústico.

O bloco *Codificação de Huffman* implementa a codificação dos valores quantizados que saem do bloco de quantização. A codificação de Huffman consiste basicamente em utilizar as probabilidades de ocorrência dos valores que se repetem em um conjunto de dados, de modo a representar estes valores com códigos menores, efetuando assim a compressão [18].

O bloco *Codificação das informações laterais* codifica valores que serão úteis para a decodificação e reprodução do sinal de áudio. E, finalmente, o bloco *Formatação da sequência de bits* é o responsável por agrupar tudo o que foi gerado anteriormente, formando as partes que compõe o arquivo MP3. No capítulo 7, será apresentada a estrutura completa de um arquivo MP3 gerado pelo processo de codificação.

## 6 DECODIFICAÇÃO MP3

O processo de decodificação de um arquivo MP3 consiste basicamente em efetuar a leitura, reconstruir e reproduzir o sinal de áudio codificado. Este processo é um pouco mais simples que o esquema de codificação e pode ser descrito pelo diagrama de blocos simplificado da Figura 5.

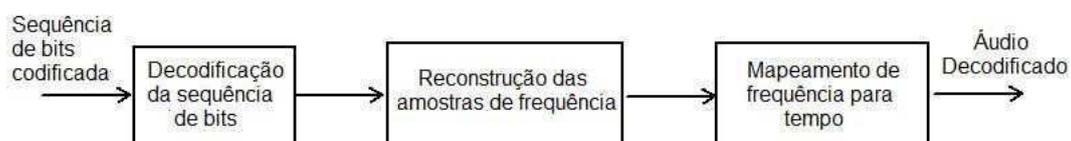


Figura 5 - Esquemático simplificado de um decodificador MP3 (adaptado de: <http://slideplayer.com/slide/7513652/>)

Pode-se observar que a decodificação é composta de três etapas principais: decodificação da sequência de bits de entrada (arquivo MP3), reconstrução das amostras de frequência e mapeamento de frequência para tempo.

A primeira parte, nomeada no diagrama de *Decodificação da sequência de bits*, sincroniza a sequência de bits codificada de entrada e extrai os coeficientes de frequência quantizados e qualquer outra informação sobre cada parte do arquivo MP3. Aqui também ocorre verificação de possíveis erros de codificação.

Logo em seguida temos o bloco *Reconstrução das amostras de frequência*. Este bloco reconstrói os dados a partir dos coeficientes de frequência gerados pelo bloco MDCT durante a codificação.

A última parte, referente ao bloco *Mapeamento de frequência para tempo*, abrange um conjunto de operações inversas da MDCT e a análise de um banco de filtros sintetizador no decodificador. O mapeamento frequência para tempo constrói o sinal de saída de áudio a partir dos coeficientes da MDCT. A MDCT Inversa (*IMDCT*) transforma os coeficientes em sinais de sub-banda no domínio do tempo. É aplicada então, a inversão na frequência, de modo a compensar a redução do número de amostras na análise do banco de filtros. Após esta etapa, a síntese do banco de filtros é aplicada aos sinais da sub-banda para formar o sinal de áudio de saída. Desta forma, o objetivo de reproduzir o arquivo codificado em MP3 é atingido.

Os decodificadores para MP3 estão especificados detalhadamente nos padrões da ISO. Estas especificações estão feitas de modo que não existam diferenças audíveis entre decodificadores que estejam em conformidade com a norma estabelecida.

## 7 ESTRUTURA DE UM ARQUIVO MP3

Todos os arquivos MP3 são divididos em fragmentos menores chamados de *frames* (quadros). Cada frame armazena 1152 amostras de áudio e tem a duração de 26 ms. Como o que determina o tamanho de cada amostra é o bitrate, ao aumentar o valor do bitrate, o tamanho do frame também aumenta. Porém, o tamanho do frame, que é dado em bytes, também é dependente do valor da taxa de amostragem. Esta dependência do tamanho do frame com estes valores pode ser descrita pela seguinte expressão:

$$\frac{144 \cdot \textit{bitrate}}{\textit{taxa\_amostragem}} + \textit{Padding\_byte}$$

Em que o *Padding\_byte* representa um byte especial, alocado no início de alguns frames, para satisfazer requisitos de bitrate. Por exemplo, ao codificar um arquivo com um bitrate de 128 kbps e uma taxa de amostragem de 44,1 kHz, obtemos, usando a expressão acima, um tamanho de frame de 417 bytes. Para que o segmento codificado cumpra exatamente o bitrate especificado, é necessário que alguns frames tenham 418 bytes, sendo neste caso utilizado o chamado *Padding\_byte*, que é um byte preenchido com zeros para simplesmente satisfazer o tamanho de frame requerido.

### 7.1 ESTRUTURA DE UM FRAME

Um frame de um arquivo MP3 consiste de cinco partes, nomeadas como: *Header* (Cabeçalho), *CRC* (*Cyclic Redundancy Check*), *Informações laterais*, *Dados principais* e *Dados auxiliares* [19]. Esta representação pode ser vista na Figura 6.



Figura 6 - Subdivisão de um frame de um arquivo MP3 (adaptado de: <http://mp3decoderfyp.blogspot.com>)

Segue abaixo a descrição, organizada em tópicos, das cinco partes de um frame:

- *Cabeçalho* - O cabeçalho basicamente tem a função de armazenar os bits de descrição do frame, assim como uma sequência de bits destinada a sincronização. Como o sinal digitalizado consiste basicamente em uma grande sequência de bits, a sincronização se faz importante para um receptor MP3 poder acessar o sinal digitalizado a qualquer ponto desta sequência de bits, bastando apenas ter que procurar primeiro pelos bits de sincronização. Este aspecto é bastante útil e possibilita a transmissão contínua de arquivos MP3. Um esquema resumido de um Cabeçalho pode ser visto na Figura 7.

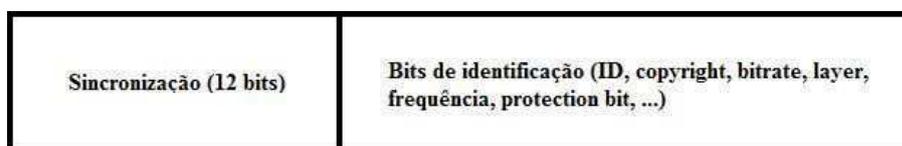


Figura 7 - Cabeçalho simplificado de um arquivo MP3 (adaptado de: [http://www.mp3-converter.com/mp3codec/mp3\\_anatomy.htm](http://www.mp3-converter.com/mp3codec/mp3_anatomy.htm))

Temos que a sequência de sincronização é composta de 12 bits, em que, para que a sincronização aconteça, estes bits precisam estar todos com o valor 1. Nos bits de identificação, temos enfatizados os mais relevantes que são:

- *ID*: Refere-se a um bit para identificar o padrão MPEG utilizado. Se o valor é 1, está sendo utilizado o MPEG-1;
- *Copyright*: Também possui apenas um bit que é utilizado para identificar se o arquivo possui direitos autorais. Se valor é 1, a cópia do conteúdo é ilegal;
- *Bitrate*: Identifica o bitrate utilizado com 4 bits, em que no padrão MP3 pode variar de 32 kbps a 320 kbps de acordo com a Tabela 1.

<i>bits</i>	<i>Bitrate(kbps)</i>
0000	-
0001	32
0010	40
0011	48
0100	56
0101	64
0110	80
0111	96
1000	112
1001	128
1010	160
1011	192
1100	224
1101	256
1110	320
1111	-

Tabela 1 - Identificação de Bitrate com 4 bits

- *Layer*: Identifica com 2 bits se o padrão é Layer 1, 2 ou 3 da seguinte forma: *01*: Layer 1; *10*: Layer 2; *11*: Layer 3. Obs.: o valor *00* não é utilizado;

- *Frequência*: Indica qual a frequência/taxa de amostragem utilizada com 2 bits. Temos assim: *00*: 44,1 kHz; *01*: 48 kHz; *10*: 32 kHz. Obs.: o valor *11* não é utilizado;

- *Protection bit*: Valor de apenas um bit que se estiver atribuído com 1, ativa o campo CRC do frame para uso.

- *CRC* - O campo CRC somente é utilizado se o valor do protection bit contido no Cabeçalho estiver igual a 1. Este campo serve para checar se erros de transmissão ocorreram e se valores contidos no Cabeçalho foram corrompidos. Se valores no Cabeçalho são corrompidos, todo o frame é comprometido, fazendo-se necessária a substituição do frame defeituoso pelo frame anterior;

- *Informações laterais* - O campo *Informações laterais* contém as informações necessárias para decodificar o que está contido no campo *Dados principais*. O tamanho em bytes deste campo vai depender da estrutura de canal que foi usada na codificação. Se a codificação foi de canal único (*Single channel*), *Informações laterais* terá 17 bytes. Para qualquer outro tipo de estrutura de canal, como *Stereo* ou *Dual Channel*, são utilizados 32 bytes. Dentre as principais informações neste campo, está a de identificar onde começa o campo *Dados principais*, pois somente após lido este valor, será possível efetuar a decodificação;
- *Dados principais* - Este campo, como o próprio nome sugere, consiste de parte dos dados principais que foram codificados e caracterizam a informação que será decodificada posteriormente. Sendo assim, *Dados principais* consiste dos bits que foram codificados usando codificação de Huffman e dos chamados *fatores de escala*. Os fatores de escala tem o propósito de reduzir o ruído causado pela quantização;
- *Dados auxiliares* - O campo *Dados auxiliares* é, na verdade, um campo opcional, ficando imediatamente após os bits de *Dados principais* codificados com codificação de Huffman.

## 7.2 ID3

Inicialmente, o padrão MP3 não seria capaz de armazenar nenhum tipo de informação em formato de texto para descrição do áudio codificado. Desta forma, foi proposta a introdução uma espécie de "etiqueta" no final dos arquivos. Esta etiqueta teria um comprimento de 128 bytes, e conteria campos de informação para o título do arquivo, artista, álbum, ano, comentário, número da faixa e gênero musical. Esta etiqueta recebeu o nome de ID3, e foi primeiramente implementada em sua versão ID3v1.1 [10].

O ID3v1.1 apresentava duas limitações significativas. O fato da informação ter de ser registrada em cada um dos campos de informação predefinidos e de estes campos estarem limitados a um número fixo de bytes, torna o ID3v1.1 muito pouco flexível. Adicionalmente, o fato da informação ID3

ser inserida no final de cada arquivo não é adequado para realizar a transmissão destes arquivos. Com o objetivo de resolver os problemas identificados no ID3v1.1, foi lançada a versão ID3v2. Nesta nova versão foram introduzidos novos campos de informação, sendo o seu tamanho dinâmico, o que permite uma introdução de informação mais completa e flexível. No ID3v2 a etiqueta é inserida no início de cada arquivo, o que facilita a operação de transmissão [10].

## 8 CONCLUSÕES

Neste trabalho, foi efetuado o estudo do padrão MPEG-1 Layer 3, mais conhecido como MP3. Foi possível constatar que se trata de um método de compressão bastante complexo e que exige uma série de procedimentos para sua total compreensão. Estes procedimentos envolvem: entender os conceitos básicos de compressão de dados, entender de que forma ocorrem as interações entre o nosso cérebro e o nosso sistema auditivo (psicoacústica), obter a compreensão das características fundamentais do padrão MP3, assim como compreender todos os processos de codificação e decodificação envolvidos neste padrão e, finalmente, entender toda a estrutura que compõe um arquivo MP3.

Desta forma, o MP3 consiste de um nível de compressão de dados bastante eficiente, provando que é capaz de reduzir consideravelmente o volume de dados exigido por sinais de áudio, conseguindo manter um nível de qualidade bastante significativo, de forma que a informação essencial do sinal original é perfeitamente assimilável após a compressão. De fato, é um padrão que faz jus a toda a popularidade que conseguiu durante todos os anos desde sua elaboração.

## 9 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ZÖLZER, U. *"Digital Audio Signal Processing"*, Second Edition, Helmut Schmidt University, Hamburg, Germany, John Wiley & Sons Ltd, pp. 273-311, 2008.
- [2] WANG, Y. *"Audio Coding"*, Polytechnic University, Brooklyn. Disponível em: <[http://eeweb.poly.edu/~yao/EE3414/audio\\_coding.pdf](http://eeweb.poly.edu/~yao/EE3414/audio_coding.pdf)>.
- [3] LIU, C., CHANG, W. *"Audio Coding Standards"*, National Chiao Tung University, Taiwan. Disponível em: < <http://www.mp3-tech.org/programmer/docs/AudioCoding.pdf>>.
- [4] SIELER, M., SPERSCHNEIDER, R., *" MPEG-Layer3 Bitstream Syntax and Decoding"*
- [5] *"Modelo psicoacústico"*, Disponível em: <[http://www.img.lx.it.pt/~fp/cav/ano2011\\_2012/Trabalhos\\_MERC\\_2012/Artigo10/MP3vsAAC\\_G10\\_JLucas\\_70685/modelo.html](http://www.img.lx.it.pt/~fp/cav/ano2011_2012/Trabalhos_MERC_2012/Artigo10/MP3vsAAC_G10_JLucas_70685/modelo.html)> .
- [6] *"Psicoacústica"*, Disponível em: <<http://www.drumchannel.com.br/psicoacustica/>>.
- [7] *"Psychoacoustics, defined"*, Disponível em: < <http://thepowerofsound.net/psychoacoustics-defined/>>.
- [8] *"How MP3 Files Work"*, Disponível em: <<http://computer.howstuffworks.com/mp3.htm>>.
- [9] *"MP3 Encoding and File Format"*, Disponível em: <<http://www.herongyang.com/CD-DVD/MP3-Audio-Encoding-and-File-Format.html>>.
- [10] ID3.org - The Tag Standard, Disponível em: <<http://www.id3.org>>.
- [11] MP3Tech, Disponível em: <<http://www.mp3-tech.org/>>.
- [12] International Organization for Standardization webpage, Disponível em: < <http://www.iso.org/>>.
- [13] *"Noise terms glossary"*, disponível em < <http://memtechacoustical.com/noise-terms-glossary/>>.

[14] "*Diferenças entre formatos de áudio*", disponível em: <  
<http://www.tecmundo.com.br/audio/7945-saiba-quais-sao-as-principais-diferencas-entre-formatos-de-audio.htm>>.

[15] "*MP3 vs AAC*", disponível em: <  
[http://www.img.lx.it.pt/~fp/cav/ano2008\\_2009/Trabalhos\\_MEEC\\_2009/Artigo\\_MEEC\\_16/Mp3Vs.AAC\\_website/Mp3Vs.AAC\\_website/MP3vsAAC/MP3\\_aplicacoes.htm](http://www.img.lx.it.pt/~fp/cav/ano2008_2009/Trabalhos_MEEC_2009/Artigo_MEEC_16/Mp3Vs.AAC_website/Mp3Vs.AAC_website/MP3vsAAC/MP3_aplicacoes.htm)>.

[16] "*MP3 Modos de operação*", disponível em: <  
[http://www.img.lx.it.pt/~fp/cav/ano2011\\_2012/Trabalhos\\_MERC\\_2012/Artigo10/MP3vsAAC\\_G10\\_JLucas\\_70685/modos.html](http://www.img.lx.it.pt/~fp/cav/ano2011_2012/Trabalhos_MERC_2012/Artigo10/MP3vsAAC_G10_JLucas_70685/modos.html)>.

[17] "*MP3 and AAC explained*", disponível em: <  
[http://www.linkwitzlab.com/Brandenburg\\_mp3\\_aac.pdf](http://www.linkwitzlab.com/Brandenburg_mp3_aac.pdf)>.

[18] "*A Codificação de Huffman*", disponível em: <  
<http://www.inf.ufes.br/~pdcosta/ensino/2009-1-estruturas-de-dados/material/CodificacaoHuffman.pdf>>.

[19] "*The MP3 Frame Layout*", disponível em: <  
<http://mp3decoderfyp.blogspot.com>>.