

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS DE CURITIBA
ENGENHARIA ELÉTRICA**

**ANDRÉ YUJI TANAKA
ELTON RIBEIRO BARBOSA
RYAN SEIYU YAMAGUCHI KIMURA**

**Análise de Qualidade de Áudio Objetiva e Subjetiva em vários
formatos digitais**

TRABALHO DE CONCLUSÃO DE CURSO

**CURITIBA
2017**

**ANDRÉ YUJI TANAKA
ELTON RIBEIRO BARBOSA
RYAN SEIYU YAMAGUCHI KIMURA**

**Análise de Qualidade de Áudio Objetiva e Subjetiva em vários
formatos digitais**

Trabalho de Conclusão de Curso apresentado
ao Departamento Acadêmico de Eletrotécnica -
DAELT como requisito parcial para obtenção do
título de Engenheiro Elétrico no curso superior
de Engenharia Elétrica da Universidade
Tecnológica Federal do Paraná (UTFPR) como
requisito parcial na obtenção do título de
Engenheiro Eletricista.

Orientador: Professor Dr. Antônio Carlos Pinho,
Dr. Eng.

CURITIBA

2017

**André Yuji Tanaka
Elton Ribeiro Barbosa
Ryan Seiyu Yamaguchi Kimura**

Análise de Qualidade de Áudio Objetiva e Subjetiva em vários formatos digitais

Este Trabalho de Conclusão de Curso de Graduação foi julgado e aprovado como requisito parcial para a obtenção do Título de Engenheiro Eletricista, do curso de Engenharia Elétrica do Departamento Acadêmico de Eletrotécnica (DAELT) da Universidade Tecnológica Federal do Paraná (UTFPR).

Curitiba, 20 de novembro de 2017.

**Prof. Emerson Rigoni, Dr.
Coordenador de Curso
Engenharia Elétrica**

**Profa. Annemarlen Gehrke Castagna, Mestre
Responsável pelos Trabalhos de Conclusão de Curso
de Engenharia Elétrica do DAELT**

ORIENTAÇÃO

**Antonio Carlos Pinho, Dr.
Universidade Tecnológica Federal do Paraná
Orientador**

BANCA EXAMINADORA

**Marcelo de Oliveira Rosa, Dr.
Universidade Tecnológica Federal do Paraná**

**Daniel Flores Cortez, Mestre.
Universidade Tecnológica Federal do Paraná**

**Ednilson Soares Maciel, Mestre
Universidade Tecnológica Federal do Paraná**

A folha de aprovação assinada encontra-se na Coordenação do Curso de Engenharia Elétrica

RESUMO

TANAKA, André Yuji. BARBOSA, Elton Ribeiro. KIMURA, Ryan Seiyu Yamaguchi. Análise de Qualidade de Áudio Objetiva e Subjetiva em vários formatos digitais, 2017. 143 f. Trabalho de conclusão de curso - Curso de graduação em Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, 2017.

Este trabalho apresenta uma análise de qualidade de áudio em vários formatos digitais, de forma objetiva e subjetiva. A análise objetiva mostra as diferenças entre áudios de formatos digitais como WAV, MP3 e OGG, usando um software específico como base desta análise. A análise subjetiva mostra as diferenças encontradas em audições em um sistema de reprodução de alta fidelidade e com grupos distintos de pessoas, tanto músicos que atuam na área quanto pessoas que não tem relação direta com a música, também com os formatos WAV, MP3 e OGG. O trabalho aborda as teorias básicas dos fenômenos da audição e do som, teorias a respeito do áudio digital e análogo, faz um histórico dos processos de conversão análogo/digital e os fenômenos associados, faz também históricos de cada formato digital analisado. Por fim, faz conclusões com os resultados obtidos nas análises a respeito da qualidade dos formatos digitais de áudio.

Palavras Chaves: Som, Áudio Digital, Análise de Qualidade de Áudio, Formatos Digitais de Áudio.

ABSTRACT

TANAKA, André Yuji. BARBOSA, Elton Ribeiro. KIMURA, Ryan Seiyu Yamaguchi. Subjective and Objective Audio Analyse with Different Digital Audio Formats, 2017. 143 f. Trabalho de conclusão de curso - Curso de graduação em Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, 2017.

This paper presents a quality analysis of audio in different digital formats, in objective and subjective way. The objective analysis shows the differences between digital audios formats like WAV, MP3 and OGG, based on a specific software. The subjective analysis shows the differences found in auditions in a High-Fidelity reproduction system, with different groups of people, musicians and people who doesn't relate with music area directly, also with WAV, MP3 and OGG formats. The paper provides basic theories about the audio and sound phenomena, theories about digital and analog audio, shows a history of analog/digital conversion process and the associated phenomena, also shows the history of each digital format analyzed. It presents conclusions with the results obtained with the analysis about digital audio form quality.

Keywords: Sound, Digital Audio, Quality Audio Analysis, Digital Audio Formats.

LISTA DE FIGURAS

Figura 1 - Mecanismo usado para gravação de áudio	16
Figura 2 - Foto tirada do gramofone nos laboratórios de Thomas Edson	17
Figura 3 - Foto de um Tríodo Audion	18
Figura 4 - Foto do microfone condensador criado por Edward C. Wente.....	18
Figura 5 - Foto do Magnetofone da AEG	20
Figura 6 - Exemplo de conversão do Áudio Analógico (linha curvada) para Áudio Digital (linha tracejada).....	21
Figura 7 - Exemplo de um sistema high-end em sala isolada, com amplificadores e caixas de som com estereofonia	23
Figura 8 - Números binários podem ser representados por PCM por uma sequência de tensões altas (high) e baixas (low).....	28
Figura 9 - Representação gráfica do sinal em um sistema PCM completo de digitalização.....	30
Figura 10 - Onda senoidal de 20 kHz amostrada à 40 kHz produzindo uma onda quadrada. O filtro passa baixa remove as harmônicas acima de 20 kHz.....	33
Figura 11 - Erro de quantização limitado à uma metade do bit menos significante	35
Figura 12 - O erro de quantização é relativamente alto em um sinal que passa por apenas alguns passos de quantização	38
Figura 13 - <i>Dither</i> pode suavizar os efeitos do erro de quantização	39
Figura 14 - Processo da conversão D/A com valores positivos.....	40
Figura 15 - Anatomia do ouvido	42
Figura 16 - Ouvido interno.....	43
Figura 17 - Mascaramento auditivo	45
Figura 18 - Os dois tipos de vibração de ondas	46
Figura 19 - Propagação da onda sonora.....	47
Figura 20 - Diminuição da pressão sonora.....	48
Figura 21 - Função senoidal.....	49
Figura 22 - Função senoidal amortecida	49
Figura 23 - Função senoidal mais específica	50

Figura 24 - Nota dó	51
Figura 25 - Oitava de dó.....	52
Figura 26 - Série harmônica.....	52
Figura 27 - Diversas ondas e respectivas frequências: (A) apenas frequência fundamental, (B) fundamental e primeiro harmônico e (C) diversos harmônicos.	
.....	53
Figura 28 - Representação do ataque, período estável e queda de um som...	54
Figura 29 - Exemplo de uma escala em dB	56
Figura 30 - Princípio de funcionamento do microfone de bobina móvel.....	59
Figura 31 - Funcionamento do microfone de fita	60
Figura 32 - Seção transversal do alto-falante com bobina móvel.....	62
Figura 33 - Diagrama de blocos de um codificador MP3Fonte: C.-H. Yen et al (2004).....	67
Figura 34 - Diagrama de blocos de um codificador AAC Fonte: BRANDENBURG (1999).....	70
Figura 35 - Disposição da sala	77
Figura 36 - Amplitude do trecho de violão no formato WAV.....	85
Figura 37 - Amplitude do trecho de violão no formato MP3 - 128 kbps.....	86
Figura 38 - Amplitude do trecho de violão no formato em MP3 - 320 kbps.....	86
Figura 39 - Amplitude do trecho de violão no formato em OGG.....	87
Figura 40 - Ponto exato do primeiro pico negativo do formato MP3 - 128 kbps do trecho de violão.....	88
Figura 41 - Ponto exato do primeiro pico negativo do formato MP3 - 320 kbps do trecho de violão.....	88
Figura 42 - Trecho com violão de todos formatos de áudio.	89
Figura 43 - Gráfico amplificado de todas amplitudes.	90
Figura 44 - Diferenças de amplitudes do trecho de violão em WAV em relação aos formatos MP3 - 128 kbps, MP3 - 320 kbps e OGG.	91
Figura 45 - Espectro de todos formatos de áudio para o trecho de violão.	92
Figura 46 - Diferença dos espectros, do trecho de violão.	92
Figura 47 - Análise na amplitude com todos formatos analisados para o trecho de vocal da música Cirandeira.	93
Figura 48 - Análise na amplitude para o formato MP3 em 128 kbps no trecho de vocal da música Cirandeira.	94

Figura 49 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de vocal da música Cirandeira.....	95
Figura 50 - Análise de todos espectros em todos formatos analisadas para o trecho de vocal da música Cirandeira.....	96
Figura 51 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de vocal da música Cirandeira.....	97
Figura 52 - Análise na amplitude com todos formatos analisados para o trecho de flauta da música Cirandeira.....	98
Figura 53 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de flauta da música Cirandeira.....	99
Figura 54 - Análise de todos espectros em todos formatos analisadas para o trecho de flauta da música Cirandeira.....	100
Figura 55 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de flauta da música Cirandeira.....	100
Figura 56 - Análise na amplitude com todos formatos analisados para o trecho de violino da música Cirandeira.....	101
Figura 57 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de flauta da música Cirandeira.....	102
Figura 58 - Análise de todos espectros em todos formatos analisadas para o trecho de violino da música Cirandeira.....	103
Figura 59 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de flauta da música Cirandeira.....	103
Figura 60 - Análise na amplitude com todos formatos analisados para o trecho de percussão da música Cirandeira.....	104
Figura 61 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de percussão da música Cirandeira.....	105
Figura 62 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de percussão da música Cirandeira.....	105
Figura 63 - Análise de todos espectros em todos formatos analisadas para o trecho de percussão da música Cirandeira.....	106
Figura 64 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de repercussão da música Cirandeira.....	107
Figura 65 - Espectrograma do formato WAV.....	107
Figura 66 - Espectrograma do formato WAV, com as três dimensões.....	108

Figura 67 - Espectrograma do formato MP3 - 128 kbps.....	108
Figura 68 - Espectrograma do formato MP3 - 128 kbps, com as três dimensões.	109
Figura 69 - Espectrograma do formato MP3 - 320 kbps.....	109
Figura 70 - Espectrograma do formato MP3 - 320 kbps, com as três dimensões.	110
Figura 71 - Espectrograma do formato OGG.	110
Figura 72 - Espectrograma do formato OGG, com as três dimensões.....	111
Figura 73 - Espectrograma da diferença entre WAV e MP3 - 128 kbps.....	111
Figura 74 - Espectrograma da diferença entre WAV e MP3 - 128 kbps com as três dimensões.....	112
Figura 75 - Espectrograma da diferença entre WAV e MP3 - 320 kbps.....	112
Figura 76 - Espectrograma da diferença entre WAV e MP3 - 320 kbps com as três dimensões.....	113
Figura 77 - Espectrograma da diferença entre WAV e OGG.	113
Figura 78 - Espectrograma da diferença entre WAV e OGG com as três dimensões.....	114

LISTA DE TABELAS

Quadro 1- Variação da velocidade do som em relação a Temperatura	48
Tabela 2 - Notas para qualidade, grupo não músicos.	79
Tabela 3 - Notas para qualidade, grupo músicos.	79
Tabela 4 - Postos das qualidades, grupo não músicos	80
Tabela 5 - Distribuição Qui-quadrado.....	81
Tabela 6 - Postos das qualidades, grupo músicos	82

LISTA DE SIGLAS

PCM	<i>Pulse Code Modulation</i>
AEG	<i>Allgemeine Elektricitäts-Gesellschaft</i>
RRG	<i>Reichs-Rundfunk-Gesellschaft</i>
AC	<i>Alternate Current</i>
CD	<i>Compact Disc</i>
VHS	<i>Video Home System</i>
MD	<i>Mini Disc</i>
IT&T	<i>International Telephone & Telegraph</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
WAV	<i>WAVEform audio format</i>
AIFF	<i>Audio Interchange File Format</i>
MP3	<i>MPEG 1 Audio Layer 3</i>
AAC	<i>Advanced Audio Coding</i>
WMA	<i>Windows Media Audio</i>
LP	<i>Long Play</i>
LSB	<i>Least Significant Bit</i>
HD	<i>Hard Drive</i>
D/A	<i>Digital/Analógico</i>
IEC	<i>International Electrotechnical Commission</i>
DVD	<i>Digital Video Disc</i>
IBM	<i>International Business Machines</i>
DSD	<i>Direct Digital Stream</i>
DAB	<i>Digital Audio Broadcasting</i>
OCF	<i>Optimum Coding in the Frequency Domain</i>
MPEG	<i>Moving Picture Expert Group</i>
MUSIC	<i>Masking-pattern Adapted Universal Subband Integrated</i>
AM	<i>Coding and Multiplexing</i>
ASPEC	<i>Adaptive Spectral Perceptual Entropy Coding</i>
CBR	<i>Constant Bit Rate</i>
VBR	<i>Variable Bit Rate</i>
AM	<i>Amplitude Modulation</i>
FM	<i>Frequency Modulation</i>
TNS	<i>Temporal Noise Shaping</i>
MDCT	Transformada de Cosseno Discreta Modificada
dB	Decibéis
kbits/s	Kilobits por segundo
kHz	Kilohertz
Hz	Hertz
V	Volts

SUMÁRIO

1. INTRODUÇÃO	15
1.1. TEMA	15
1.1.1. DELIMITAÇÃO DO TEMA	23
1.3. OBJETIVOS	25
1.3.1. OBJETIVO GERAL	25
1.3.2. OBJETIVOS ESPECÍFICOS	25
1.4. JUSTIFICATIVA	26
2. FUNDAMENTAÇÃO TEÓRICA	27
2.1. ÁUDIO DIGITAL/ANALÓGICO	27
2.1.1. Amostragem	31
2.1.2. Quantização	34
2.1.3. Dither	37
2.1.4. Armazenamento de áudio digital	39
2.1.4. Conversão analógica/digital	40
2.2.1. Fisiologia da audição	41
2.2.1.1. Anatomia do Ouvido	41
2.2.1.2. Mascaramento	44
2.2.2. Som	45
2.2.4. Frequências das vibrações sonoras	50
2.3. SISTEMAS DE GRAVAÇÃO E REPRODUÇÃO	58
2.3.1. MICROFONE	58
2.3.1.1. Microfones de bobina móvel	58
2.3.1.2. Microfones de fita	59
2.3.2. ALTO-FALANTES	60
2.3.2.1 - Alto-falante de bobina móvel	61

2.3.3. Resposta de Frequência	62
2.4. FORMATOS DE ÁUDIO DIGITAL	63
2.4.3. Red Book – Formato padrão dos CDs de áudio	63
2.4.4. Formato de arquivo WAV	64
2.4.5. Super Audio CD	65
2.5. COMPRESSÃO DIGITAL	65
2.5.1 Formato de arquivo MP3	65
2.5.2 Formato de Arquivo AAC	70
2.5.3 Formato OGG Vorbis	72
3. ANÁLISE SUBJETIVA	73
3.1 Metodologia	73
3.1.1 Desenvolvimento do questionário	73
3.1.2 Grupos de amostras	74
3.1.3 Faixas de reprodução	74
3.2 Aplicação e Descrição da Prática	75
3.2.1 Equipamentos	75
3.2.2 Aplicação do Teste	76
3.2.2.1 Grupo não músicos	77
3.2.2.2 Grupo de músicos	78
3.3 Resultados	78
3.3.1 Análise dos resultados do grupo de Não Músicos	79
3.3.1 Análise dos resultados do grupo de Músicos	81
4. ANÁLISE OBJETIVA	84
4.1. METODOLOGIA	84
4.1.2 ANÁLISE DA AMPLITUDE	84
4.1.3 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE	90
4.1.4 ANÁLISE DOS ESPECTROS	91

4.1.5 ANÁLISE DAS DIFERENÇAS DE ESPECTROS	92
4.2 ANÁLISE DOS TRECHOS ESPECÍFICOS	93
4.2.1 ANÁLISE DO TRECHO DE VIOLÃO	93
4.2.2 ANÁLISE DO TRECHO DE VOCAL	93
4.2.2.1 ANÁLISE DA AMPLITUDE	93
4.2.2.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE	95
4.2.2.3 ANÁLISE DOS ESPECTROS	96
4.2.2.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS	96
4.2.3 ANÁLISE DO TRECHO DE FLAUTA	97
4.2.3.1 ANÁLISE DA AMPLITUDE	97
4.2.3.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE	98
4.2.3.3 ANÁLISE DOS ESPECTROS	99
4.2.3.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS	100
4.2.4 ANÁLISE DO TRECHO DE VIOLINO	101
4.2.4.1 ANÁLISE DA AMPLITUDE	101
4.2.4.2 - ANÁLISE DAS DIFERENÇAS DE AMPLITUDE	101
4.2.4.3 ANÁLISE DOS ESPECTROS	102
4.2.4.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS	103
4.2.5 ANÁLISE DO TRECHO DE PERCUSSÃO	104
4.2.5.1 ANÁLISE DE AMPLITUDE	104
4.2.5.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDES	104
4.2.5.3 ANÁLISE DOS ESPECTROS	106
4.2.5.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS	106
4.6 ANÁLISE DO ESPECTROGRAMA	107
5. CONCLUSÃO	115
REFERÊNCIAS	118
ANEXO A1 – Formulários do primeiro dia de pesquisa	122

ANEXO A2 – Formulários do Segundo dia de pesquisa	132
ANEXO A3 – CÓDIGO 1	138
ANEXO A4 – CÓDIGO 2	141

1. INTRODUÇÃO

1.1. TEMA

A percepção do som (audição) foi uma das principais características evolutivas adquiridas pelo ser humano; ela é fundamental para sua sobrevivência, porque podemos escutar ou perceber sons naturais do ambiente, nos ajudando a encontrar comida ou de se tornar em alimento para outro predador, além também de influenciar em nosso senso de balanço e também na nossa comunicação com outros seres (Steven Errede, 2002). “Pela definição da Física o som é uma vibração que se propaga naturalmente em uma onda mecânica audível de pressão e deslocamento, através de um meio como ar ou água” (Houghton Mifflin, 2002). Enquanto que em Fisiologia ou Psicologia, o som é a recepção destas ondas através da percepção do cérebro (Ballou G, 2008).

Na Era da Informação atual, caracterizada pela mudança da indústria tradicional para uma economia baseada em computação da informação (Castells, 1999), o som faz parte também de culturas e tecnologias criadas pelo homem, como música e telecomunicações por exemplo, na forma de áudio digital (majoritariamente devido aos computadores), ou analógico (da gravação de conversão do sinal analógico para digital). Sendo áudio definido como o processo de recepção, reprodução e de transmissão de som (Douglas Harper, 2010). Será apresentada uma revisão da história do áudio para contextualização e também para mostrar suas implicâncias nos meios de comunicações, conflitos (guerra) e também na arte.

As tecnologias atuais de telecomunicações têm como sua base o telégrafo inventado por Samuel F. B. Morse, que conseguiu enviar um sinal elétrico por um fio, de Washington para Baltimore em 1844. Em seguida Alexander Graham Bell em 1876, na Exposição Mundial que ocorreu em Filadélfia, inventou o telefone, que transformava ondas sonoras mecânicas para corrente elétrica e então de volta (Steven Schoenherr, 2002).

Os primeiros vestígios sobre gravação ou reprodução de som começaram com Thomas Edison e Emile Berliner, criadores do fonógrafo e do gramofone

respectivamente. O objetivo do fonógrafo descrito na patente “US 200521 A”, era gravar em características permanentes, a voz humana e outros sons, sendo que características como o som serão reproduzidas e consideradas audíveis novamente em um tempo futuro (Thomas Edson, 1877). Estes aparelhos utilizavam-se do mesmo mecanismo para gravação de som, era usado um cone (como os de megafones antigos) para que o som emitido fosse concentrado em um diafragma estendido (membrana sensível à vibração sonora), que era responsável pela captação do som emitido por alguma fonte, e. Finalmente era conectado junto com o diagrama, uma agulha (*stylus*) que fazia o entalhamento do áudio analógico em uma folha delgada de estanho (enrolada em algum cilindro) no caso do fonógrafo, ou em um disco de cera no caso do gramofone. Porém o som tinha um intervalo de frequência limitado e também era distorcido. Este mecanismo, de acordo com seu funcionamento, era totalmente mecânico ou também referenciado como acústico, apenas a partir de 1920 equipamentos elétricos começaram a ser usados para gravação ou reprodução do som (Francis Rumsey, 1992).

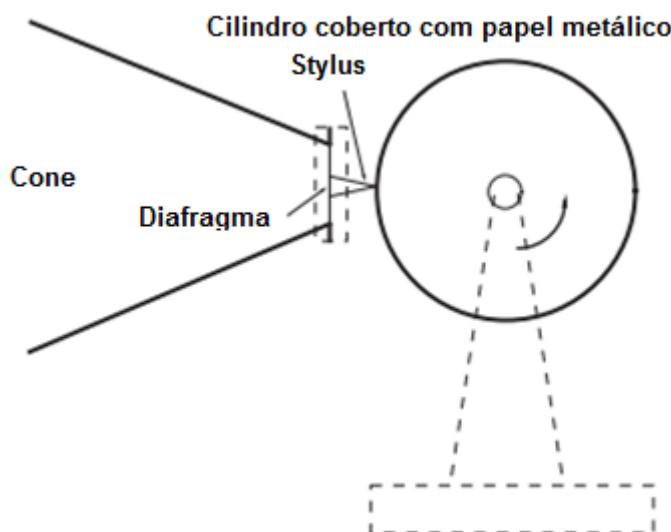


Figura 1 - Mecanismo usado para gravação de áudio
Fonte: Sound and Recording (Francis Rumsey, 1992).



Figura 2 - Foto tirada do gramofone nos laboratórios de Thomas Edison
Fonte: O autor, 2016.

Em torno de 1920, as principais gravadoras dos EUA nesta época (Victor e Columbia) começaram a adotar gradativamente microfones condensadores, amplificadores, gravadores e reprodutores eletromagnéticos, sendo que estas novas tecnologias foram desenvolvidas nos laboratórios da Bell Telephone Company (pertencentes à AT&T na época) e seu antecessor, a Western Electric Engineering Department. O objetivo para a invenção destas tecnologias era o aperfeiçoamento da rede de telefonia americana, com pesquisa e desenvolvimento para melhores telefones e transmissão de informações pelo telefone. Desse jeito, os engenheiros da Western Electric queriam solucionar o problema de longa distância para comunicações em telefone; portanto era necessário um dispositivo para amplificação do sinal de corrente, que era muito fraco com a tecnologia da época (T. Shaw, 1944). Lee de Forest inventou o Audion em 1906, que é um amplificador eletrônico de tubo em vácuo com 3 eletrodos (triodo). Porém ele ainda não era prático o suficiente para amplificar o sinal na magnitude desejada. Harold D. Arnold se interessou pelo Audion e demonstrou que poderia fazê-lo um dispositivo prático de amplificador eletrônico de tensão. Com seus conhecimentos em física relacionada à elétrons ele

aumentou o nível de vácuo no tubo e desse jeito eles foram testados com sucesso nas linhas telefônicas de Nova Iorque para São Francisco, possibilitando a comercialização de linhas telefônica à longa distância e também o desenvolvimento do alto-falante (Sheldon Hochheiser, 1992).

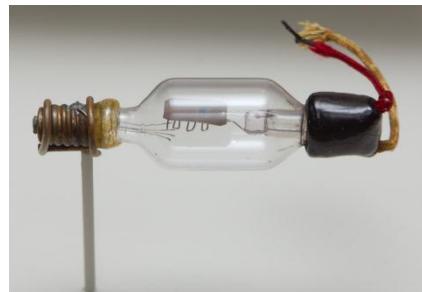


Figura 3 - Foto de um Tríodo Audion

Fonte: Gregory F. Maxwell em *The History of Audio: The Engineering of Sound*.

Ainda assim, no Departamento de Engenharia da Western Electric eram feitos estudos científicos sobre o som com a crença de que o conhecimento deste poderia melhorar a qualidade em transmissões telefônicas. Em 1914, Edward C. Wente se juntou ao departamento para pesquisa relacionada ao desenvolvimento e calibragem de um transmissor uniformemente sensível para uso em estudos do som, sendo que os transmissores usados nos receptores dos telefones tinham uma resposta à frequência muito desigual, e também muito ruído de fundo. Em 1922, ele produziu um transmissor condensador com 100 vezes mais sensitividade que os transmissores convencionais com a utilização dos amplificadores aperfeiçoados de Harold D. Arnold, criando assim o microfone condensador.



Figura 4 - Foto do microfone condensador criado por Edward C. Wente.

Fonte: Timetoast.

Neste mesmo ano, J. Maxfield foi o responsável, com a ajuda de uma equipe, pelo desenvolvimento de uma versão eletromagnética do fonógrafo. No

meio do ano de 1922, sua equipe teve êxito no desenvolvimento de um sistema protótipo experimental, com a utilização do microfone condensador de Edward C. Wente e com amplificadores baseados no projeto de Harold D. Arnold. Neste sistema o microfone condensador convertia a energia sonora do músico para energia elétrica com a utilização de transdutores, que então era amplificado antes de ser convertido para energia mecânica para a agulha do gravador (Sheldon Hochheiser, 1992). A gravação e reprodução de som desta época, foi marcada pelos desenvolvimentos da utilização de estudos eletrônicos e eletromagnéticos, em contraste dos aparelhos utilizados anteriormente que eram totalmente mecânicos, possibilitando em gravações e reproduções com intervalo de frequência maior e também de faixa dinâmica maior.

Após a era de gravações e reproduções elétricas, as tecnologias com fitas magnéticas foram a nova frente de desenvolvimento para o áudio. Suas origens são datadas entre as fases analógica e elétrica do áudio, sendo que primeiramente, Oberlin Smith em 1888 teve a ideia de gravar sons utilizando algodão ou fios de seda, com poeira de aço suspensa nos materiais. Assim as partículas eram magnetizadas de acordo com a corrente ondulatória entregue pelo microfone. O primeiro dispositivo prático com a ideia de gravação com teorias eletromagnéticas, foi feita pelo dinamarquês Valdemar Poulsen, que inventou o telephone que utilizava fios de aço magnetizados, porém a reprodução era ainda muito fraca e, portanto, outras formas de reprodução eram popularmente utilizadas (Semi J. Begun, 1949).

Foi aproximadamente no final da Segunda Guerra Mundial que os processos magnéticos começaram a mostrar suas vantagens em relação às tecnologias popularmente utilizadas naquela época, inicialmente com o Magnetofone que era o modelo pioneiro de gravador de fitas *reel-to-reel* (carrete para carrete), pela companhia de eletrônicos AEG em 1930 (Friedrich Engel, 2006; Peter Hammar, 2006). Porém durante seus primeiros modelos, ele era considerado um produto de qualidade baixa, com distorções e ruídos quanto às reproduções das gravações. Foi com pesquisas e experimentos feitos por Hans Joachim von Braunmühl and Walter Weber em 1941, que ocorreram inovações significativas para as gravações com fitas magnéticas, ambos trabalhavam para a RRG (Reichs-Rundfunk-Gesellschaft) que era a principal transmissora de rádio da Alemanha. Assim descobriram a técnica de *bias* de alta-frequência (AC *bias*),

que era a adição de alto nível (10 vezes do nível de áudio máximo) de tom com alta frequência inaudível, superando a histerese inerente dos materiais magnéticos contidos na fita. Desse jeito o Magnetofone aplicado com a técnica de AC *bias*, era utilizado nos discursos de Adolf Hitler, em reproduções ao vivo para uma cidade, enquanto que na realidade ele estava seguro em outra. Com o final da guerra esta tecnologia ficou exposta ao Mundo, e então os métodos de gravação de fitas magnéticas começaram a ser adotado popularmente pelas principais emissoras de transmissão (Steven Schoenherr, 2002).



Figura 5 - Foto do Magnetofone da AEG
Fonte: Site de "Museum of Magnetic Sound Recording".

As inovações de gravação magnética, foram adotadas no campo da computação, em discos rígidos, CD's e disquetes para armazenamento de dados e também para gravação de áudio com fitas cassetes, VHS e MD's, progredindo então para a nova era de som digital. O áudio digital tem sua história embasada no descobrimento da Modulação por Código de Pulso, ou Pulse Code Modulation (PCM) em inglês, descoberto por Alec H. Reeves em 1938, enquanto trabalhava na IT&T. Sua patente mudou o Estado da Arte de telecomunicações, pois podia se obter transmissões tanto de vídeo como de áudio com uma qualidade e fidelidade maior do que com as tecnologias utilizadas anteriormente afetando então consequentemente toda a cultura que se havia prevalecente no período desde a décadas de 40 até a era atual contemporânea (Era da Informação). O objetivo principal desta invenção era prover sistemas elétricos de sinal com

praticamente nenhum ruído de fundo, mesmo em condições em que a taxa de sinal-ruído fosse normalmente entre 20 dB ou menos (Alec H. Reeves, 1942). Sua ideia era usar a representação binária de som para superar os ruídos provenientes de longa distância em transmissões analógicas de telefone. Porém todo este avanço não aconteceu imediatamente após a publicação de sua patente nomeada de “Electric Signaling System (US 2272070 A)”, seu conceito era muito diferenciado da tecnologia atual de telecomunicações e sua importância não foi largamente entendida ou apreciada (E. Maurice Delorraine, 1965), adicionando também com as palavras do próprio: “Eu sabia que não havia ferramentas existentes que poderiam conceber este mecanismo economicamente viável para o uso geral da população. Apenas nos últimos anos, nesta era de semicondutores, seu valor comercial tem sido percebido” (Alec H. Reeves, 1965). O primeiro uso do PCM começou na Segunda Guerra Mundial, a tecnologia não poderia ser concebida para a população em geral porém os militares dos Estados Unidos tinham recursos econômicos e intelectuais o suficiente para sua concretização, os Alemães haviam descoberto como decifrar o único sistema seguro de comunicação por voz utilizado pelo Aliados (A-3 Scrambler) e os principais oficiais do exército americano já desconfiavam dessa façanha, foi então criado o SYGSALY, que foi o sistema de voz criptografada utilizado pelo exército americano (IEEE Spectrum, 1984).



Figura 6 - Exemplo de conversão do Áudio Analógico (linha curvada) para Áudio Digital (linha tracejada).

Fonte: Digital Technical Journal. 1993.

Durante e após a Segunda Guerra Mundial, houveram grandes desenvolvimentos na área da computação moderna, devido à Teoria de Números Computáveis por Alan Turing em 1936 e do desenvolvimento de transistores mais práticos aplicados com circuitos integrados. Sendo que Jack

Kilby foi referenciado como o principal responsável pela invenção de circuitos integrados, sua ideia era de criar um corpo com materiais semicondutores, onde todos os componentes do circuito eletrônica eram completamente integrados entre si, recebendo o prêmio Nobel de Física em 2000 (Jack S. Kilby, 2000).

A procura pelo aperfeiçoamento da gravação e reprodução de som teve diversas implicações na história da humanidade, afetando indiscutivelmente nossos meios de comunicações, conflitos (guerra) pelo histórico apresentado. Atualmente há uma vertente de audiófilos que procuram pela alta-fidelidade de áudio, ou seja, de uma reprodução de áudio por um aparelho de som high-end (aparelhos com alta performance e qualidade), com a maior fidelidade possível do som real e é relacionado com a apreciação pela música em seu mais alto nível. Sendo uma paixão pela arte da música, procurando-se recriar para o ouvinte o máximo de realismo, emoção e intensidade do compositor (Robert Harley, 1994).

Os equipamentos high-end se utilizam de estereofonia para reprodução do áudio. Desse jeito é possível identificar a localização de cada músico com sua profundidade, altura e distância em relação ao equipamento de gravação também captura a ambientação da reprodução com suas reverberações e respostas acústicas. Usualmente para criação de um sistema de alta-fidelidade, o sistema é modulado; assim procura-se obter componentes de diferentes produtoras com especialização do tipo de componente, em vez de se obter um sistema integrado de som. Atualmente procura-se pelo aperfeiçoamento de equipamentos sonoros; desse jeito amplificadores de alta-fidelidade (ou high-end) utilizam válvulas eletrônica ou transistores de estado sólido, há também equipamentos high-end que utilizam de fibra óptica para transmissão de dados. Um sistema básico para reprodução sonora necessita de uma fonte de sinal, um CD-player, um Servidor de Músicas para músicas digitais ou um toca-discos para músicas analógicas, um amplificador integrado e um par de caixas acústicas, com os devidos cabos de interconexão, de força e de caixas acústicas.



Figura 7 - Exemplo de um sistema high-end em sala isolada, com amplificadores e caixas de som com estereofonia

Fonte: Robert Haley. *The Complete Guide to High-End Audio*. 2010.

1.1.1. DELIMITAÇÃO DO TEMA

Com o advento de computadores pessoais, foi possível a armazenagem de sons nos computadores, convertendo o sinal analógico do áudio para sinal digital, com a ajuda da Teoria de Modulação por Código de Pulso e também de conversores analógicos para digitais ou conversores digitais para analógicos. Sua qualidade é determinada pela taxa de amostragem (quantas vezes por segundo uma amostragem é tomada) e também pela profundidade de bits (a quantidade de amplitudes que um valor vai assumir). Porém a qualidade do áudio é subjetiva, assim deve-se verificar a qualidade do áudio de forma objetiva e subjetiva.

Desse jeito a codificação de formato de áudio pode ser descompactada, ou comprimida para reduzir o tamanho do arquivo em bit, usualmente utilizado na compressão *Lossy* (com perdas) (Gordon Laing, 2004).

Os formatos descompactados, WAV e AIFF, são considerados também

como containers de formatos de áudio *lossless* (sem perdas). Eles utilizam-se da Teoria de Modulação por Código de Pulso e assim eles são os tipos de arquivos que serão gravados primariamente quando se deseja gravar o som analógico para o computador. São os formatos com maior qualidade, pois ainda não foram comprimidos para otimização (diminuição) do tamanho de arquivo.

Os formatos de áudio *lossless* (sem perdas) atualmente podem ser do tipo: *Free Lossless Audio Codec*, *Apple Lossless Audio Codec* ou *Monkey Audio Codec*. São todos formatos de áudio comprimidos por algoritmos, a partir de formatos descompactados, com nenhuma perda de dados. Desse jeito eles podem ser reconstruídos de volta para os arquivos originais, em WAV ou AIFF. Codecs de áudio, em software, são programas de computadores usados para implementar algoritmos que irão comprimir ou descompactar dados digitais de áudio, reduzindo o tamanho de arquivo do tipo WAV ou AIFF (Yatri Trivedi, 2011).

A maioria dos formatos usados atualmente são do tipo *lossy* (com perdas), onde se há perdas da qualidade do áudio em troca de um arquivo com tamanho reduzido com o uso de compressores de arquivos *lossless*. Os tipos que são mais usados são MP3 (*MPEG 1 Audio Layer 3*), Vorbis, AAC (*Advanced Audio Coding*), ou WMA (*Windows Media Audio*). Diferentemente dos arquivos *lossless*, não é possível reverter o arquivo comprimido para o arquivo original, pois ocorrem perdas graduadas de qualidade conforme o número de compressões realizadas no arquivo de áudio. Sua taxa de bits é reduzida em 192 kbit/s ou 192 kbps usualmente, gerando menos dados para a reprodução do áudio, perdendo a preservação de detalhes do arquivo original (Yatri Trivedi, 2011).

Desse jeito procura-se o formato com a maior alta-fidelidade em reprodução sonora, com a comparação com os diferentes tipos de formatos de áudio. A definição de alta-fidelidade foi criada por Hartley H. A. em 1958, sendo está definida como uma reprodução com alta qualidade do som, idealmente os equipamentos de alta-fidelidade têm quantidades mínimas de ruído e distorção além de uma resposta à frequência mais precisa.

1.2. PROBLEMAS E PREMISSAS

As facilidades da digitalização nos processos de gravação e reprodução de áudio são um fato que tem apoiado o grande uso dessa tecnologia no mundo atual. Tendo em vista o grande uso dos formatos digitais em arquivos de áudio, e as tendências da indústria fonográfica, é natural que a busca pelo melhor tipo de formato de arquivo fique cada vez mais intensa.

A discussão a respeito dos tipos de compactação dessa mídia versus a qualidade sonora de reprodução é o alvo desta análise. Procurando-se avaliar quais são os pontos de maior relevância nessa comparação.

Tendo como objetivo a melhor reprodução, isto é, a maior fidelidade de reprodução em comparação com a fonte original (arquivo sem compressão), buscaremos identificar os fatores que afetam a reprodução, tanto objetivamente quanto subjetivamente.

Deseja-se identificar as diferenças entre os formatos de áudio compactado comparando com o formato sem compressão e também com outros formatos de compressão. Pretende-se comparar também comparar o próprio arquivo digital sem compressão com a fonte original, e analisar se apesar dos teoremas de componentes em frequência e taxa de amostragem dizerem que é possível recuperar o sinal em sua totalidade, cumpridas certas condições, se isso é verificado como verdade, mesmo que subjetivamente, pelos ouvintes.

1.3. OBJETIVOS

1.3.1. OBJETIVO GERAL

Avaliação dos diferentes tipos de formatos de áudio de maneira objetiva e subjetiva para identificar/qualificar a qualidade de áudio.

1.3.2. OBJETIVOS ESPECÍFICOS

- Estabelecer o histórico dos formatos de mídia digital para arquivos de áudio.
- Entender e conceituar que é uma boa reprodução de áudio.
- Realizar uma pesquisa sobre as diferenças de áudio em ambiente controlado, com escolha de amostras que permitam fazer diferença dos tipos de formatos.
- Analisar matematicamente o sinal de onda para os diferentes tipos de formatos.

1.4. JUSTIFICATIVA

Um dos motivos de formatos de compressão de áudio terem sido desenvolvidos era a tamanho dos arquivos e dificuldade da transmissão de arquivos maiores. Atualmente esses motivos que eram grandes empecilhos já não representam tanta dificuldade de modo que não há mais a necessidade de escolher o formato de compressão que resulta no menor tamanho de arquivo. Além disso, existem vários formatos de arquivos de áudio comprimido e dentro de cada formato há diversos parâmetros que podem ser alterados. Assim inicia-se o cuidado e a dúvida de qual método, qual formato de compressão de áudio permite maior fidelidade ao arquivo original sem compressão.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. ÁUDIO DIGITAL/ANALÓGICO

Na gravação analógica o sinal gravado é incapaz de distinguir a diferença entre sinais desejados e indesejados. Sendo que os sinais indesejados podem ser considerados como distorções, ruídos e outras formas de interferências introduzidas no processo de gravação. Por exemplo, um toca disco não consegue distinguir a diferença entre o movimento da agulha no disco (indesejado) e aquele causado por um alto transiente na música (desejado). Imperfeições no disco podem ser reproduzidos por clicks, rachaduras e outros; porém são estas características que fazem da reprodução analógica uma experiência mais natural e fiel da reprodução do som (Francis Rumsey, 2009).

Diferentemente da gravação analógica, a gravação digital converte a forma de onda elétrica utilizando-se de um microfone, para uma série de números binários, sendo que cada número representa a amplitude do sinal em um único instante do tempo. Desse jeito os números serão gravados de uma forma codificada (podendo ser binário ou hexadecimal por exemplo), permitindo que o sistema detecte se o sinal é coerente ou não. O dispositivo de reprodução pode então distinguir entre sinais desejados e não desejados, excluindo os ruídos, distorções e outras formas de interferência (Francis Rumsey, 2009).

O áudio digital possibilitou aos engenheiros de som, com o auxílio do desenvolvimento da indústria de computadores, a produção em massa de discos ou outras formas de reprodução em uma escala que não era possível anteriormente. Atualmente é comum que o som seja gravado, processado e editado de uma forma relativamente barata em computadores, e esta é uma tendência provável de se continuar (Francis Rumsey, 2009).

Todos sistemas de áudio digital usam o sistema de numeração binário. Binário significa a possibilidade de 2 valores: 0 ou 1. Em áudio digital, o sinal analógico representa uma sequência destes números binários.

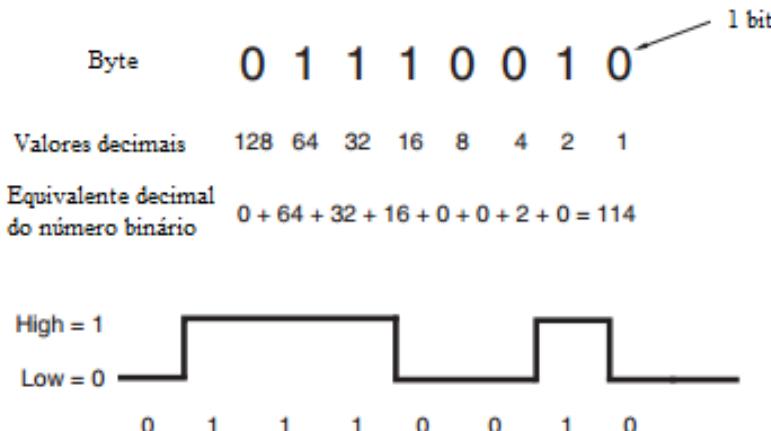


Figura 8 - Números binários podem ser representados por PCM por uma sequência de tensões altas (high) e baixas (low)
Fonte: Sound and Recording (Francis Rumsey, 1992).

A amostragem e a quantização foram primordiais para o áudio digital, eles preservam o sinal analógico em uma série de números, sendo que o sinal representado é uma tensão que varia ao longo do tempo. Quanto mais rápido a variação ao longo do tempo, maior a frequência do sinal do áudio e também quanto mais oscila a amplitude, mais alto será o sinal. Desse jeito o sinal de áudio tem duas variáveis que devem ser codificadas para a preservação do sinal, o tempo e a amplitude.

O disco LP (*Long Play*, popularmente conhecido com disco de vinil) é um bom exemplo de como as informações de tempo e amplitude são preservadas, pois suas modulações gravadas nos sulcos do disco, codificam a informação da amplitude, assim quanto maior a modulação, maior será a amplitude do sinal. Já a informação do tempo é codificada pela rotação do LP, que deve ser igual à reprodução de quando o disco forá gravado, assim se for mudada a velocidade do LP muda-se a relação do tempo e, portanto, a frequência do sinal de áudio (Robert Harley, 2010).

O áudio digital deve preservar as informações de amplitude e tempo de um sinal de áudio, mas ao invés de codificar e gravar estas características continuamente para o LP, o áudio digital preserva o tempo e a amplitude em sinal discreto (Robert Harley, 2010).

O sistema de áudio digital codifica a informação do sinal de tempo, fazendo sua amostragem em intervalos de tempo discreto. A informação da amplitude é codificada pela geração de um valor para cada instante da amostragem, representando a amplitude da onda analógica em um certo tempo da amostragem. Chamamos este processo de quantização. Desse jeito a amostragem armazena a informação de tempo, enquanto que a quantização armazena a informação da amplitude (Robert Harley, 2010).

Tanto a amostragem quanto a quantização, produzem "palavras" (série de números binários) que representam a onda analógica. Quando estas palavras binárias são convertidas de volta para tensão, com o mesmo tempo de referência de quando foram amostrados, a onda de áudio analógico é reconstruída. A forma de onda original é recuperada pela suavização do sinal discreto por um filtro passa-baixas, que permite a passagem de baixas frequências e atenua (ou reduz) a amplitude das frequências maiores que a frequência de corte (Robert Harley, 2010).

A amostragem e a quantização convertem a função analógica contínua (a variação contínua de tensão na forma de onda analógica) em uma série de números binários discretos.

Na figura 9, pode-se ver uma conversão de sinal analógica para digital com variação de tensão de 0 a 7 volts, sendo que na amostragem apenas vê-se valores inteiros para o sinal digital.

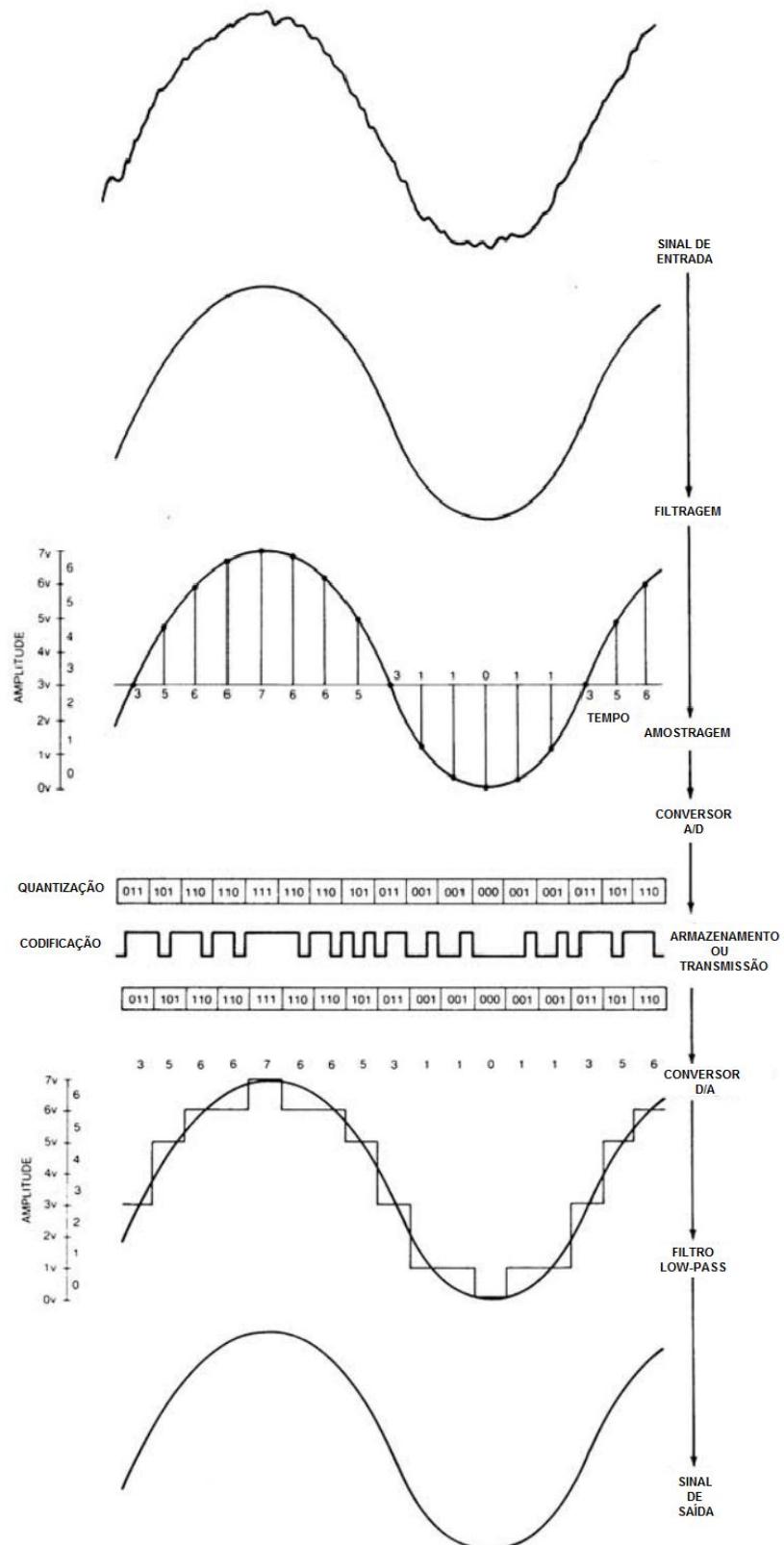


Figura 9 - Representação gráfica do sinal em um sistema PCM completo de digitalização
Fonte: Ken Pohlmann

Deste exemplo pode-se inferir que na conversão de um sinal analógico para um sinal digital, a amostragem é um processo sem perdas. A onda pode

ser reconstruída de modo idêntica à forma de onda original. Isto é verdade porque o sinal de entrada é propriamente limitado por faixa e uma taxa de amostragem apropriada para as características do filtro escolhido. Outros problemas em áudio digital serão discutidos a seguir (Robert Harley, 2010).

2.1.1. Amostragem

A velocidade com que as amostras de som são capturadas é chamada de taxa de amostragem, ou frequência de amostragem, ela determina a maior frequência de áudio para se obter uma codificação com sucesso. Especificamente a taxa de amostragem deve ser duas vezes maior do que o sinal de áudio que se deseja preservar (Nyquist, 1928). O formato do disco compacto, por exemplo, faz a amostragem do sinal analógico à 44.100 palavras por segundo (44,1 kHz), o que resulta em um áudio com largura de banda (audio bandwidth) de aproximadamente 20 kHz. Uma amostragem de 96 kHz, resultaria em uma largura de banda de aproximadamente 48 kHz. Desse jeito a taxa de amostragem deve ser duas vezes maior do que a maior frequência que se deseja codificar.

Se o teorema de Nyquist for violado em uma amostragem aonde o sinal tem uma frequência maior do que a metade da taxa de amostragem, ocorre uma forma de distorção chamada *aliasing*. Ela cria novas frequências no sinal amostrado que não estavam presentes no sinal original. Por exemplo se fizermos a amostragem de uma onda senoidal com 33 kHz com uma taxa de amostragem de 48 kHz, é criado um novo sinal com 15 kHz (48 kHz - 33 kHz) que não tem como ser removido posteriormente. O termo "alias" de aliasing vêm do fato de que o sinal incorreto da amostragem assume uma nova forma como sinal de baixa frequência (Robert Harley, 2010).

Para prevenir isto é necessário que o conversor não veja um sinal maior em frequência do que metade da taxa de amostragem. É utilizado então um filtro de entrada passa baixa que remove as irregularidades (harmônicos) que estejam acima da metade da taxa de amostragem (Robert Harley, 2010).

Como exemplo será usado uma amostragem de 10 kHz e uma onda senoidal de 20 kHz. Apesar de o amostrador enxergar apenas 2 amostras por onda, ele tem informação suficiente para codificar a onda de 20 kHz. Como mostrado na figura abaixo. Uma onda quadrada é uma onda senoidal combinada com os harmônicos ímpares, estes harmônicos são representados como as 4 formas de onda de alta frequência, abaixo da onda quadrada de 20 kHz. Quando estes harmônicos são removidos por um filtro passa baixa, sobra apenas a onda de 20 kHz que é a onda inicial da amostragem.

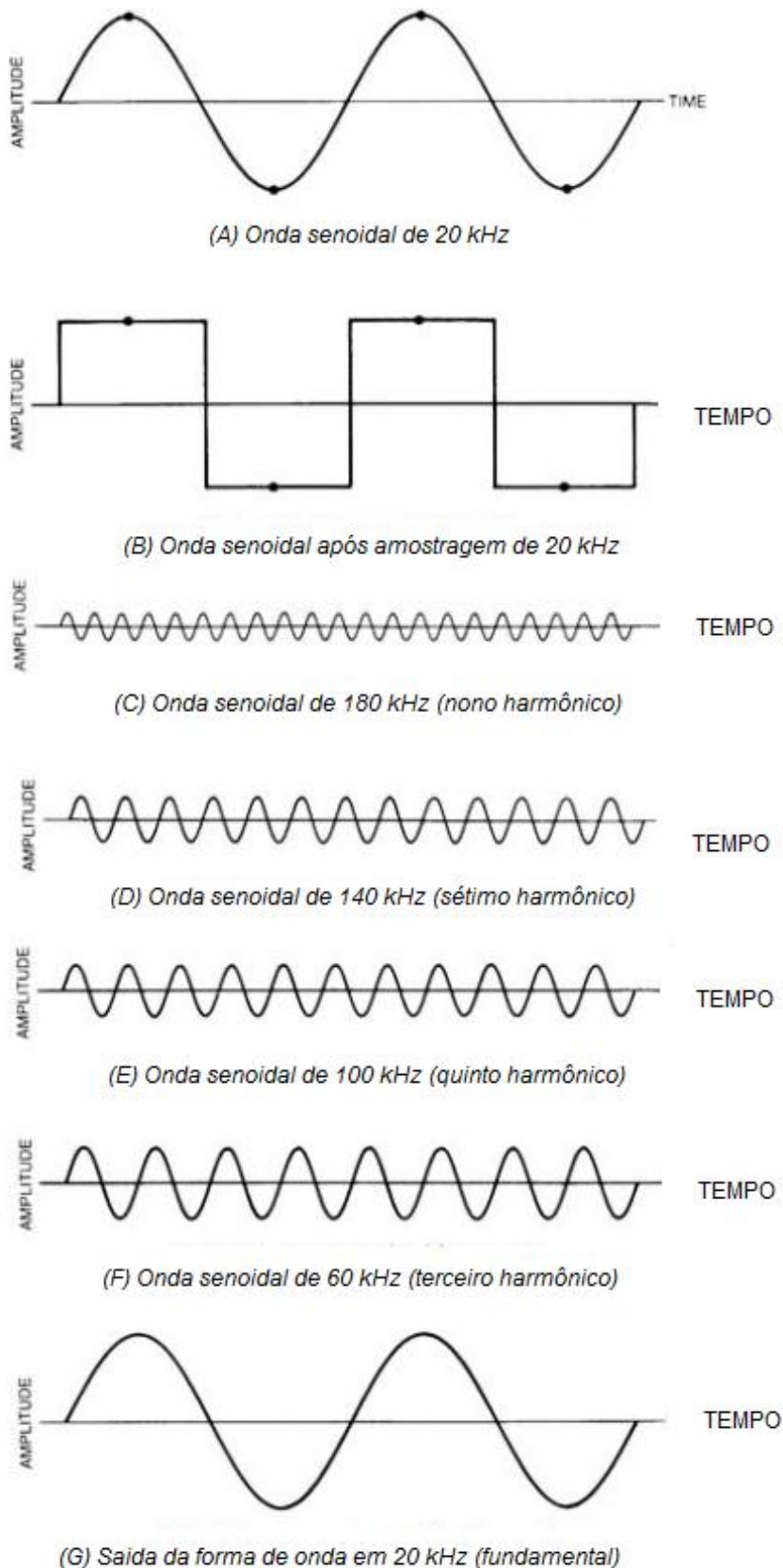


Figura 10 - Onda senoidal de 20 kHz amostrada à 40 kHz produzindo uma onda quadrada. O filtro passa baixa remove as harmônicas acima de 20 kHz, recuperando a onda senoidal de 20 kHz
Fonte: Ken Pohlmann

2.1.2. Quantização

A quantização gera o número binário que representa a amplitude da forma de onda analógica em um dado tempo da amostragem. O número binário é uma representação digital por tensão do sinal analógico quando a amostragem é feita. O número de bits disponíveis para se codificar a forma de onda de um áudio, é chamada de “quantização do comprimento da palavra” (*quantization word length*). Ela determina o intervalo dinâmico, a resolução e o nível de distorção em um sistema de áudio digital. Os sistemas de áudio digital mais comuns, incluindo os CD's, usam palavras de 16-bit. Enquanto que sistemas de áudio com alta resolução (*high-resolution*), utilizam palavras de 20 bits ou mais. Quanto maior o comprimento da palavra, maior será a fidelidade ao sinal original (Robert Harley, 2010).

O comprimento da palavra define o número de degraus disponíveis para codificar o sinal de áudio. Especificamente podemos chamar o número de degraus por 2^x , aonde x é o número de bits na palavra. Por exemplo uma quantização de 16-bits resulta em 2^{16} degraus (256) para quantizar a amplitude do sinal analógico (Robert Harley, 2010).

Diferente da amostragem, que é um processo sem perdas, a quantização produz erros no sinal codificado. Converter uma função analógica de variável infinita em um número discreto é implicitamente uma aproximação. Este erro ocorre porque a amplitude representada por uma palavra quantizada nunca é exatamente a tensão do sinal analógico. Esta diferença entre a amplitude do sinal analógico e da amplitude representada pela palavra quantizada é chamado de erro de quantização.

A figura 11 mostra como erros de quantização ocorrem. A amplitude do sinal analógico não é exatamente o mesmo valor representado pelas palavras quantizadas. Por exemplo, na primeira quantização (que seria o ponto entre a ordenada 101000 e 100111), não existe o valor de 100111.25, então o quantizador arredonda para o valor discreto de quantização mais próximo (neste

caso seria de 200111), apesar de o número não ser corretamente o valor que se procura. A diferença entre a tensão representada por 100111 (1.3V) e a tensão atual de 1.325V é chamada de erro de quantização. Quando o sinal de áudio é reconstruído, o número binário 100111 irá representar uma amplitude diferente do valor real. O resultado é uma distorção da forma de onda original.

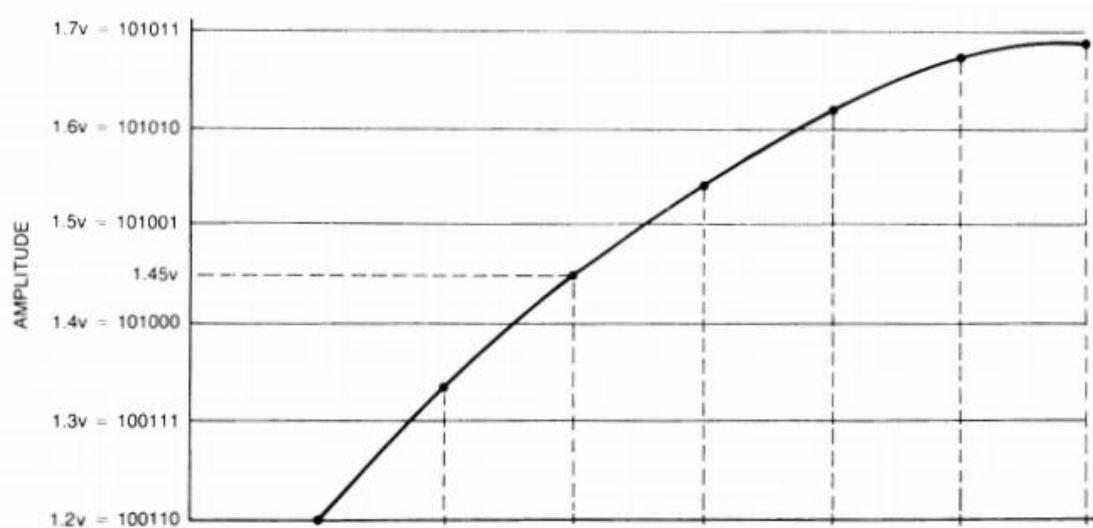


Figura 11 - Erro de quantização limitado à uma metade do bit menos significante
Fonte: Ken Pohlmann

O pior caso acontece quando a amplitude analógica está exatamente entre 2 níveis de quantização, como mostrado na figura 11. A diferença entre a amplitude analógica e a palavra quantizada representando esta amplitude é a maior neste caso.

O erro de quantização pode ser expressado como uma porcentagem do bit menos significante (LSB de *least significant bit*). Na primeira amostragem da figura, o erro era de 1/4 do LSB (1.325V), já na segunda amostragem da figura, o erro foi de 1/2 do LSB (1.45V). O erro de quantização nunca pode exceder a metade do valor de amplitude de um LSB. Consequentemente, quanto menor for o tamanho do passo de quantização, menor será o erro de quantização. Adicionar um bit para o comprimento da palavra de quantização irá dobrar o número de passos de quantização e diminuir os erros de quantização pela metade. Como este processo representa uma diferença de 6 dB, a razão de sinal-ruído se aprimora em 6 dB para cada bit adicional na palavra de

quantização. Um sistema digital com quantização de 18-bits irá ter 12 dB à menos de ruído da quantização do que um sistema com quantização de 16-bits (Robert Harley, 2010).

Pode-se também aproximar a razão de sinal-ruído de um sistema digital pela multiplicação do número de bits em uma palavra de quantização em 6 vezes. Uma quantização de 16-bits resulta em uma faixa dinâmica teórica de 96 dB. Um sistema digital de 20-bits tem uma faixa dinâmica de 120 dB, ou 24 dB maior do que o sistema de 16-bits.

O erro de quantização é audível como um som granular e áspero em sinal de baixo nível, com decaimento de reverberação. Em vez de ouvir o som decaindo para um silêncio, o decaimento da reverberação se transforma em ruídos enquanto o sinal enfraquece. Isto acontece porque o erro de quantização se torna numa grande porcentagem do sinal enquanto a amplitude do sinal diminui (Robert Harley, 2010).

Este aumento da distorção enquanto o nível do sinal cai é único no áudio digital. O problema é o aumento da distorção enquanto os níveis do sinal diminuem, fazendo a distorção mais audível. Porém o aumento do comprimento da palavra de quantização de 16-bits para 20-bits pode reduzir este problema.

É importante notar que a razão de sinal-ruído e distorção são fatores específicos para sistemas de áudio digital em um sinal de escala natural. Na maioria das vezes, o sinal da música está abaixo da escala natural e então perto do nível de distorções. A distorção não é função de quantos bits o sistema tem disponível, mas sim do número de bits que são usados para o sinal em um dado momento. Portanto a distorção e o ruído são inversamente proporcionais à amplitude do sinal, e é devido a isto que em áudio digital existem problemas com sinais baixos (Robert Harley, 2010).

Estes fatores fazem da configuração dos níveis de gravação em sistemas digitais completamente diferente da configuração dos níveis de gravação analógica. Idealmente, o maior pico na gravação de áudio deve alcançar uma

escala natural (*full-scale*) digital (usando todos bits disponíveis). Se o nível da gravação está configurado para que o maior pico alcance -6 dB, é o mesmo que jogar fora um bit da palavra de quantização, com auxílio da redução de 6 dB para a razão de sinal-ruído. E se a amplitude do sinal analógico é maior do que a tensão representada pelo maior número, então o quantizador fica sem bits necessários e repete o maior número, fazendo então com que a forma de onda fica com um pico plano. Esta forma de onda distorcida produz um som de "trituração" em picos. Olhando para os níveis de gravação de um CD, se os maiores picos nunca alcançam a escala natural, isto quer dizer que parte da resolução do áudio foi perdida.

É importante notar que um programa de áudio com faixa dinâmica mais abrangente estará mais perto do nível de ruído na quantização do que um sinal com faixa dinâmica mais limitada. Os picos do sinal com faixa mais abrangente estarão abaixo da escala natural, e os sinais com níveis muito baixos estarão consequentemente codificados com menos bits. Este é um erro bem conhecido na música clássica, que tem uma faixa dinâmica bem abrangente. Diminuir a faixa dinâmica da música faz com que ela seja alta o tempo todo, ao custo do dinamismo e ritmo da música (Robert Harley, 2010).

Os níveis de sinal digital são referenciados como sinal de escala natural, que ocorrem quando todos bits são usados e não há sinais altos que não podem ser codificados. Este nível de referência é chamado de 0 dBFS, sendo que FS significa *full-scale* (escala natural). Por exemplo -20 dBFS seriam 20 dB abaixo da escala natural.

2.1.3. *Dither*

O caso mais extremo do erro de quantização, ocorre quando a amplitude do sinal é menor do que 1 LSB, como na figura 12. Este sinal de baixo nível não é codificado pelo quantizador, que fornece o mesmo código a cada ponto de amostragem e então a informação acaba sendo perdida como nos casos do item (C) e (D). Se o sinal ignorar os passos de quantização, o sinal codificado será

uma onda quadrada, que representa uma significante distorção do som original. Desse jeito qualquer informação com amplitude menor do que o LSB será perdida (Robert Harley, 2010).

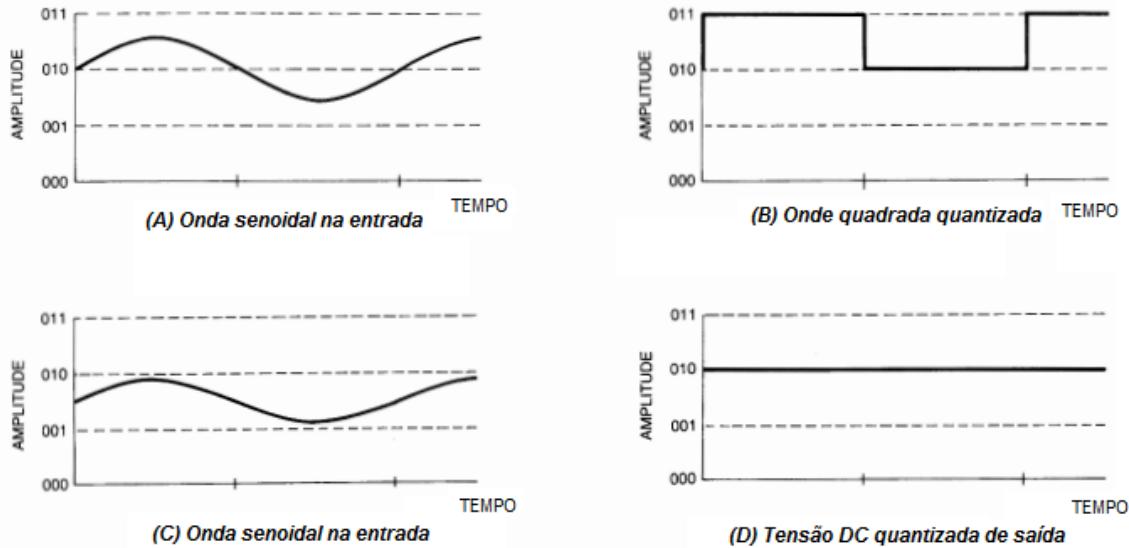


Figura 12 - O erro de quantização é relativamente alto em um sinal que passa por apenas alguns passos de quantização
 Fonte: Ken Pohlmann.

Felizmente este limite pode ser superado pela adição de pequenas quantidades de ruídos, chamados de *dither*, no sinal de áudio. O *dither* permite que o quantizador resolva sinais abaixo do LSB, e melhorem o som do áudio digital. Na figura a 13 foi adicionado uma pequena quantidade de ruído branco para o sinal de áudio. Este ruído faz com que o sinal de áudio ignore os níveis de quantização, permitindo que o sinal seja codificado. A onda senoidal original é melhor preservada na modulação de pulso por comprimento como mostrado na figura 13.

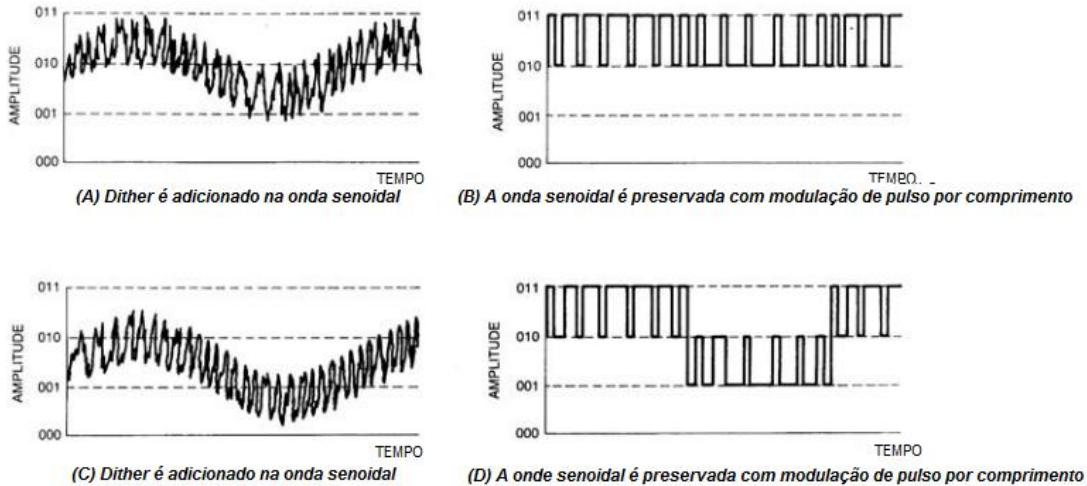


Figura 13 - Dither pode suavizar os efeitos do erro de quantização
Fonte: Ken Pohlmann

O *dither* reduz os efeitos do erro de quantização, permite que o sistema resolva informações com amplitudes menores do que uma metade do LSB, e faz com que o áudio digital fique com um som mais "análogo". Entre outros benefícios, o *dither* aprimora resoluções de baixo nível e também suaviza o decaimento da reverberação, responsável pelo som granulado. Apesar destas vantagens o *dither* requer uma pequena penalidade na razão de sinal-ruído.

2.1.4. Armazenamento de áudio digital

Assim uma gravação digital é produzida pela conversão das propriedades físicas do som original, em uma sequência de números, que então será armazenada em um meio para então ser reproduzida posteriormente. Normalmente o som é convertido em um sinal analógico elétrico e então convertido em sinal digital pela amostragem e quantização, através de um conversor analógico-digital, para então ser armazenado em um meio digital como discos compactos ou HD's.

Após feito a conversão do sinal de áudio em sua forma digital, é necessário seu armazenamento em algum meio. Os meios mais comuns para armazenamento são o disco compacto, memória *solid-state*, discos óptico-magnéticos, e HD's. O sistema de gravação de áudio digital pode ser pensado

como o armazenamento de informação binária representando o sinal de áudio diretamente no meio de armazenamento. Por exemplo, um ponto refletivo no disco pode representar o número binário "1", enquanto um ponto não refletivo representa o "0". Porém na prática o áudio digital sempre é codificado com um processo de modulação.

2.1.4. Conversão analógica/digital

O processo básico de conversão D/A é exemplificado na figura a seguir. As palavras da amostragem são convertidas de volta em níveis de tensões, em forma de escada, correspondentes aos valores amostrados. Isto é feito em conversores, usando os estados dos bits para ligar ou desligar as fontes correntes, compensando os pulsos de amplitude requeridos pela combinação da saída de cada uma das fontes. Esta onda em forma de escada é então "reamostrada" para reduzir a largura dos pulsos antes que eles passem pela reconstrução do filtro passa baixa, onde a frequência de corte é a metade da frequência de amostragem. A reconstrução tem a finalidade de unir os pontos de amostragem para fazer uma forma de onda mais contínua. A reamostragem é necessária para evitar quaisquer descontinuidades do sinal em limites da amostra e também porque o efeito de média do filtro iria resultar numa redução da amplitude em sinais de alta frequência ("aperture effect" ou efeito de abertura). O efeito de abertura pode ser reduzido pela limitação do comprimento dos pulsos de amostragem para 1/8 do período de amostragem (Francis Rumsey, 2009).

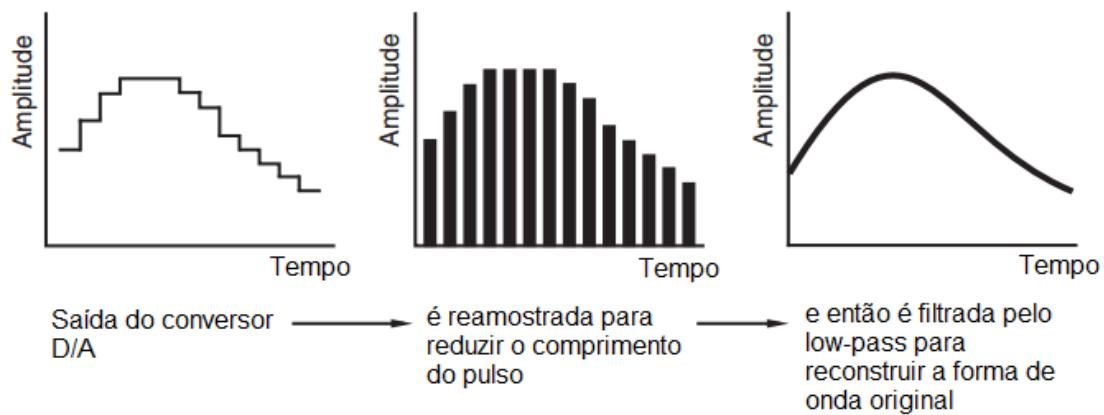


Figura 14 - Processo da conversão D/A com valores positivos
Fonte: Francis Rumsey e Tim McCormick

2.2.QUALIDADE DO SOM

Os termos qualidade do som e design sonoro começaram a surgir nas décadas de oitenta e noventa, quando as técnicas de engenharia que reduziam a intensidade dos ruídos emitidos por um equipamento ficaram insatisfatórios, então surge a necessidade na indústria de quantificar melhor as características mais complexas dos sons e das sensações de correspondentes ao ouvir.

2.2.1. Fisiologia da audição

A compreensão da estrutura sonora exige o conhecimento preliminar do sentido físico da audição, ou sensação auditiva, que é o fenômeno objetivo, bem como do fenômeno psicológico da percepção e apreciação do som, o qual ocorre na mente, e constitui o fenômeno subjetivo. No primeiro caso, estuda-se o órgão anatômico responsável pela sensação de audição, que é o ouvido ALVARENGA (2008).

2.2.1.1. Anatomia do Ouvido

Para um estudo esquemático, o ouvido divide-se em três partes: ouvido externo; ouvido médio; ouvido interno (ALVARENGA, 2008).

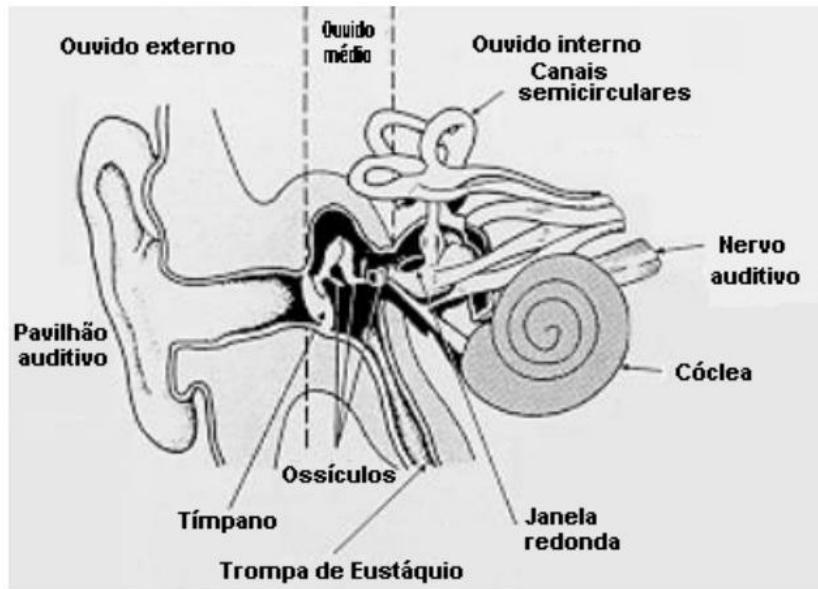


Figura 15 - Anatomia do ouvido
Fonte: ALVARENGA (2008, p. 32).

O ouvido externo é constituído pelo pavilhão auditivo (orelha) e o canal auditivo externo, ou meato auditivo externo, o qual termina no tímpano (ALVARENGA, 2008).

O ouvido externo tem características direcionais, serve também como uma espécie de amplificador sonoro. Quanto ao canal auditivo, sendo um tubo fechado em uma das extremidades, ele possui uma frequência de ressonância, frequência esta que foi verificada experimentalmente situar-se em torno de 3.800 Hz (ALVARENGA, 2008).

O ouvido médio está numa cavidade do osso temporal denominada de cavidade timpânica. Na parede interna do tímpano ficam contidas as janelas ovais, na parede superior, e redonda, na inferior. O ouvido médio também contém os ossículos denominados de martelo, bigorna e estribo, cuja função é acoplar ou unir mecanicamente o tímpano ao caracol ou cóclea, através da janela oval (ALVARENGA, 2008).

Uma das funções do ouvido médio é proteger o ouvido interno de sons de alta intensidade e mudanças súbitas de pressão, seja contraíndo o tímpano, seja afastando o estribo da janela oval, através de dois conjuntos de músculos. No

ouvido médio se encontra também a Trompa de Eustáquio, que se liga à garganta e equilibra a pressão interna de acordo com a pressão atmosférica (ALVARENGA, 2008).

O ouvido interno também chamado de labirinto, consiste em uma série de espaços, ou labirinto ósseo no interior dos ossos, dentro do qual se encontra o labirinto membranoso, que é um sistema fechado de bolsas e canais que se comunicam. O labirinto membranoso se encontra cheio de um líquido, a endolinfa, e é rodeado pela perilinfa, a qual preenche o labirinto ósseo; é composto de duas bolsas, outrículo e o sáculo. Iniciando no sáculo encontra-se o canal coclear, ou condutor coclear, que é um tubo longo enrolado dentro da perilinfa da cóclea (ALVARENGA, 2008).

A cóclea é enrolada helicoidalmente em torno de um eixo central chamado modíolos, tomando a forma de um cone arredondado, ou concha de caracol. Ela possui uma membrana mediana, que a divide em dois canais (ou rampas): o canal vestibular e o canal timpânico, os quais se comunicam na extremidade da cóclea (chamada Apex ou Helicotrema) (ALVARENGA, 2008).

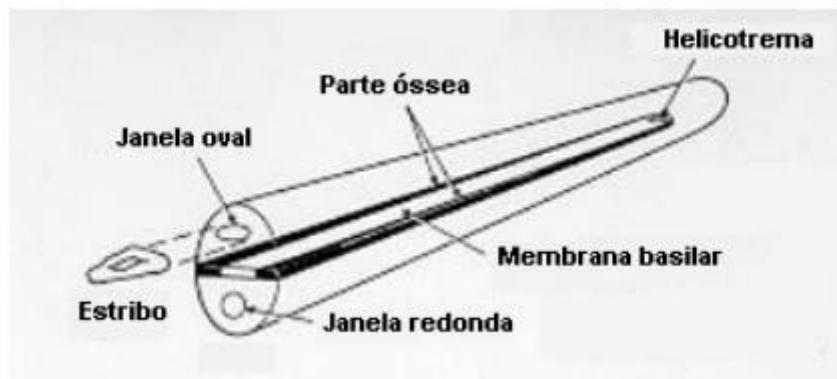


Figura 16 - Ouvido interno
Fonte - ALVARENGA (2008, p. 33).

Esta membrana central, que contém o canal central, chamado ducto ou canal coclear, o qual abriga o órgão essencial da audição, o órgão de Corti, é chamada de membrana basilar ou lâmina espiral. Ela possui cerca de 25.000 filamentos ou fibras de comprimento decrescente na direção do seu vértice. No órgão de

corti se encontram dois tipos especializados de células nervosas chamados de células ciliadas - fileira única de células ciliadas internas, em número de cerca de 3.500 e medindo por volta de 12 micrômetros de diâmetro e três ou quatro fileiras de células ciliadas internas em número cerca de 12.000 e tendo diâmetros de apenas 8 micrômetros. As bases e os lados das células ciliadas fazem sinapse com a rede de terminações nervosas da cóclea. Entre 90% e 95% dessas terminações terminam nas células ciliadas internas, o que demonstra sua importância na detecção do som (Hall, 2011).

Essas células ciliadas presentes no órgão de corti são eletromecanicamente sensíveis e constituem os órgãos receptores finais que geram impulsos nervosos em resposta às vibrações sonoras (Hall, 2011).

No ouvido interno, os canais semicirculares se comunicam com outrículo, e relacionam-se com o sentido de equilíbrio ou orientação corporal (Hall, 2011).

2.2.1.2. Mascaramento

Mascaramento é o fenômeno pelo qual os sinais altos impedem que os ouvidos ouçam sons suaves. O maior efeito do mascaramento ocorre quando a frequência do som e a frequência do ruído de mascaramento estão próximas uma da outra. Por exemplo, um tom de 4 kHz irá mascarar um tom mais suave de 3,5 kHz, mas terá pouco efeito sobre a audibilidade de um silencioso tom de 1 kHz.

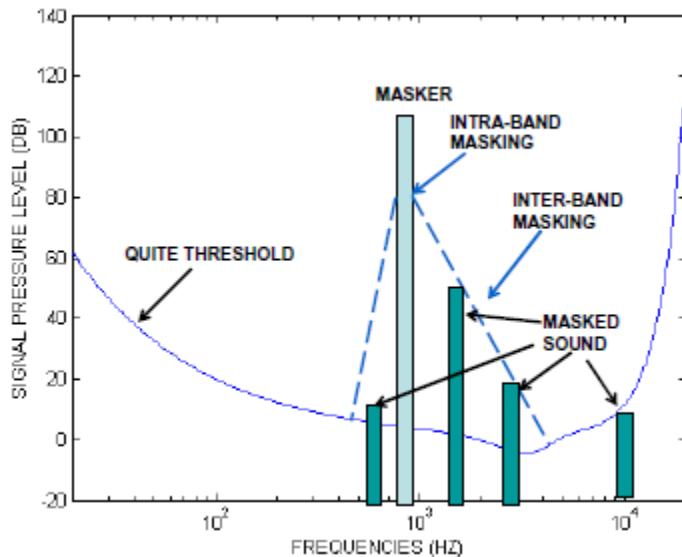


Figura 17 - Mascaramento auditivo
Fonte: Li (2002).

O mascaramento também pode ser causado por harmônicas do tom mascarador, por exemplo, um tom de 1 kHz com uma harmônica de 2 kHz forte poderia mascarar um tom de 1,9 kHz. Esse fenômeno é uma das principais razões porque o posicionamento do som estéreo e a equalização são tão importantes para o processo de mixagem. Um instrumento que por si só soa bem pode ser completamente ocultado ou alterado por instrumentos mais altos que têm um timbre semelhante (Taylor & Francis, 2011).

Alguns métodos de compressão de áudio, como o MP3, se aproveitam desse fenômeno em sua implementação.

"Não serve de nada falar de engenharia de som sem antes falar do Som" (FONSECA, 2008).

2.2.2. Som

Entende-se então que o som corresponde às variações da pressão do ar que conseguem ser captadas pelo ouvido. Para que o ouvido humano consiga captar uma dessas variações, ela deve estar entre o intervalo de 20 e 20.000 Hz, ou seja, variar entre 20 e 20.000 vezes por segundo. Utiliza-se a unidade de medida *hertz* (Hz) para representar frequência.

O som é produzido pelo choque entre corpos rígidos (sólidos), fluidos

(líquidos) ou elásticos (gases). Além de produzir o som, estes corpos podem também transportá-lo, ou seja, o som pode se propagar através deles, pela vibração das moléculas do meio (ALVARENGA, 2008).

Existem diversas formas de se estudar o fenômeno sonoro, embora todas as formas estejam interligadas, cada uma delas apresenta um foco específico. A acústica física trata sobre o aspecto material, já a psicoacústica estuda a percepção do fenômeno sonoro. A acústica musical relaciona os dados dessa disciplina com a atividade artística. A ondulatória é a parte da física que trata dos fenômenos que apresentam dessa maneira, as ondas mecânicas (longitudinais, associado ao fenômeno sonoro) e eletromagnéticas (transversais, associado à luz) (BACKUS, 2006).

Existem duas formas de vibração, que podem ser visualizadas através de uma experiência. Tomando-se uma mola, e faz-se com que ela vibre. A mola pode vibrar de dois modos diferentes: no sentido transversal, perpendicular à direção do movimento, e no sentido longitudinal, que ocorre na própria direção do movimento, como se vê na Figura 18 (ALVARENGA, 2008).

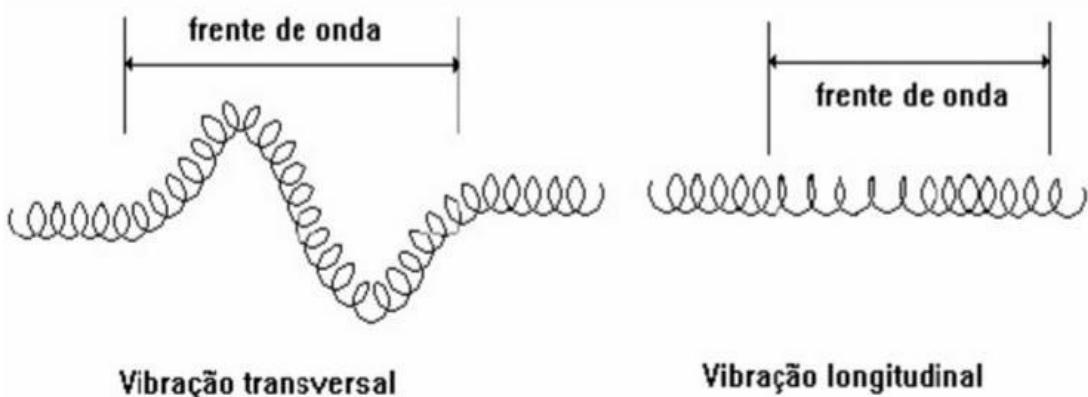
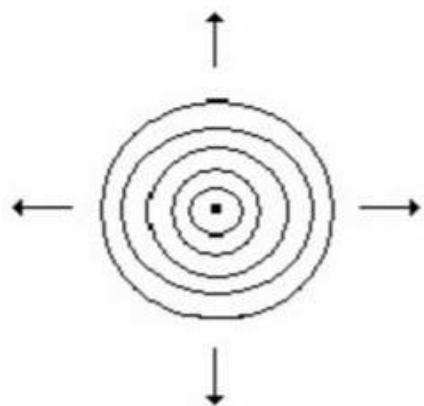


Figura 18 - Os dois tipos de vibração de ondas
Fonte: ALVARENGA. (2008, p. 13).

A vibração das moléculas no ar pelo qual o som se propaga é no sentido longitudinal, na mesma direção à qual o som se propaga, as moléculas de ar, ao receberem a vibração sonora, agrupam-se em partes concentradas seguidas de partes rarefeitas, num movimento que se comunica às moléculas vizinhas, fazendo com que o som se propague. Este movimento é denominado de *onda de compressão*, ou *onda sonora*. É evidente que o que se desloca é esta onda

sonora, e não as moléculas de ar. O corpo que recebe a onda sonora é sacudido pela vibração das moléculas de ar que lhes estão próximas, e não pelas moléculas de ar adjacentes ao corpo que produziu o som. No vácuo, o som não se propaga, porque não existe o meio para o deslocamento da onda. Os sólidos transmitem melhor o som que os líquidos, e estes, melhor que os gases. Esta propagação segue em todos os sentidos vistos da figura 19.



Propagação da onda sonora

Figura 19 - Propagação da onda sonora
Fonte - ALVARENGA (2008, p. 14).

A maioria das fontes sonoras tem uma propagação esférica, ou seja, o som vai-se expandindo em todas as direções. Isso significa que à medida que a distância aumenta, o som terá de cobrir uma área maior, tornando-se mais fraco (FONSECA, 2007).

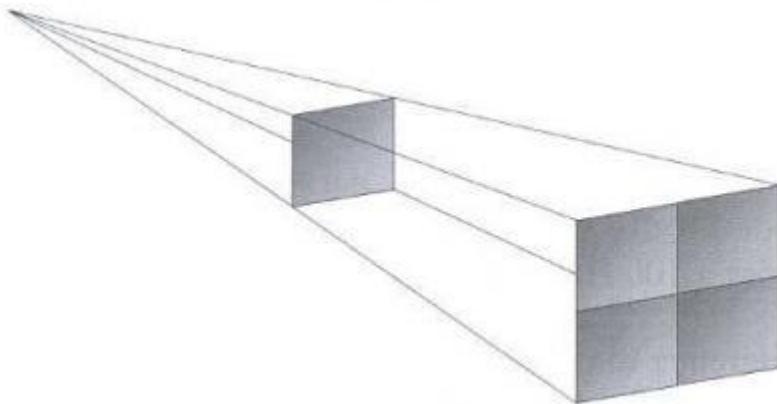


Figura 20 - Diminuição da pressão sonora
Fonte - FONSECA (2007, p. 10)

Em condições normais atmosféricas, a velocidade média de propagação do som é de cerca de 330 metros por segundo, mas essa velocidade de propagação varia com a temperatura, como se vê no Quadro 1:

Temperatura	Velocidade do som
0 °C	331,4 m/s
15 °C	340,5 m/s
30 °C	349,6 m/s

Quadro 1- Variação da velocidade do som em relação a Temperatura
Fonte - FONSECA (2007, p. 10).

2.2.3. Estudo Matemático do Som

O estudo das vibrações sonoras pode ser facilitado pelo uso de representações gráficas equivalentes, estas representações gráficas são denominadas de *funções senoidais* ou mais simplesmente, *senoides* (ALVARENGA, 2008).

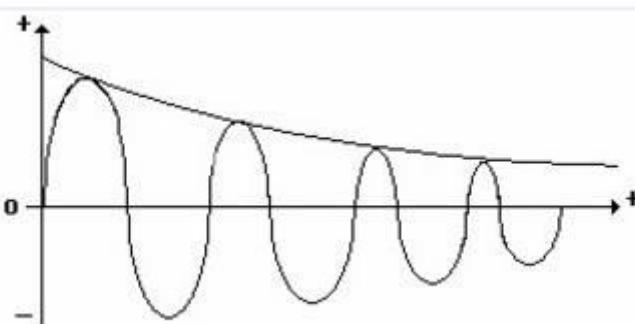
A senoide é uma função matemática que define uma variação simétrica e periódica qualquer em relação ao tempo; esta função é colocada nas

coordenadas retangulares XY, onde o ponto em que os eixos se cortam representa o zero da função.



Figura 21 - Função senoidal
Fonte - ALVARENGA (2008, p. 16).

A figura 21 representa a vibração ou oscilação cujo o movimento é constante. O gráfico, visto na figura 22 a seguir mostra uma oscilação cujo o movimento periódico não se mantém, ou seja, a oscilação amortecida com o tempo.



Função senoidal amortecida
Fonte - ALVARENGA (2008, p. 16).

Dessa forma, o eixo Y é representado a intensidade das oscilações, e o eixo X representa o tempo decorrido. Podemos aprofundar a relação entre a função senoidal e a efeito sonoro.

Observa-se o a figura 23:

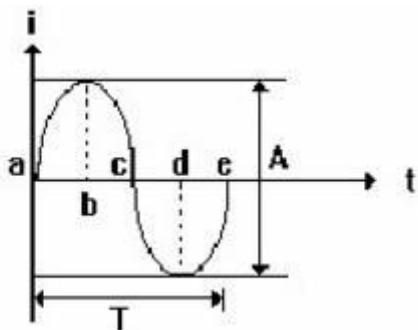


Figura 23 - Função senoidal mais específica
Fonte - ALVARENGA (2008, p. 17).

Os pontos denominados **a**, **c**, **e** são denominados *nós da senoide*; os pontos **b**, **d** são denominados *ventre, vales ou antiondas* da senoide. A distância entre os nós **a**, **e**, define o período T da onda sonora, ou seja, o intervalo de tempo que decorre entre o início do primeiro movimento da vibração e o fim do segundo movimento. O inverso do período define a frequência (também denominada *altura - pitch*) da onda sonora, que é o número ou quantidade de vibrações por unidade de tempo. Sendo o inverso do período, então:

$$f = \frac{1}{T}$$

Onde,

T é o período medido em ciclos.

f é a frequência medida em Hertz (Hz).

O estudo do comprimento de onda é importantíssimo, pois é vital para o comportamento que o som terá ao encontrar determinadas superfícies/obstáculos. Por exemplo, se o som encontrar um obstáculo com dimensões inferiores ao comprimento de onda, irá conseguir contorná-lo. Caso o obstáculo tenha dimensões superiores ao comprimento de onda, irá existir uma espécie de “sombra sonora” (FONSECA, 2007).

2.2.4. Frequências das vibrações sonoras

Praticamente todos os sons são constituídos por diversas frequências distintas, e não por uma única. Mesmo uma nota musical tocada por um baixo elétrico pode ter frequências espalhadas por quase toda a gama audível, embora obviamente as frequências mais graves tenham mais intensidade que as mais agudas. Por isso, é praticamente impossível remover um determinado instrumento de uma determinada gravação - como os instrumentos adicionam tantas frequências ao longo da gama audível, não é possível saber exatamente que frequências são de que instrumento (FONSECA, 2007).

Essas frequências múltiplas são denominadas de **Harmônicos**, ALVARENGA (2008, p. 25) afirma que, harmônicos são as frequências múltiplas da frequência fundamental F. Por exemplo, uma onda sonora de frequência igual a 200Hz terá como frequências múltiplas: 400Hz (segundo harmônico); 600 Hz (terceiro harmônico); 800 Hz (quarto harmônico); etc. Existem, então, os harmônicos de ordem ímpar e de ordem par.

Pode-se entender melhor os harmônicos a partir do exemplo a seguir. Imagina-se que uma corda produza o som correspondente à seguinte nota dó, como mostra a figura 24.

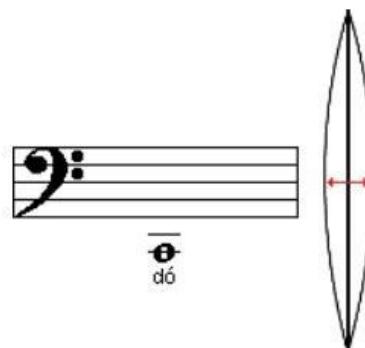


Figura 24 - Nota dó
Fonte: Dirsom– Harmônicos.

Ao mesmo tempo, a corda também produz uma vibração duas vezes mais rápida e que gera um novo som correspondente a um dó uma oitava acima do primeiro, conforme figura 25.

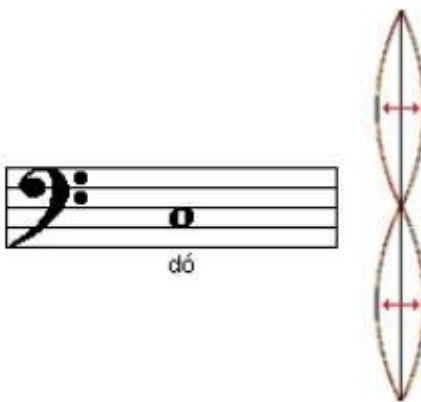


Figura 25 - Oitava de dó
Fonte - Dirsom - Harmônicos.

E ao mesmo tempo, vibra também em 3 terços, 4 quartos, 5 quintos, ..., como mostra a figura 26, e cada uma dessas vibrações secundárias produz um novo som harmônico e assim consecutivamente, até que a corda pare de vibrar.

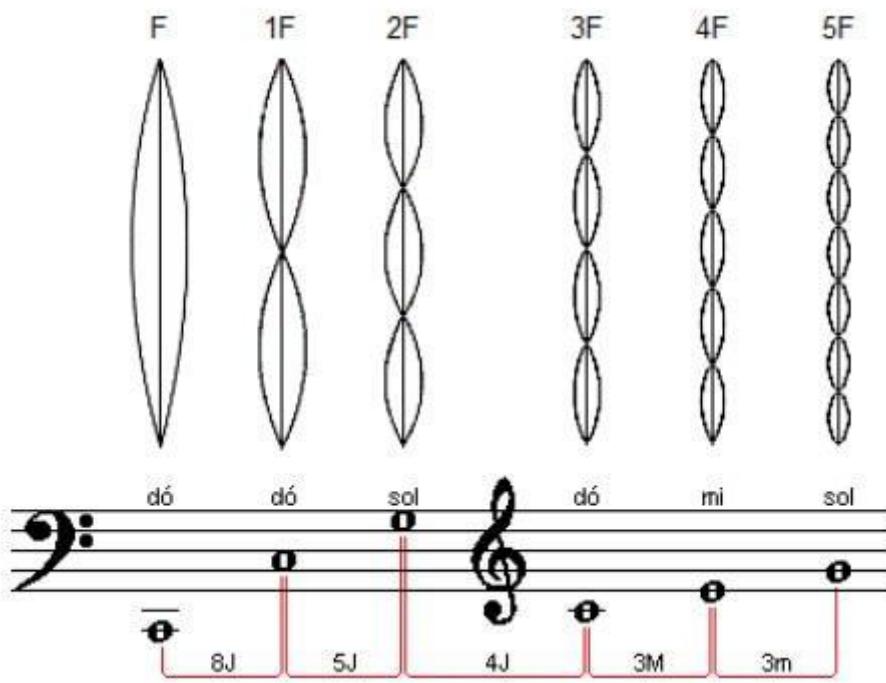


Figura 26 - Série harmônica
Fonte - Dirsom - Harmônicos

Também é com base na frequência fundamental que os músicos atribuem as notas aos diversos sons. Se um som tiver uma frequência fundamental de 440 Hz, então será um “Lá”. Um “Lá” tocado uma oitava acima terá o dobro do valor (880 Hz), e uma oitava abaixo terá metade do valor (220 Hz).

2.2.5. O Timbre

Notas iguais produzidas por instrumentos musicais permitem que se distinga a sua origem, em função do número de harmônicos que possuem. Isso resulta no timbre de cada instrumento. O timbre é o som característico de cada instrumento. O timbre também depende do número, da ordem e da intensidade dos diferentes harmônicos associados ao som fundamental (mas não depende das suas fases). A mistura harmônica com o som fundamental não modifica a altura, somente o timbre (ALVARENGA, 2008). A figura 27 ilustra uma fundamental e seus harmônicos.

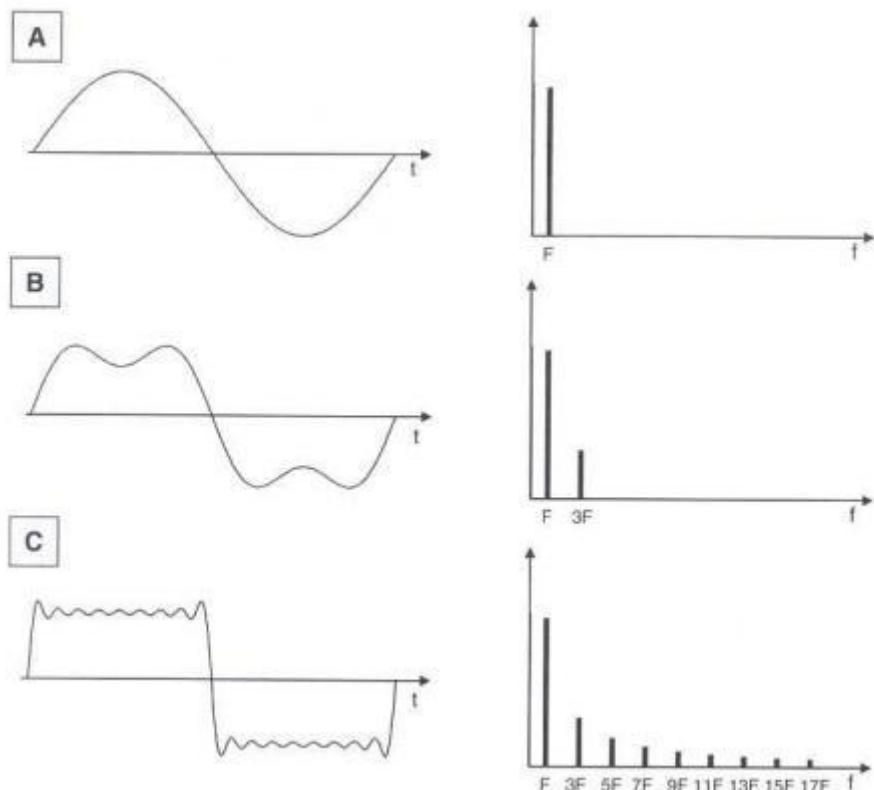


Figura 27 - Diversas ondas e respectivas frequências: (A) apenas frequência fundamental, (B) fundamental e primeiro harmônico e (C) diversos harmônicos.
Fonte: FONSECA (2006, p. 7).

Flávio Adami, em seu artigo *Os Instrumentos e seus Timbres* para revista Audionews, descreve o timbre como “a característica sonora que nos permite distinguir os vários instrumentos entre si”.

Hermann von Helmholtz, no livro “On the Sensations of Tone” formula uma teoria que é fundamento do que hoje sabe-se respeito de timbre, pode-se então caracterizar os sons consistentes de uma onda de forma arbitrária fechada em um *envelope de amplitude* constituídos de três partes: **ataque, período estável e queda.**

O ataque é o tempo que a amplitude de um som leva para sair do zero e subir até seu valor de pico. Para ADAMI, em seu artigo O Timbre e seu Envelope (Audionews52), o ataque é o início de cada nota musical. Dependendo do instrumento, o ataque pode durar alguns centésimos de segundo ou até pode chegar a durar um pouco mais de 1 segundo, sendo este o item mais crítico na reprodução, algo que caracteriza a microdinâmica.

O período estável é o tempo em que a amplitude é idealmente constante. Durante este tempo de sustentação, a intensidade é mantida no mesmo nível.

A queda é o tempo em que a amplitude cai a zero. É o final da nota, quando a intensidade sonora diminui até desaparecer completamente. Pode ser bastante brusco, como acontece em um instrumento de sopro, quando o instrumentista corta o fluxo de ar, ou muito lento, como em um gongo, ou em um piano com o pedal acionado.

A figura 28 exemplifica essas três partes.

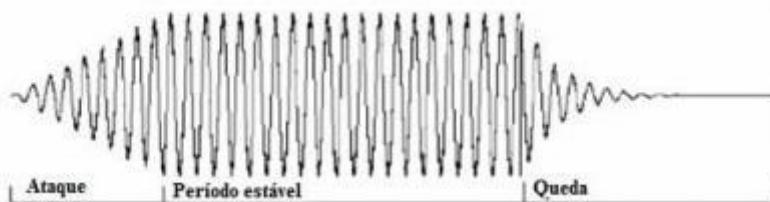


Figura 28 - Representação do ataque, período estável e queda de um som
Fonte: HORIZONTAL e JAMES. (1995)

Diferentes sons têm diferentes envoltórios de amplitude.

2.2.6. Decibel

A unidade do decibel é amplamente utilizada em engenharia de som, muitas vezes de preferência a outras unidades como volts, watts, ou outras tais como unidades absolutas, uma vez que é conveniente representar a razão da amplitude de um sinal para outro e o ouvido diferencia (perceptualmente) a pressão sonora em escala logarítmica. Também resulta em números de tamanho prático, por exemplo, a gama de intensidades sonoras que pode ser manuseado pelo ouvido humano cobre cerca de 14 potências de dez, de 0,00000000001Wm⁻² para cerca de 100Wm⁻², mas o intervalo equivalente em Decibéis é apenas de 0 a 140dB (RUMSEY e McCORMICK, 2006)

O decibel é baseado no logaritmo da razão entre dois números. Pode também ser utilizado como uma unidade absoluta de medição se o valor de referência é conhecido. Algumas referências normalizadas têm sido estabelecidas para decibel para escalas em diversos campos de engenharia acústica.

No mundo do áudio, o decibel é usado de diversas formas possíveis como, por exemplo, para definir a pressão sonora (SPL). Mas existem outras situações, especialmente no que diz respeito a circuitos eletrônicos. Umas dessas situações é quando se pretende medir a amplificação/atenuação de um sinal ao passar por um dispositivo eletrônico:

$$dB = 20 \log \frac{V_{out}}{V_{in}}$$

O que significa que se um circuito colocar à saída o mesmo sinal que existe a entrada, então a fórmula anterior dará o valor de 0 dB (FONSECA, 2008).

Quando se olha para um controle de volume e vê uma escala definida em dB, como mostra a figura 29, tem se indicação da atenuação ou amplificação que irá ocorrer.



Figura 29 - Exemplo de uma escala em dB
Fonte: FONSECA (2008, p. 59).

Se esse circuito colocar à saída o dobro do sinal de entrada, então o circuito terá um comportamento de +6dB, e na situação inversa (se colocar metade do sinal) terá -6 dB. No entanto, se o circuito não colocar nada à saída, a fórmula anterior resultará em $-\infty$ dB.

FONSECA (2008, p. 59) ainda diz que, a utilização do dB não fica por aqui. É também usado para definir níveis elétricos. Devido a um conjunto de razões, nem todas as entradas/saídas dos equipamentos de áudio funcionam com os mesmos níveis elétricos. Por exemplo, o nível elétrico à saída de um microfone é muito inferior ao nível elétrico à saída de um amplificador de potência. Como FONSECA, normalmente dividem-se os sinais de áudio em 3 grandes categorias:

LowLevel- Sinais muito fracos, como microfones e captadores de guitarra;

LineLevel - Sinais médios, usados essencialmente para interligar aparelhos de áudio;

High Level - Sinais de potência, usados essencialmente para ligação de colunas de caixas de som.

Quando se interligam dispositivos de áudio, não se convém ligar uma saída de umas destas categorias a uma entrada de um equipamento de uma

outra categoria. Se o nível de um sinal for muito forte, poderão existir distorções; por outro lado, se o nível do sinal for muito fraco, o ruído de fundo pode aumentar. Ainda, dentro de uma mesma categoria podem haver diferenças. Pode-se encontrar equipamentos a designação de **+4 dBu** ou **-10 dBV**. Embora estes sinais sejam considerados *line levels*, os seus níveis são diferentes.

O dBu e o dBV utilizam o mesmo princípio de comparação e servem para indicar uma tensão com base num valor de referência (0.775 V no caso do dBu e 1V no caso do dBV).

2.3. SISTEMAS DE GRAVAÇÃO E REPRODUÇÃO

2.3.1. MICROFONE

O microfone é um transdutor que converte energia acústica sonora em energia elétrica. Pela lei de Faraday onde uma corrente elétrica é induzida num fio que está movendo-se em um campo magnético, perpendicular às linhas de fluxo entre os pólos do imã. A direção do movimento controla a direção do fluxo de corrente no fio, deste jeito uma corrente alternada é induzida no fio, com frequência e amplitude dependentes do movimento do fio. Ou uma corrente é adicionada no fio de maneira que ele se move pelo campo magnético. Os 3 princípios mais comuns de operação do microfone são de bobina móvel (ou dinâmico), de fita e capacitores (ou condensadores), sendo que apenas os 2 primeiros serão explicados adiante.

2.3.1.1. Microfones de bobina móvel

O microfone de bobina móvel consiste de um diafragma rígido de 20 à 30mm de diâmetro, que fica suspenso na frente do imã. Um molde cilíndrico, onde é feito a bobina por fios de calibre fino, é conectada ao diafragma. Este conjunto fica situado na fenda de fortes imãs permanentes, a figura 30 ilustra o modelo. Quando o diafragma vibra devido às ondas sonoras, ocorre o movimento da bobina de um lado para outro nas fendas dos imãs. Desse jeito uma corrente alternada flui na bobina produzindo a corrente. Alguns modelos têm enrolamentos suficientes na bobina para produzir um sinal diretamente nos terminais de saída, porém outros modelos com menos enrolamentos tem a saída alimentada por um transformador *step-up* no revestimento do microfone (Francis Rumsey, 2009).

Eles são usualmente utilizados na indústria de reforçamento do som, sua robustez é mais adequada para uso portátil. Geralmente eles contêm malhas de

fio e espuma para atenuar barulhos de vento e alguns ruídos provenientes da voz do vocalista.

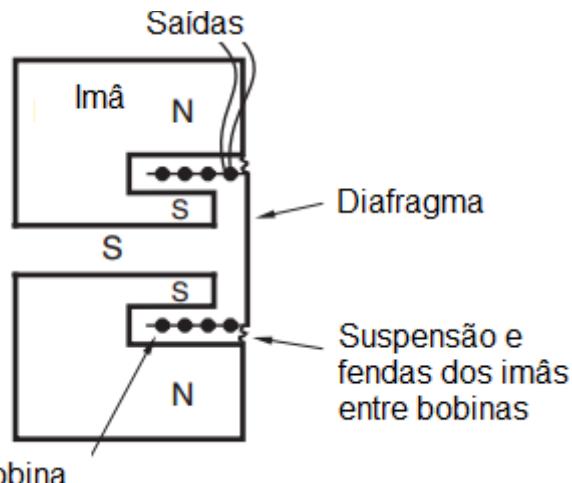


Figura 30 - Princípio de funcionamento do microfone de bobina móvel
Fonte: Francis Rumsey

2.3.1.2. Microfones de fita

O microfone de fita consiste de uma faixa bem fina e longa, laminada com metal condutivo e plissada para dar uma maior rigidez e elasticidade, como se vê na figura 31. Esta faixa fica tensionada entre 2 grampos. Os pólos magnéticos criam campos magnéticos que atravessam a fita, desse jeito ela é excitada por ondas sonoras e então a corrente elétrica é induzida na fita, pela Lei de Faraday. A corrente de saída da fita é muito baixa, portanto é usado um transformador no microfone para aumentar esta corrente.

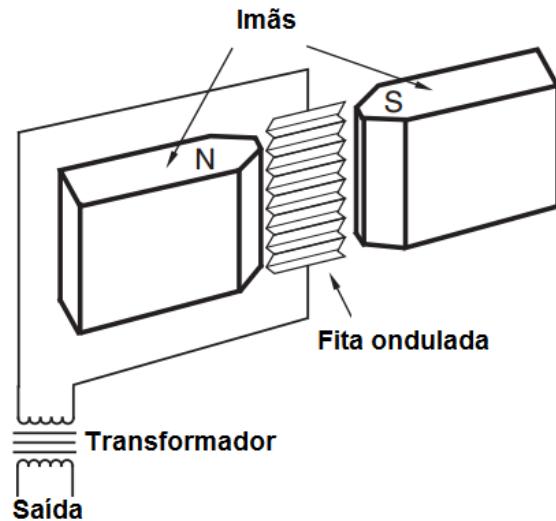


Figura 31 - Funcionamento do microfone de fita
Fonte: Francis Rumsey.

O microfone de fita é mais utilizado para resultados de alta qualidade. A suspensão comparativamente "flexível" da fita dá-lhe uma ressonância de baixa frequência a cerca de 40 Hz, abaixo da qual a sua resposta de frequência cai rapidamente. Na alta frequência a sua resposta de frequência fica mais suave. O microfone de fita é mais delicado do que a bobina móvel, e é mais adequado para aplicações com resposta de frequência suave, como de instrumentos acústicos e conjuntos clássicos (Francis Rumsey, 2009).

2.3.2. ALTO-FALANTES

O alto-falante é um transdutor que converte energia elétrica em energia acústica. Ele é composto de um diafragma que vibra quando energizado para produzir ondas sonoras que serão similares ao som original. Alguns tipos são mais apropriados para algum gênero musical enquanto outros não serão tão adequados quanto.

2.3.2.1 - Alto-falante de bobina móvel

O princípio da bobina móvel é o mais utilizado em alto-falantes; é usado desde alto-falantes de rádio com transistores baratos, até estúdio de alto desempenho. O driver é o componente que produz as ondas sonoras no ar. Basicamente o alto-falante consiste de um imã permanente de forte intensidade entre uma fenda anular para acomodar a bobina de fio em torno do molde cilíndrico. Este molde é conectado ao cone ou diafragma que é segurado por um sistema de suspensão que geralmente consiste de um tecido dopado, nas bordas da base do cone. Nas bases do cone é utilizado um tipo de borracha ou tecido dopado para permitir a movimentação do cone. A figura 32 mostra a seção transversal de um alto falante de bobina móvel.

O revestimento consiste normalmente em aço prensado ou fundido, sendo este último mais desejável quando são utilizados grandes ímãs pesados, uma vez que a distância muito pequena entre a bobina e o espaço do ímã requer uma estrutura rígida para manter o alinhamento e um revestimento de aço prensado pode às vezes ser distorcido se o alto-falante estiver sujeito a ações bruscas.

O cone pode, em princípio, ser feito de praticamente qualquer material: as escolhas mais comuns são, a pasta de papel, plásticos de vários tipos (como usado em alto-falantes *hi-fi* devido à maior consistência do que com a pasta de papel e à coloração potencialmente mais baixa do som, geralmente à custa de aumento de peso e, portanto, menor eficiência que não é crucial em um alto-falante doméstico) e às vezes de folha metálica (Francis Rumsey, 2009).

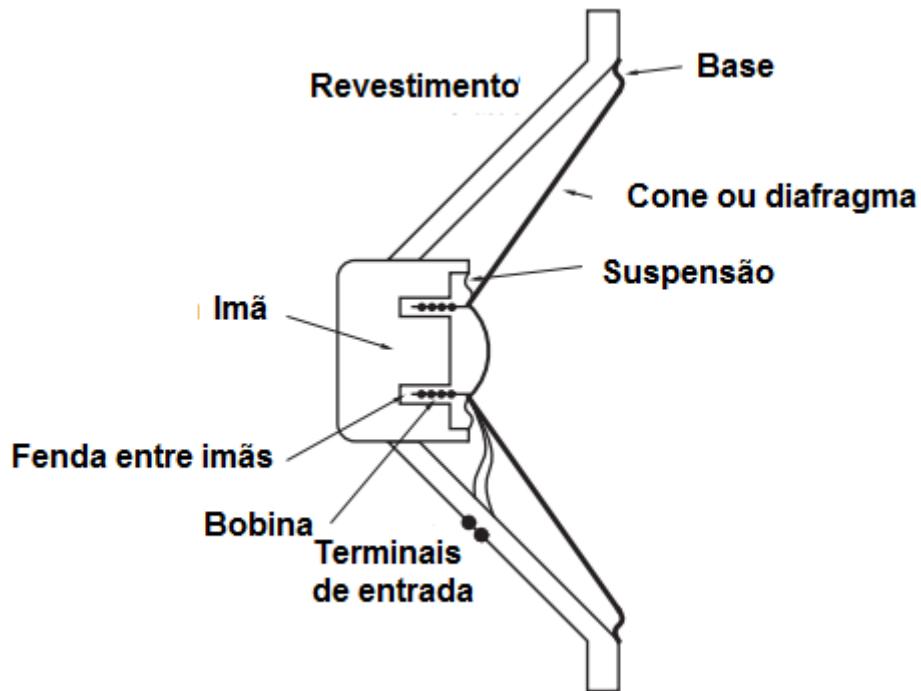


Figura 32 - Seção transversal do alto-falante com bobina móvel
Fonte: Francis Rumsay, 2009

2.3.3. Resposta de Frequência

A resposta de frequência é usada para descrever a frequência dinâmica audível que um alto-falante pode reproduzir ou que um microfone pode produzir, sendo que para cada valor de frequência haverá um valor em ganho (dB). A figura 33 exemplifica uma curva de frequência, no caso, um microfone SHURE 56A.

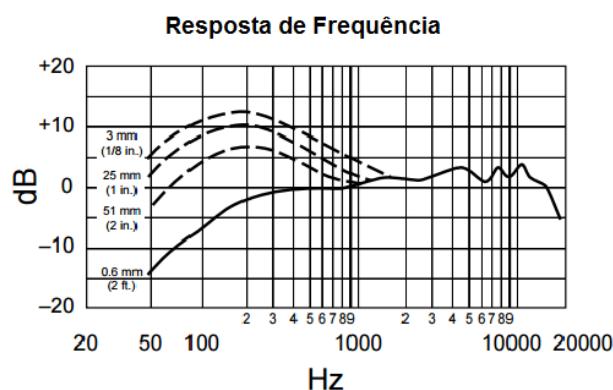


Figura 33 - Resposta de frequência de um microfone SHURE 56A
Fonte: SHURE

2.4. FORMATOS DE ÁUDIO DIGITAL

Som é uma onda mecânica longitudinal que necessita de meios físicos para se propagar, podendo ser captado de diversas maneiras e convertido em sinais elétricos analógicos para transmissão e reprodução. Para realizar o processamento digital desses sinais, como por exemplo gravar músicas em um CD de áudio, se faz necessário a conversão desses sinais analógicos em sinais digitais.

A técnica mais utilizada para conversão de sinais analógicos de áudio para sinais digitais é a modulação por código de pulsos, PCM (*Pulse code modulation*). A técnica PCM foi patenteada, em 1939, pelo Sr. Alec. Reeves, quando era engenheiro da IT&T na França (VARDALAS, 2012)

A modulação por código de pulsos ocorre em três etapas: amostragem, quantização e codificação. O sinal é amostrado em intervalos regulares de tempo, quantizado em níveis finitos de amplitude e então codificado em uma sequência de bits, os pulsos. As duas primeiras etapas exercem impacto direto na fidelidade do sinal digital ao sinal analógico original: A taxa de amostragem, que é a frequência em que o sinal é amostrado, e a profundidade de bit que determina quantos níveis de amplitude possíveis um determinado sinal pode assumir (MOECKE, 2006).

2.4.3. *Red Book* – Formato padrão dos CDs de áudio

O formato padrão dos CDs de áudio (*Compact Disc Digital Audio*) é conhecido como *Red Book*. Este padrão não é gratuito e tem de ser licenciado pela sua detentora Phillips ou através da IEC (*International Electrotechnical Commission*).

Este padrão dita alguns requisitos quanto à qualidade do áudio, são eles:

- 2 canais

- Método de modulação Linear PCM (LPCM), que é o método PCM que já descrevemos, mas com os níveis de quantização linearmente uniformes;
- Taxa de amostragem de 44,1 kHz, assim atendendo com folga o Teorema de Nyquist para garantir a reconstrução do sinal sem perda de informação
- Profundidade de bit de 16-bits, totalizando 65.536 níveis de amplitude que cada amostra pode assumir.

O padrão definido pela IEC 60908 contém vários outros aspectos, como por exemplo, a espessura mínima que um disco deve ter para receber o selo de *Compact Disc Digital Audio*, mas serão analisados os aspectos já citados e que impactam de maneira direta a qualidade do áudio.

Em um CD de áudio não há arquivos como em um CD-ROM ou DVD, existe um fluxo de dados em LPCM e dados de subcanal que servem para controle do leitor de onde começa ou termina as faixas do CD. Se colocar um CD de áudio em um computador, ele exibirá um arquivo para cada música, porém esse é um processo realizado pelo software do sistema operacional.

2.4.4. Formato de arquivo WAV

O formato WAV é um dos formatos mais conhecidos de áudio digital, foi desenvolvido pela Microsoft e pela IBM. O formato de arquivo *.wav pode ser usado para várias aplicações, e em áudio pode conter áudio comprimido, porém é mais comum conter áudio não comprimido PCM. E essa é a sua similaridade com os CDs de áudio, que utilizam o *Red Book Audio* que já apresentado. Portanto um leitor normal de CDs não irá tocar um CD que contém arquivos WAV.

É um formato não muito definido, podendo conter áudio com as mais diversas taxas de amostragem, profundidade de bit, duração e etc. Uma das suas principais limitações é o tamanho do arquivo que não pode exceder 4 gigabytes, devido ao cabeçalho do arquivo ser do tipo *32-bit unsigned integer*. Também é usado para guardar dados que não sejam áudio (WATERS, 2009).

2.4.5. Super Audio CD

Introduzido em 1999, foi desenvolvido em conjunto pela Sony e Philips Electronics, e pretendia ser o sucessor do formato de CD (*Compact Disc*). Ele suporta áudio multicanal (*surround*) e utiliza uma técnica chamada *Direct Digital Stream* (DSD) para conversão do sinal analógico em digital, com uma taxa de amostragem de 2,82MHz, 64 vezes a taxa de amostragem do CD (GEUSTSKENS, 2009).

O formato acabou não se popularizando como o CD ficando voltado ao nicho da comunidade audiófila.

2.5. COMPRESSÃO DIGITAL

2.5.1 Formato de arquivo MP3

Em 1987 uma aliança entre a universidade de Erlangen-Nuremberg e Fraunhofer Institute for Integrated Circuits dentro da estrutura de trabalho da união europeia fundou o “*EUREKA project EU147 for Digital Audio Broadcasting (DAB)*”. Essa equipe conseguiu desenvolver o codec do LC-ATC, que permitia a codificação em tempo real de áudio estéreo.

Já em 1998 o desenvolvimento do algoritmo OCF (*optimum coding in the frequency domain*) foi mais um marco histórico no desenvolvimento do MP3. O OCF já tinha várias características que o MP3 viria a ter. Com algumas melhorias, a estrutura do OCF foi transformada em uma técnica realizável. Segundo Karlheinz Brandenburg: “*O desenvolvimento do OCF em 1998 foi um marco importante porque tornava real a visão da década de 70 de transmitir música através de conexões telefônicas. Pela primeira vez era possível codificar música em boa qualidade a uma taxa de 64 kbit/s para um sinal mono. O OCF foi o começo do caminho para a normatização MPEG.* ”

Em 1989 a *Moving Picture Expert Group* (MPEG), uma organização internacional de normatização, planejava introduzir um padrão de áudio e o OCF foi apresentado. A MPEG recebeu um total de 14 propostas para codificação de áudio, e os participantes foram encorajados a combinar suas contribuições. Disso resultou em 4 potenciais candidatos, dois deles eram baseados no codificador MUSICAM (*Masking-pattern Adapted Universal Subband Integrated Coding and Multiplexing*) do Institute for Broadcasting Technology IRT e da Philips. O codec ASPEC (*Adaptive Spectral Perceptual Entropy Coding*) nasceu de melhorias no algoritmo OCF realizados pelo Fraunhofer IIS em adição às contribuições da University of Hanover, AT&T and Thomson. Depois de incontáveis testes a MPEG propôs de combinar o ASPEC e o MUSICAM e estabelecer uma família de três técnicas de codificação: *Layer 1* seria uma variante de pouca complexidade do MUSICAM, *Layer 2* seria um codificador MUSICAM e a *Layer 3*, que mais tarde seria chamada de MP3 (*MPEG-1 Layer 3*), seria baseada no ASPEC. O desenvolvimento técnico do padrão MPEG-1 foi completado em dezembro de 1991. MPEG-1 *Layer 3* foi padronizado para maiores taxas de amostragem de 32 kHz, 44,1 kHz e 48 kHz em MPEG-1 em 1992.

Em 1995 o nome “MP3” é cunhado. Em uma enquete interna os pesquisadores da Fraunhofer votam por unanimidade para .mp3 como a extensão de arquivo para o MPEG-1 *Layer 3*. No mesmo ano, a Fraunhofer fornece o primeiro codec *Layer 3* baseado em PC como *shareware* (FRAUNHOFER IIS, 2017).

A figura 33 mostra um diagrama de blocos de um codificador MP3:

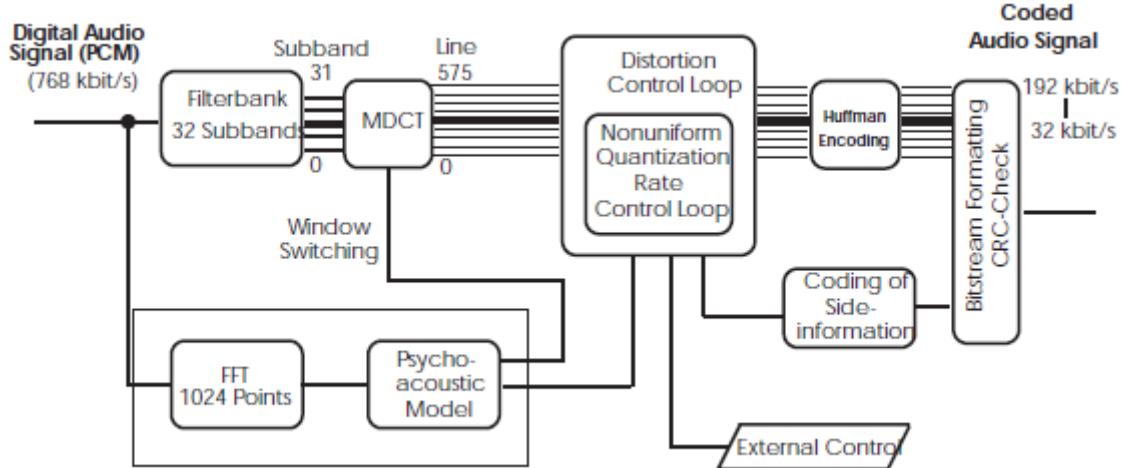


Figura 33 - Diagrama de blocos de um codificador MP3 Fonte: C.-H. Yen et al (2004)

Descrição dos blocos:

Banco de filtros (Filterbank): O banco de filtros usado no MPEG Layer-3 é um banco de filtros híbridos que consiste em um banco de filtros polifásicos e uma Transformada de Cosseno Discreta Modificada (MDCT).

Modelo perceptual (Perceptual Model): Usa um banco de filtros separado ou combina o cálculo de valores de energia (para os cálculos de mascaramento) e o banco de filtros principal. A saída do modelo perceptual consiste em valores para o limiar de mascaramento ou o ruído permitido para cada partição do codificador. Se o ruído de quantização pode ser mantido abaixo do limiar de mascaramento, os resultados de compressão devem ser indistinguíveis do sinal original.

Junção estéreo (Joint stereo): Codificação de junção estéreo se aproveita do fato que ambos canais de um par de canais estéreo contêm informações muito parecidas. Essas redundâncias estereofônicas são exploradas a fim de reduzir a taxa de bits. Junção estéreo é usada em casos onde apenas baixas taxas de bits estão disponíveis, mas deseja-se sinais em estéreo.

Quantização e codificação (Quantization and Coding): A quantificação é feita através de um quantificador de Lei de potência. Desta forma, os valores

maiores são codificados automaticamente com menos precisão e alguma configuração de ruído já está integrada no processo de quantificação. Os valores quantificados são codificados pela codificação de Huffman. Como um método específico para a codificação de entropia, a codificação de Huffman é sem perdas. Isso é chamado de codificação *noiseless* porque nenhum ruído é adicionado ao sinal de áudio. O processo para encontrar o ganho ideal e os fatores de escala para um determinado bloco, taxa de bits e saída do modelo perceptual geralmente é feito por dois laços de iteração aninhados de uma maneira de análise por síntese:

Laço de iteração interna (*Inner iteration loop - rate loop*): As tabelas de código Huffman atribuem palavras de código mais curtas a (mais frequentes) valores quantificados menores. Se o número de bits resultantes da operação de codificação exceder o número de bits disponíveis para codificar um dado bloco de dados, isso pode ser corrigido ajustando o ganho global para resultar em um tamanho de etapa de quantificação maior, levando a valores quantificados menores. Essa operação é repetida com diferentes tamanhos de etapa de quantificação até que a demanda de bits resultante para a codificação de Huffman seja pequena o suficiente. O loop é chamado de loop de taxa porque modifica a taxa global do codificador até ser suficientemente pequeno.

Laço de iteração externa (*Outer iteration loop - noise control/distortion loop*): Para moldar o ruído de quantização de acordo com o limiar de mascaramento, os fatores de escala são aplicados a cada faixa de fator de escala. Os sistemas começam com um fator padrão de 1,0 para cada banda. Se o ruído de quantização em uma determinada faixa exceder o limiar de mascaramento (ruído permitido), conforme fornecido pelo modelo perceptual, o fator de escala para esta banda é ajustado para reduzir o ruído de quantização. Uma vez que para alcançar um menor ruído de quantização precisa-se de um número maior de etapas de quantização e, portanto, uma taxa de bits maior, o ciclo de ajuste de taxa deve ser repetido sempre que novos fatores de escala são usados. Em outras palavras, o loop de taxa está aninhado dentro do loop de controle de ruído. O loop externo (controle de ruído) é executado até que o ruído atual (calculado a partir da diferença dos valores espectrais originais menos os

valores espectrais quantizados) esteja abaixo do limiar de mascaramento para cada faixa de fator de escala (ou seja, faixa crítica) (QUACKENBUSH, 2005).

Outra característica usualmente encontrada em faixa de áudio no formato MP3 é o tipo de codificação relacionada à taxa de bits, que pode ser CBR (*constant bit rate*) ou VBR (*variable bit rate*). No CBR a taxa de bits é constante enquanto no VBR ocorre uma variação na quantidade de informação guardada por segundo, assim, uma taxa de bits maior é utilizada para segmentos mais complexos e uma taxa de bits reduzida é utilizada para aqueles segmentos de menor complexidade, apresentando uma melhor proporção de qualidade por tamanho de arquivo que o CBR. A taxa de bits representa a quantidade de informação ou detalhe que está guardada por segundo numa gravação digital, a seguir uma tabela de parâmetros de taxas de bit relacionadas à qualidade:

- 800 bps — Qualidade mínima para ter uma voz reconhecível.
- 8 kbps — Qualidade de transmissão de voz telefônica.
- 24-32 kbps — Qualidade AM.
- 96–128 kbps — Qualidade FM.
- 160 kbps — Qualidade comparável às fitas K-7 tipo II (cromo), no limite máximo de bias/headroom.
- 192 kbps — Qualidade DAB (Digital Audio Broadcasting).
- 224–320 kbps — Qualidade aproximada à de CD. (Maior taxa de bits que o MP3 suporta)
- 500 kbps a 1 Mbps — Áudio sem qualquer perda de qualidade.
- 1411 kbps — Formato de som PCM, equiparável ao CD "*Compact Disc Digital Audio*".

2.5.2 Formato de Arquivo AAC

O Advanced Audio Coding (AAC) é um padrão de codificação de áudio proprietário para compressão de áudio digital com perdas. Projetado para ser o sucessor do formato MP3, o AAC geralmente consegue uma melhor qualidade de som que o MP3 na mesma taxa de bits.

AAC segue a mesma base de codificação do MP3 (Banco de filtros de resolução de alta frequência, quantização não-uniforme, codificação de Huffman, estrutura de loop de iteração usando análise por síntese), mas apresenta melhorias em comparação ao MP3 em vários detalhes e usa novas ferramentas de codificação para melhorar a qualidade em baixas taxas de bits.

A figura 34 mostra um diagrama de blocos de um codificador AAC:

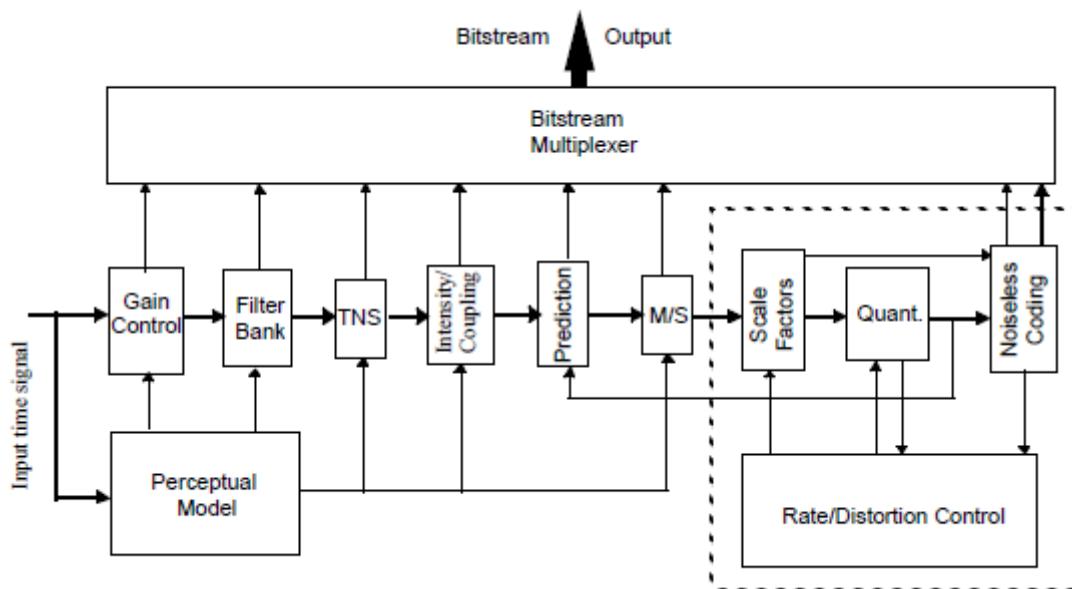


Figura 34 - Diagrama de blocos de um codificador AAC
Fonte: BRANDENBURG (1999).

As seguintes alterações em comparação com o MP3 para obter a mesma qualidade em taxas de bits mais baixas:

Predição (Prediction): Uma previsão inversa opcional, obtém uma melhor eficiência de codificação. Esse recurso está disponível apenas no perfil principal do AAC.

Junção estéreo melhorada: Tanto a codificação do meio/lado quanto a codificação da intensidade são mais flexíveis, permitindo aplicá-las para reduzir a taxa de bits com mais frequência.

Codificação de Huffman aperfeiçoada: Em AAC, o número de linhas de frequência é maior. Além disso, a atribuição de tabelas de código Huffman para partições de codificadores pode ser muito mais flexível.

Existem ainda outras melhorias no AAC que visam manter a qualidade do sinal codificado:

Comutação de blocos aprimorada: Em vez do banco de filtros híbrido (cascata) no MP3, o AAC usa um banco de filtros MDCT comutado padrão com uma resposta ao impulso (para blocos curtos) de 5,3 ms com frequência de amostragem de 48 kHz. Isso se compara favoravelmente com os 18,6 ms do MP3 e reduz a quantidade de ruídos.

Moldagem de ruído temporal - TNS (*Temporal Noise Shaping*): Esta técnica executa a moldagem do ruído no domínio do tempo através de uma predição de laço aberto no domínio da frequência. TNS é uma técnica que prova ser especialmente bem-sucedida na melhoria da qualidade de fala para taxas de bits mais baixas. BRANDENBURG (1999)

No geral, o formato AAC permite, aos desenvolvedores, maior flexibilidade para projetar codecs do que o MP3, e corrige muitas das opções de design feitas na especificação de áudio MPEG-1 original. Essa maior flexibilidade leva a diferentes estratégias de codificação, como resultado, a uma compressão mais eficiente. Isto é especialmente verdadeiro em taxas de bits muito baixas, onde a codificação estéreo superior, MDCT puro e tamanhos de janela de transformação melhores deixam o MP3 incapaz de competir.

Embora o formato MP3 tenha suporte de hardware e software quase universal, principalmente devido ao fato de o MP3 ser o formato de escolha durante os primeiros anos cruciais de compartilhamento de arquivos de música pela internet, o AAC é um forte concorrente devido ao apoio e suporte de alguns

grandes nomes da indústria, por exemplo, o AAC é o formato de áudio padrão do iPhone da Apple.

2.5.3 Formato OGG Vorbis

Ogg Vorbis é um formato de áudio comprimido de uso geral aberto, não proprietário, livre de patente e livre de royalties para áudio e música de mídia a alta qualidade (8kHz-48.0kHz, 16+ bit, polifônico) em taxas de bits fixas e variáveis de 16 a 128 kbps / canal. Isso coloca a Vorbis na mesma classe competitiva que as representações de áudio, como o AAC, e um desempenho similar, mas superior ao MP3.

O desenvolvimento mais intensivo começou depois de uma Carta de Infração, recebida em setembro de 1998 da Fraunhofer Society, anunciando planos para cobrar taxas de licenciamento pelo formato de áudio MP3. Então o código-fonte foi refinado até o formato de arquivo Vorbis ser congelado na versão 1.0 em maio de 2000.

As especificações do Ogg e do Vorbis são de domínio público, dessa forma não existe cobrança monetária por conta da distribuição ou venda de conteúdo em Vorbis (XIPH.ORG FOUNDATION, 2017).

Alguns videogames usam o Vorbis como formato dos sons utilizados *in-game*, e de maior destaque, o serviço de *streaming* de música, Spotify.

3. ANÁLISE SUBJETIVA

A análise subjetiva dos diferentes formatos de arquivos digitais de áudio tem como objetivo mostrar a percepção de qualidade que o indivíduo pode encontrar ao ouvir esses formatos, verificando assim se é possível diferenciar os formatos de áudio e como ele caracteriza a qualidade de cada áudio em relação ao seu formato, ou áudios escolhidos são WAV 16bits com 44.1 kHz de amostragem, MP3 com 128Kbits e MP3 320Kbits.

Esse estudo será dividido entre as etapas do processo de desenvolvimento da metodologia, aplicação e descrição da prática e análise dos resultados.

3.1 Metodologia

3.1.1 Desenvolvimento do questionário

O processo metodológico se inicia na definição do tipo de questionário utilizado. O questionário contém perguntas objetivas e subjetivas abertas, primeiramente, pergunta-se ao entrevistado se ele é capaz de sentir alguma diferença entre os áudios reproduzidos com uma questão dicotômica simples.

As questões seguintes são em relação a qualidade dos áudios reproduzidos, existe então uma questão para cada áudio ouvido em que o entrevistado dará uma nota de 0 a 5, onde 0 significa baixa qualidade e 5 significa alta qualidade. Dessa forma será possível estabelecer numa escala de 0 a 5 qual a nota que o entrevistado classifica cada áudio, possibilitando a comparação entre as notas das amostras.

Por fim, tem-se uma questão subjetiva aberta, em que se busca entender

quais foram os principais aspectos de diferenças encontrados pelo entrevistado. Dessa forma será possível analisar os critérios que cada indivíduo usou para classificar as qualidades dos áudios.

3.1.2 Grupos de amostras

A aplicação do questionário será feita para dois distintos grupos de populações, com o objetivo de se verificar se há divergências nos resultados obtidos quando há uma relação direta entre o entrevistado e o objeto de estudo, o primeiro grupo será composto por indivíduos não músicos já o segundo grupo será composto por músicos, tanto profissionais quanto amantes.

Os grupos serão compostos por indivíduos que se assemelham em outros aspectos, buscou-se convidar somente adultos entre as faixas de 20 a 50 anos, para um estudo específico para esse tipo de população, levando em consideração a influência negativa que o envelhecimento traz ao sistema auditivo. (ALVARENGA, 2008)

3.1.3 Faixas de reprodução

Para aplicação do questionário, é necessário separar e descrever as faixas utilizadas no teste. Usou-se um trecho de 1 minuto de duração da música Cirandeira da cantora Elba Ramalho, gravadora Rio Mix. Escolheu-se essa música pela diversidade de instrumentos usados em sua gravação, mas também pela relação que se faz nesse mesmo trabalho com a parte de análise objetiva que se encontra no capítulo 4.

Neste minuto pode-se ouvir os instrumentos violão, flautas, percussões, cordas e também a voz da cantora, esse grupo de instrumentos mais voz é bem

apropriado para o tipo de estudo pois há uma grande cobertura de frequências que fazem parte dos estudos, além disso, a diversidade de instrumentos traz diversidade de timbres que também tem importante influencia na percepção de qualidade.

3.2 Aplicação e Descrição da Prática

Para compreensão e análise dos resultados dos questionários se faz necessário descrever todo o processo de aplicação do teste e equipamentos utilizados.

3.2.1 Equipamentos

O Conjunto de equipamentos, ou *setup*, são parte fundamento e que traz maior relevância ao estudo subjetivo pois existe um grande diferencial em sua concepção, esses equipamentos têm como objetivo reproduzir os áudios com alta-fidelidade, ou seja, o mais próximo possível do que se foi gravado. Dessa forma, pode-se considerar que as diferenças entre as faixas ouvidas se darão principalmente em relação ao formato de arquivo executado, podendo assim desprezar diferenças em função de ruídos não desejados oriundos de mal funcionamento dos equipamentos.

Para o primeiro dia, foram usados os seguintes equipamentos.

- Amplificador integrado CurrentA da Etalon Acoustics.
- Caixas acústicas Dynaudio Special 25.
- Computador Asus Eee PC – 2GB de RAM
- Conversor Digital Analógico Solo da Etalon Acoustics.

A ligação entre os aparelhos ocorria na seguinte ordem: o arquivo digital armazenado no computador é enviado por cabo USB para o conversor digital analógico que faz a conversão do arquivo para uma saída analógica que por

cabos RCA vai do conversor D/A até a entrada RCA do amplificador integrado. Este amplifica o sinal, elevando a corrente alimenta as caixas acústicas e nelas ocorre a conversão de sinal elétrico, em corrente, em ondas sonoras que atingem o ouvido do audiente. É este sujeito que será convidado a identificar diferenças auditivas entre os vários formatos digitais tocados pelo sistema descrito anteriormente. Deve-se ressaltar que o computador não é um equipamento dedicado à reprodução sonora de alta qualidade.

E no segundo dia, foram os mesmos amplificador integrado CurrentA, o conversor digital analógico Solo, as caixas acústicas Special 25 com a diferença que a fonte sonora foi um transporte The Signature Data CD da MSB Technology e uma fonte dedicada Signature Transport Power Base. E as mídias digitais foram gravadas em um CD. Neste segundo caso, para audição com os músicos, buscou-se apresentar as músicas em um sistema completo de Alta Fidelidade.

3.2.2 Aplicação do Teste

Para aplicação do teste de forma mais eficaz foi necessária uma visita ao estúdio onde o teste seria aplicado. Nesta visita inicial pode-se avaliar quais seriam os melhores lugares para os participantes ficarem durante a reprodução dos áudios, o tempo de duração total do teste e a quantidade de entrevistado em cada grupo de aplicação.

Os locais em que os participantes deveriam ficar foram estabelecidos em função das principais propriedades dos equipamentos e propriedades do alcance do som, esses locais foram definidos com o auxílio do Prof. Dr. Antônio Carlos Pinho, que gentilmente cedeu o seu estúdio pessoal para aplicação dos testes e orientou toda parte prática de aplicação dos testes. Os participantes foram dispostos de acordo com o esquema da figura 35.

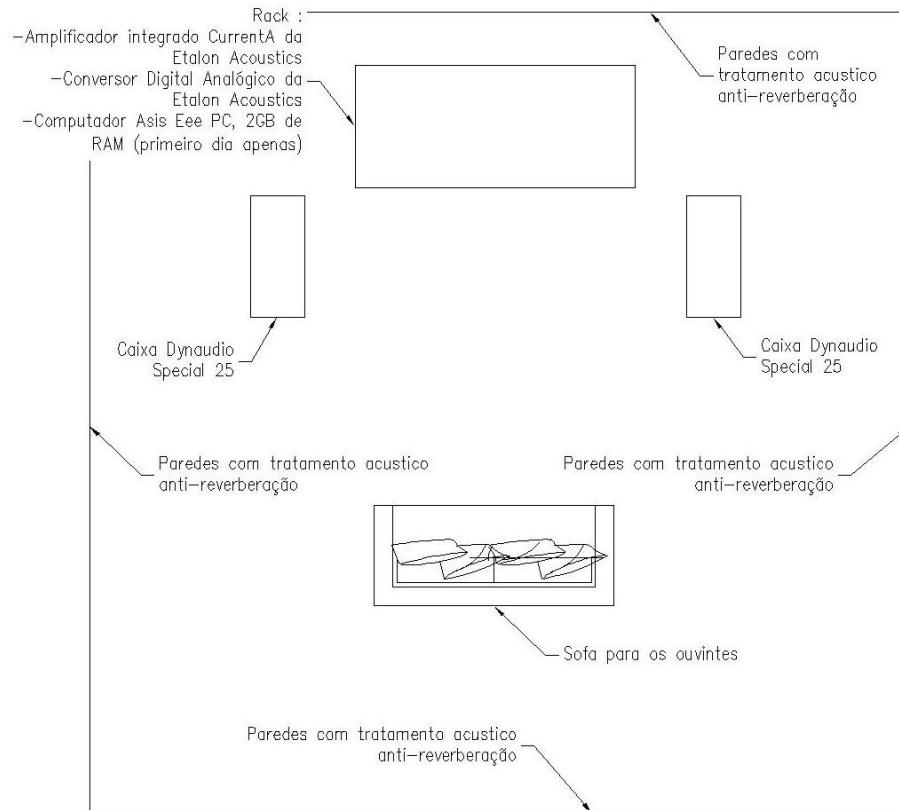


Figura 35 - Disposição da sala
Fonte: Própria

3.2.2.1 Grupo não músicos

No dia 30 de setembro de 2017 foi realizada aplicação do teste auditivo com o grupo de não músicos, no total 10 pessoas participaram dessa prática. Os entrevistados foram divididos em 3 grupos para sessões de aproximadamente 10 minutos. Nos primeiros 4 minutos foram realizadas as reproduções dos trechos dos diferentes formatos, no restante do tempo, cada entrevistado respondeu ao questionário conforme as orientações iniciais ao teste.

Os questionários foram recolhidos e verificados para possíveis erros de preenchimento, agrupados para análises posteriores.

3.2.2.2 Grupo de músicos

Com o intuito de se manter parâmetros semelhantes para comparação de resultados, os procedimentos de aplicação do teste foram muito semelhantes aos procedimentos com o grupo de não músicos.

No dia 2 de dezembro de 2017 foi realizada a aplicação do teste auditivo com o grupo de músicos, com o total de 7 músicos. Os participantes foram divididos em duas sessões de aproximadamente 15 minutos. Os primeiros 5 minutos os músicos escutavam ao sistema de som para se acostumarem ao tipo de reprodução que estariam prestes a comparar, criando assim um parâmetro mais claro do que seria uma boa reprodução. Após esses minutos iniciais, a reprodução das faixas foi iniciada, à medida que as faixas eram reproduzidas os participantes respondiam ao questionário. Ao final, os entrevistados usavam alguns minutos para responder ao último questionamento.

Os questionários foram recolhidos e verificados para possíveis erros de preenchimento, agrupados para análises posteriores.

3.3 Resultados

Para facilitar a análise, os resultados serão separados para os dois grupos entrevistados, e a comparação será feita ao final. Além dessa separação, serão feitos comentários para a sessão de perguntas objetivas e seus resultados bem como da pergunta subjetiva.

Os questionários foram lidos e as respostas das perguntas objetivas foram dispostas em forma de tabelas. Nas tabelas 2 e 3, onde n é o número do participante e as notas fornecidas para a audição dos 4 formatos de áudio, wav, MP3 com 128 kbits, com 320 kbit/s e formato OGG, formato do Spotify de alta qualidade com tamanho de arquivo semelhante ao do MP3 320 kbits.

Tabela 2 - Notas para qualidade, grupo não músicos.

n	wav	128	320	ogg
1	3	4	4	5
2	4	5	5	5
3	5	5	4	4
4	3	4	4	5
5	3	5	4	3
6	4	2	3	1
7	3	4	3	2
8	2	5	4	4
9	4	5	1	2
10	4	2	4	3
Soma	35	41	36	34

Tabela 3 - Notas para qualidade, grupo músicos.

n	wav	128	320	ogg
1	3	4	3	3
2	4	3	3	5
3	1	0	5	4
4	4	2	3	5
5	5	3	4	4
6	5	3	4	5
7	5	2	3	4
Soma	27	17	25	30

Aplicou-se então o Teste de Friedman para análise. O teste de Friedman propõe uma comparação entre as médias de vários conjuntos de amostras, no nosso caso, conjunto de notas para diferentes formatos digitais de áudio, sendo possível verificar se a diferença entre as médias foi causada pela aleatoriedade dos eventos ou se elas consistem em diferenças oriundas das percepções das populações.

3.3.1 Análise dos resultados do grupo de Não Músicos

Atribuir postos às notas de cada linha. A mais baixa nota de cada linha atribui-se o posto 1, ao seguinte em cada linha o posto 2, etc. Se existir porem a possibilidade de empates entre as notas, nesse caso, atribui-se a média

aritmética dos postos para cada um dos empates. A substituição dos valores empatados pelo seu valor médio não afeta a validade do teste. (Prof. Lorí Viali, Dr., 2008)

Tabela 4 - Postos das qualidades, grupo não músicos

n	wav	128	320	ogg
1	1	2,5	2,5	4
2	1	3	3	3
3	3,5	3,5	1,5	1,5
4	1	2,5	2,5	4
5	1,5	4	3	1,5
6	4	2	3	1
7	2,5	4	2,5	1
8	1	4	2,5	2,5
9	3	4	1	2
10	3,5	1	3,5	2
R	22	30,5	25	22,5

Se a hipótese nula é de fato, verdadeira, então a distribuição de postos em cada coluna será aleatória, sendo então de se esperar que os postos 1, 2, 3 e 4 apareçam em todas as colunas com frequências aproximadamente iguais. Isso indica que a condição com os menores escores é uma questão de acaso. Se os escores fossem realmente dependentes da condição, isto é, se a hipótese de nulidade for falsa, então os totais de postos variam de uma coluna para outra. (Prof. Lorí Viali, Dr., 2008)

A prova de Friedman determina se as somas de todos os postos de cada condição diferem significativamente, usou-se então a variável R para representar as somas das notas de cada condição.

Calcula-se o valor de X^2 , que deve ter uma distribuição aproximadamente qui-quadrado, com grau de liberdade ($k-1$), sendo:

$$X^2 = \frac{12}{nk(k+1)} \sum R^2 - 3n(k+1), \text{ onde}$$

n = número de linhas,

k = número de colunas,

R = Soma dos postos da coluna.

Realizando os cálculos obteve-se $X_{calc1}^2 = 2,73$ para o grupo de não músicos.

Para comparar o resultado, buscou-se a aproximação da distribuição qui-quadrado para confiança de 0,05 e G.L. de 3. Obteve-se o valor de referência através da tabela 5.

Tabela 5 - Distribuição Qui-quadrado

ν	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.25}$	$\chi^2_{.50}$	$\chi^2_{.75}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$	$\chi^2_{.999}$
1	,0000	,0002	,0010	,0039	,0158	,102	,455	,1,32	,2,71	,3,84	,5,02	,6,63	,7,88	,10,8
2	,0100	,0201	,0506	,103	,211	,575	,1,39	,2,77	,4,61	,5,99	,7,38	,9,21	,10,6	,13,8
3	,0717	,115	,216	,352	,584	1,21	,2,37	,4,11	,6,25	,7,81	,9,35	,11,3	,12,8	,16,3
4	,207	,297	,484	,711	1,06	,9,92	,3,36	,5,39	,7,78	,9,49	,11,1	,13,3	,14,9	,18,5
5	,412	,554	,831	1,15	,61	,2,67	,4,35	,6,63	,9,24	,11,1	,12,8	,15,1	,16,7	,20,5
6	,676	,872	1,24	1,64	,2,20	,3,45	,5,35	,7,84	,10,6	,12,6	,14,4	,16,8	,18,5	,22,5
7	,989	1,24	1,69	2,17	,2,88	,4,25	,6,35	,9,04	,12,0	,14,1	,16,0	,18,5	,20,3	,24,3
8	1,34	1,65	2,18	2,78	,3,49	,5,07	,7,34	,10,2	,13,4	,15,5	,17,5	,20,1	,22,0	,26,1
9	1,73	2,09	2,70	3,33	,4,17	,5,90	,8,34	,11,4	,14,7	,16,9	,19,0	,21,7	,23,6	,27,9
10	2,16	2,56	3,25	3,94	,4,87	,6,74	,9,34	,12,5	,16,0	,18,3	,20,5	,23,2	,25,2	,29,6

Fonte: PEARSON e HARTLEY, 1966.

Segundo a tabela, o valor crítico é $X_n^2 = 7,81$.

$$X_{calc1}^2 < X_n^2$$

Conclui-se que a hipótese de nulidade não deve ser descartada, ou seja, não existe diferença estatisticamente significativa para os conjuntos de amostras do grupo de não músicos.

Por fim, é possível analisar as respostas a pergunta aberta, em que se queria saber em quais aspectos os entrevistados notaram as diferenças. É possível notar ao ler as respostas da última pergunta que os entrevistados têm observações bastante diversificadas em relação a cada faixa apresentada, as respostas dos questionários estão à disposição do leitor no primeiro apêndice deste trabalho, os nomes dos participantes foram retirados para manter o sigilo de identificação dos entrevistados.

3.3.1 Análise dos resultados do grupo de Músicos

A análise estatística feita para os resultados do grupo de músicos foi muito

semelhante ao grupo anterior. Da mesma forma, foi aplicado o teste de Friedman para comparação das médias de cada conjunto de amostras. Da mesma forma, nossa hipótese de nulidade será de que não existe diferença entre as médias das amostras.

Após atribuir os postos tem-se a tabela abaixo.

Tabela 6 - Postos das qualidades, grupo músicos

n	wav	128	320	ogg
1	2	4	2	2
2	3	1,5	1,5	4
3	2	1	4	3
4	3	1	2	4
5	4	1	2,5	2,5
6	3,5	1	2	3,5
7	4	1	2	3
R	21,5	10,5	16	22

A prova de Friedman determina se as somas de todos os postos de cada condição diferem significativamente, usou-se então a variável R para representar as somas das notas de cada condição.

Calcula-se o valor de X^2 , que deve ter uma distribuição aproximadamente qui-quadrado, com grau de liberdade ($k-1$), sendo:

$$X^2 = \frac{12}{nk(k+1)} \sum R^2 - 3n(k+1), \text{ onde}$$

n = número de linhas,

k = número de colunas,

R = Soma dos postos da coluna.

Realizando os cálculos obteve-se $X_{calc2}^2 = 7,50$ para o grupo de músicos.

Para comparar o valor calculado, usou-se a aproximação da distribuição qui-quadrado para confiança de 0,05 e G.L. de 3. Segundo a distribuição, o valor crítico é $X_n^2 = 7,81$.

$$X_{calc2}^2 < X_n^2$$

Dessa forma, não se pode descartar a hipótese de nulidade. Entretanto, é possível notar uma grande aproximação do valor calculado ao valor crítico tabelar, o que indica uma melhor compreensão das diferenças de qualidade por parte deste grupo.

Pode-se considerar também as respostas subjetivas ao último questionamento, onde as diferenças apontadas aparecem de forma mais clara, mostrando a familiaridade do grupo com o objeto de estudo. Em geral, as diferenças apontadas estão relacionadas a pequenas sutilezas percebidas em algumas faixas, e que não apareciam em outras.

4. ANÁLISE OBJETIVA

4.1. METODOLOGIA

Para a análise objetiva foi utilizado o Matlab para análise dos sinais do áudio em diferentes formatos. Desse jeito foi necessário estabelecer um formato como referência para comparação com os outros formatos, sendo assim foi escolhido o WAV (32-bit float) por ser o formato sem compactação, ou seja, sem perdas de conversões. Para a conversão dos formatos digitais foi utilizado o software Audacity, sendo que as conversões foram todas feitas a partir do formato WAV (32-bit float).

Primeiramente foi analisado a amplitude dos formatos de áudio à serem analisados, o algoritmo é mostrado nos anexos A1 e A2. Em seguida verificou-se a diferença da amplitude do formato WAV em relação aos outros formatos de áudio, MP3 de 128 kbps, MP3 de 320 kbps e OGG. Após as diferenças, é analisado o espectro do áudio em todos formatos e também calculadas as diferenças do espectro de áudio em formato WAV com os outros formatos de áudio. Finalmente é obtido o spectrograma do áudio estudado.

Foram escolhidos trechos da música Cirandeira de Elba Ramalho para a análise objetiva, com a condição de utilizar-se de sinais periódicos, todos com o mesmo tempo definido de 0,024 s. Desse jeito foi definido os seguintes instrumentos à serem analisados: violão, vocal, flauta, violino e trombete. Para a obtenção dos trechos, é ampliado o sinal de áudio no software Audacity e em seguida a velocidade de reprodução é reduzida para isolar o som do instrumento desejado.

Por último será feito o spectrograma da faixa completa. Para a verificação dos padrões de amplitude, frequência e energia no áudio.

4.1.2 ANÁLISE DA AMPLITUDE

Para obtenção das amplitudes dos diferentes formatos de áudio, são

obtidas as amostras do áudio no Matlab em forma de vetor, e também do tempo do áudio criando outro vetor pela seguinte equação:

$$[t_{\text{áudio}}] = [\frac{1}{F_S}, \frac{2}{F_S}, \dots, \frac{n}{F_S}]$$

Sendo n o tamanho do vetor. Para todas análises foi utilizado o canal 1 do áudio, já que o áudio é estéreo, pois será feita a comparação dos resultados com a mesma fonte.

Com os vetores de amostra e tempo, é feito o gráfico com os intervalos definidos desde a primeira amostra até n, para obter-se os gráficos das seguintes figuras 36, 37, 38 e 39.

As amplitudes dos áudios são normalizadas, sendo -1 e +1 os valores máximos de tensão que podem ser reproduzidas.

Foi observado que os formatos em MP3, apresentavam tempos diferentes, em relação ao formato WAV, representados na figura 37 e figura 38 com ruídos criados antes e depois do áudio WAV original. O formato OGG não apresentou diferenças no tempo em relação ao formato WAV.

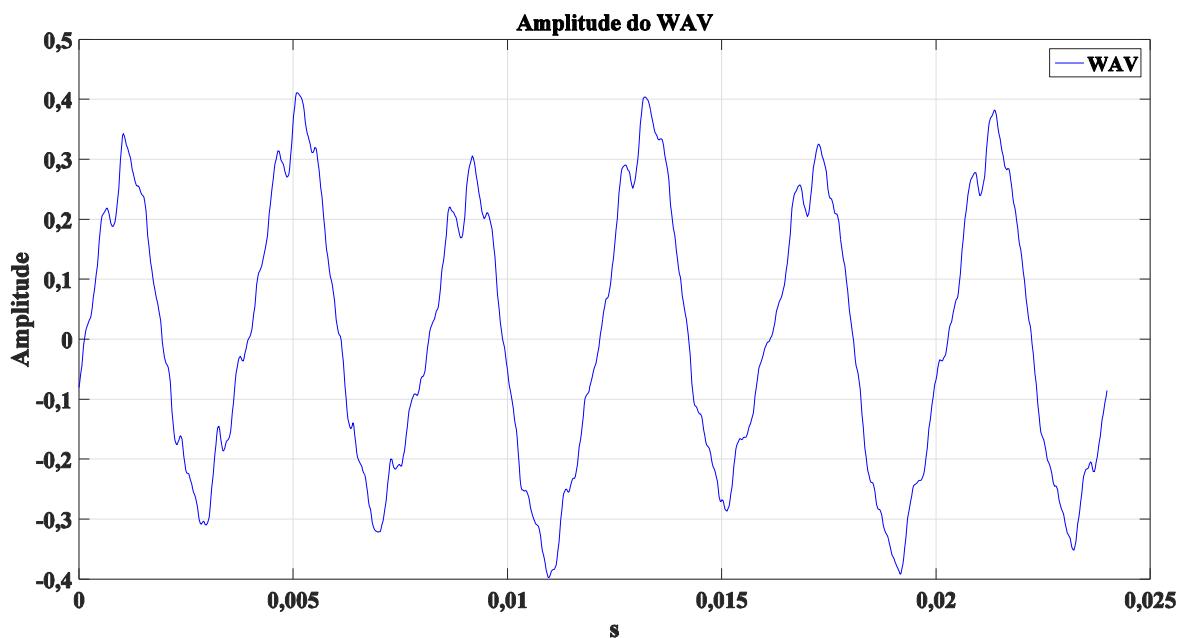


Figura 36 - Amplitude do trecho de violão no formato WAV.
Fonte: Própria.

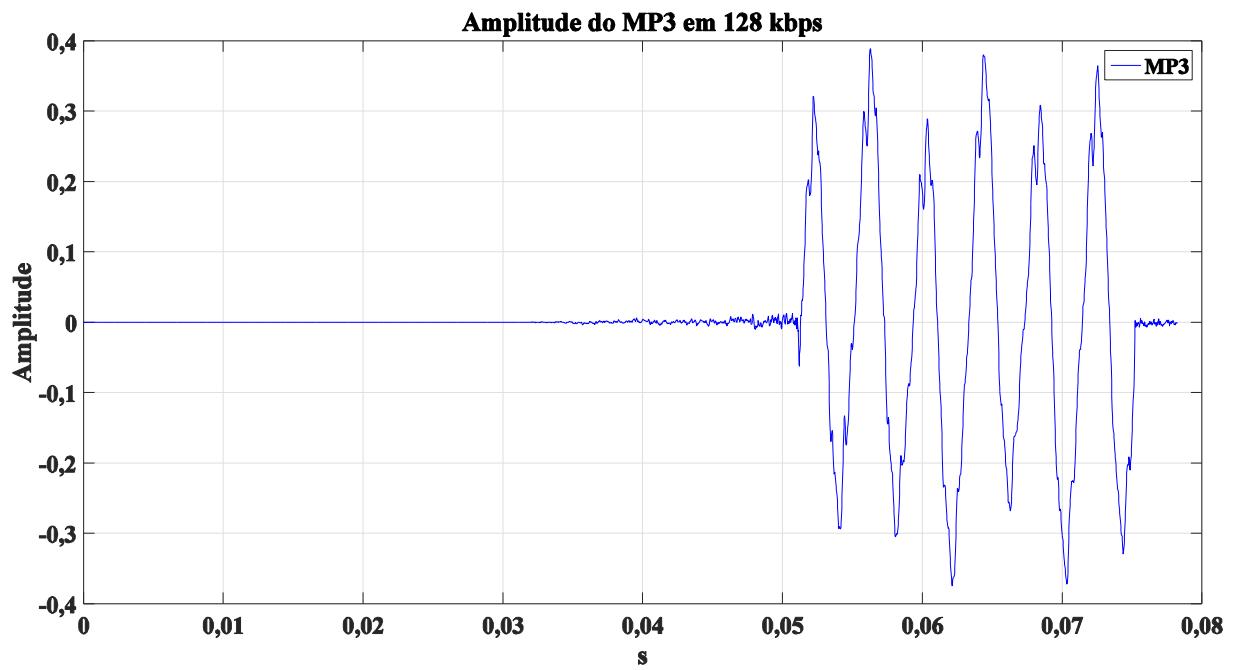


Figura 37 - Amplitude do trecho de violão no formato MP3 - 128 kbps.
Fonte: Própria.

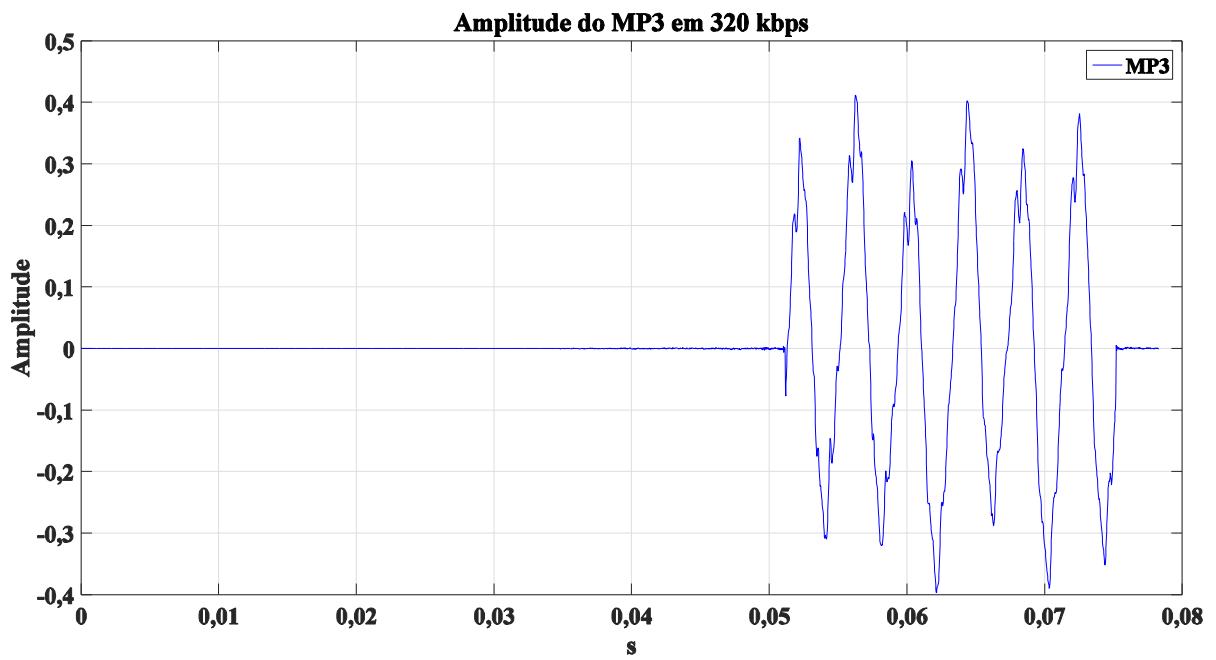


Figura 38 - Amplitude do trecho de violão no formato em MP3 - 320 kbps.
Fonte: Própria.

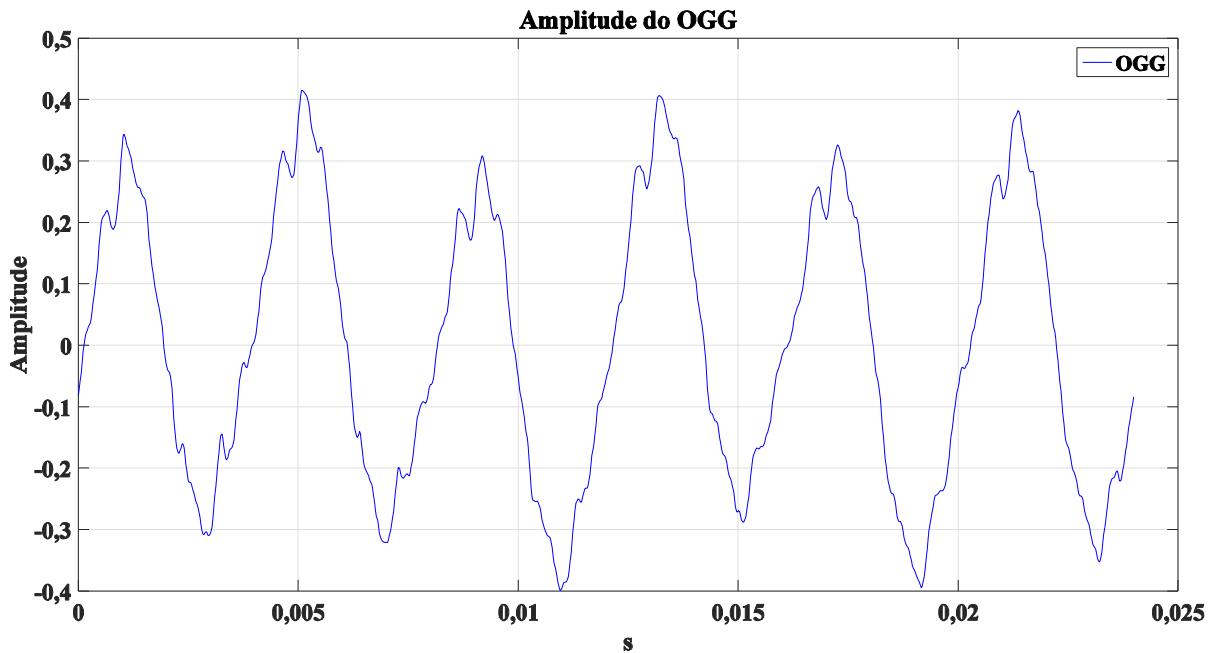


Figura 39 - Amplitude do trecho de violão no formato em OGG.
Fonte: Própria.

A primeira amostra do áudio em formato WAV é negativa, assim é pego o primeiro valor de pico negativo para o formato MP3, esta amostra se encontra no tempo de 0,0512 segundos tanto no formato MP3 de 128 kbps e 320 kbps como demonstrado na figura 40 e 41.

Portanto foi concluído que são adicionados ruídos até o tempo de 0,0512 s e também após 0,0752 s devido a conversão para formato MP3 (em ambos kbps). Desse jeito pode-se afirmar que os áudios convertidos têm tempos maiores em relação ao formato de referência WAV. Foram obtidas 3451 amostras para o formato MP3 em 128 kbps e 3454 amostras para o formato MP3 em 320 kbps.

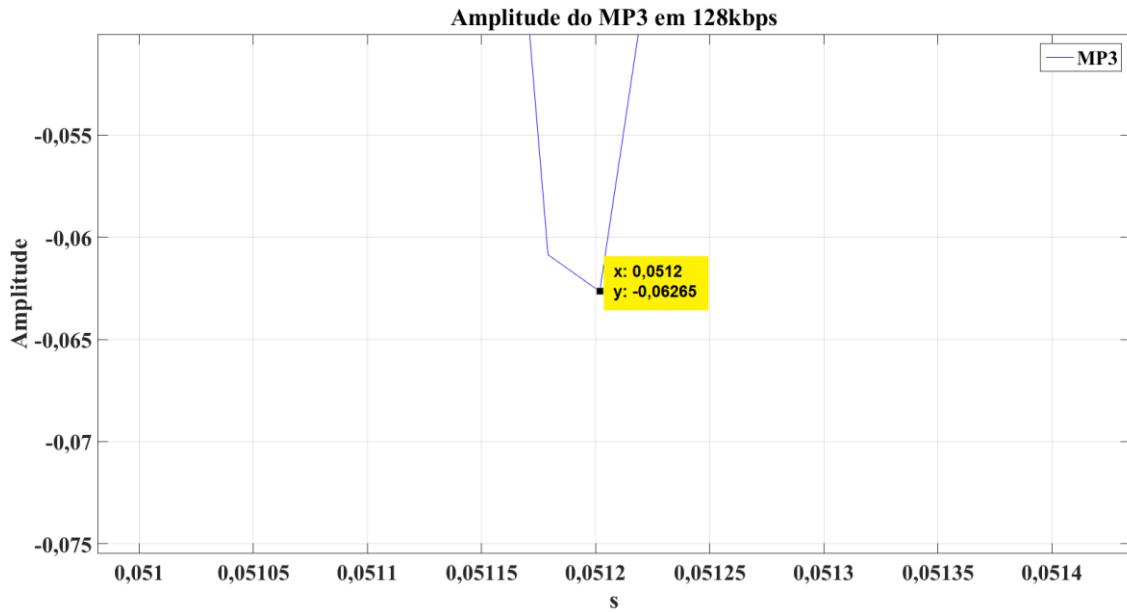


Figura 40 - Ponto exato do primeiro pico negativo do formato MP3 - 128 kbps do trecho de violão.
Fonte: Própria.

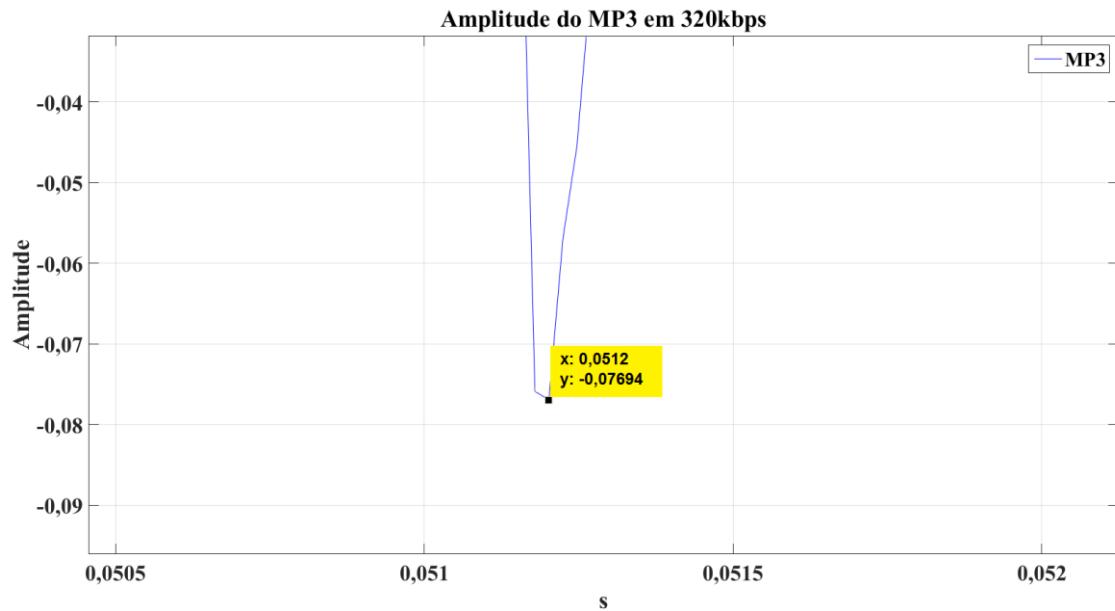


Figura 41 - Ponto exato do primeiro pico negativo do formato MP3 - 320 kbps do trecho de violão.
Fonte: Própria.

Apesar dos atrasos nos formatos MP3 serem iguais, o tamanho de ambos era diferente. O tempo final após a conversão do formato em MP3 de 128kbps era de 0,0782 segundos enquanto que o tempo da conversão do MP3 de 320kbps era de 0,0783 segundos. Estes atrasos foram identificados para todos trechos analisados em formato MP3.

Para se obter a amostra exata que será utilizado para o gráfico no Matlab, o instante do atraso que é de 0,0512 segundos é multiplicado pela taxa de amostragem, resultando na posição do vetor aonde ocorre o atraso. Desse jeito seu intervalo é escolhido a partir do atraso, sendo que sua duração deve ser a mesma da faixa analisada, pois não é contabilizado os ruídos após 0,0752 segundos.

Após a modificação dos gráficos dos áudios em MP3 no mesmo eixo do tempo, é possível uni-las com os gráficos dos outros formatos de áudio como mostrado na figura 42.

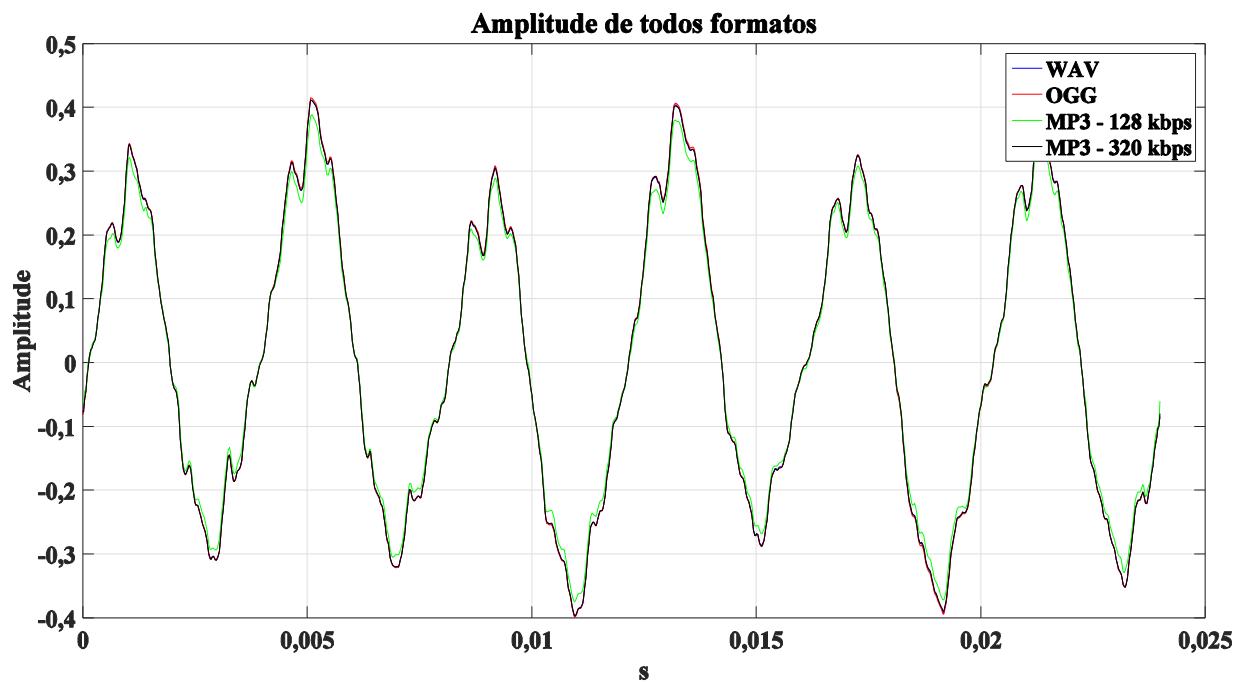


Figura 42 - Trecho com violão de todos formatos de áudio.
Fonte: Própria.

Amplificando o gráfico, é observado que o formato em MP3 de 128 kbps possui a maior diferença em relação ao formato WAV, como mostrado na figura 43.

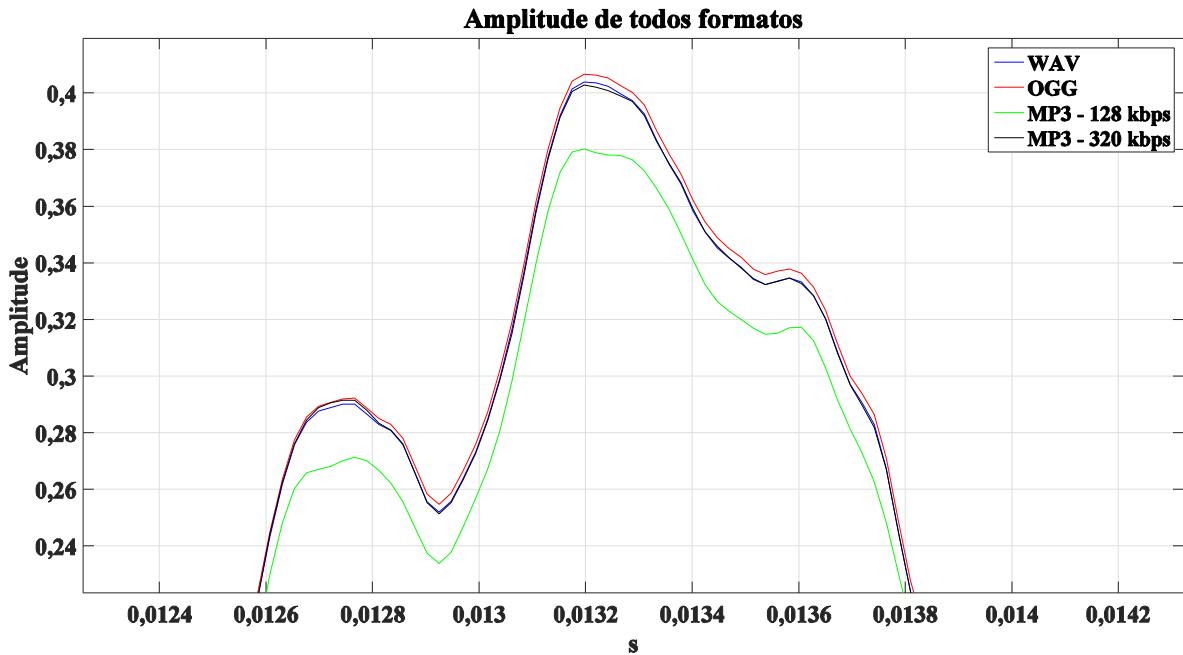


Figura 43 - Gráfico amplificado de todas amplitudes.
Fonte: Própria

4.1.3 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE

Em seguida deseja-se quantificar as diferenças de cada formato de áudio em relação ao formato WAV. Portanto é feito a subtração da amplitude do formato em WAV com as outras amplitudes e então multiplicado por 100 para resultar em porcentagens. Variações positivas significam que a amplitude do formato WAV é maior do que formato à ser analisado, enquanto que variações negativas significam que a amplitude do formato WAV é menor. Quanto maior a magnitude da porcentagem maior é a diferença entre as amplitudes.

Notamos que as principais diferenças estão situadas justamente nos instantes de picos máximos e mínimos, notavelmente nos formatos de MP3 - 128 kbps e OGG como mostrado na figura 44.

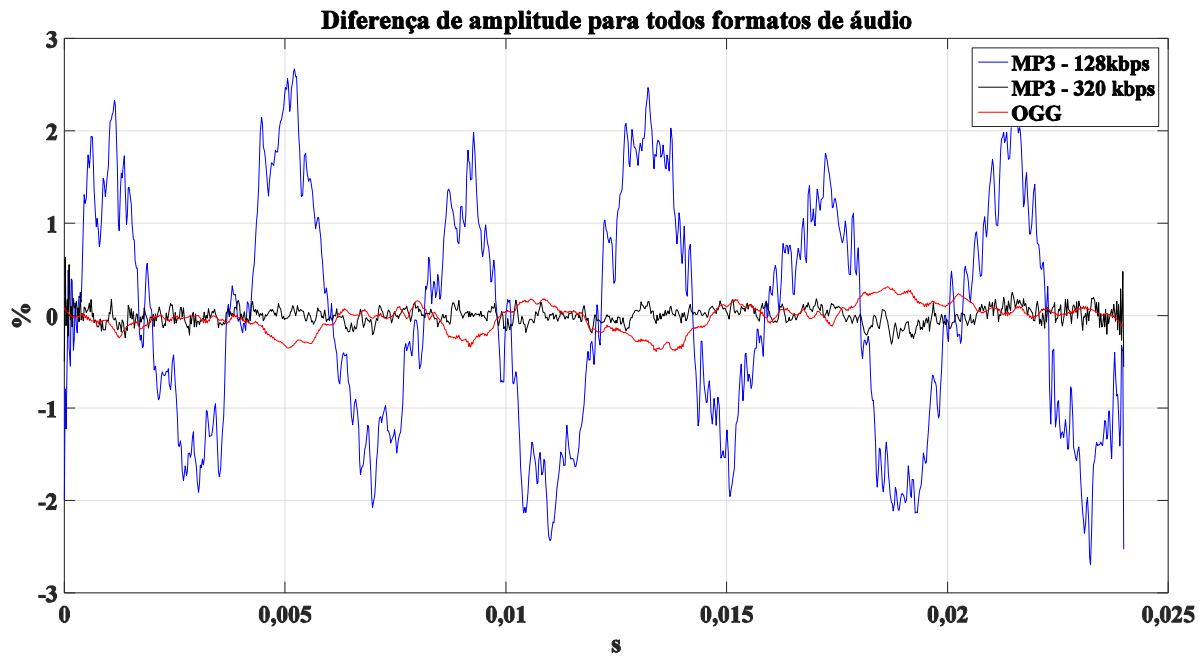


Figura 44 - Diferenças de amplitudes do trecho de violão em WAV em relação aos formatos MP3 - 128 kbps, MP3 - 320 kbps e OGG.
Fonte: Própria

4.1.4 ANÁLISE DOS ESPECTROS

Após a análise das diferenças, foi feita a análise em espectro de frequência dos formatos de áudio. Foi criada uma janela de Hann, normalizando o número de amostras, para assim ser aplicada a Transformada de Fourier e então obter-se o espectro dos formatos de áudio a serem analisados. Assim obtemos a figura 45, aonde é possível notar que as diferenças mais significativas ocorrem após 10.000 Hz.

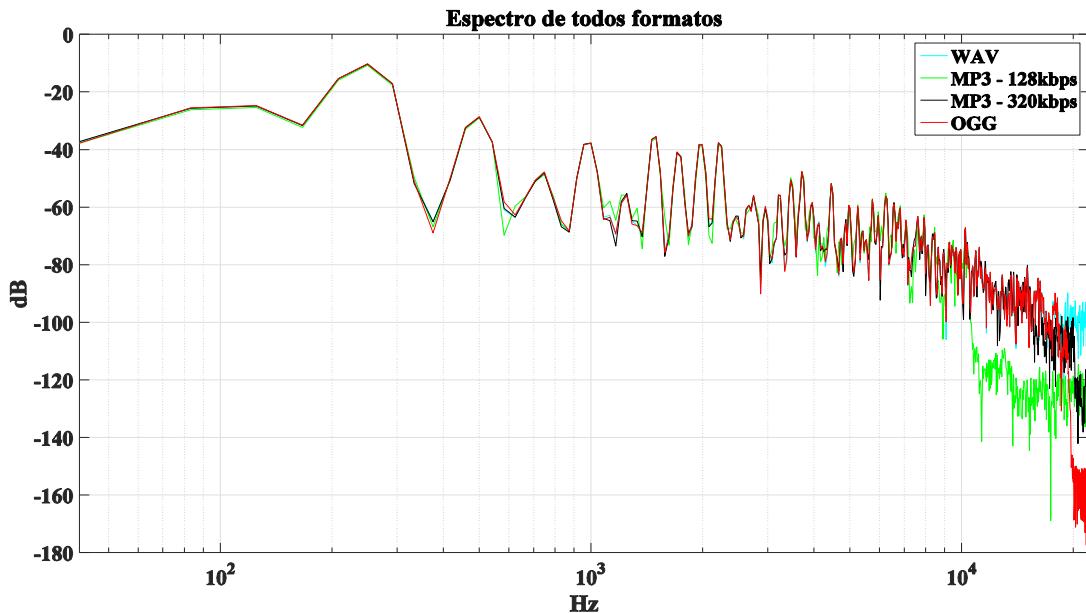


Figura 45 - Espectro de todos formatos de áudio para o trecho de violão.
Fonte: Própria

4.1.5 ANÁLISE DAS DIFERENÇAS DE ESPECTROS

Foi calculado também as diferenças de espectro entre os diferentes formatos de áudio, pela subtração das amostras em frequência de cada formato de áudio, em comparação com o formato WAV como mostrado na figura 46.

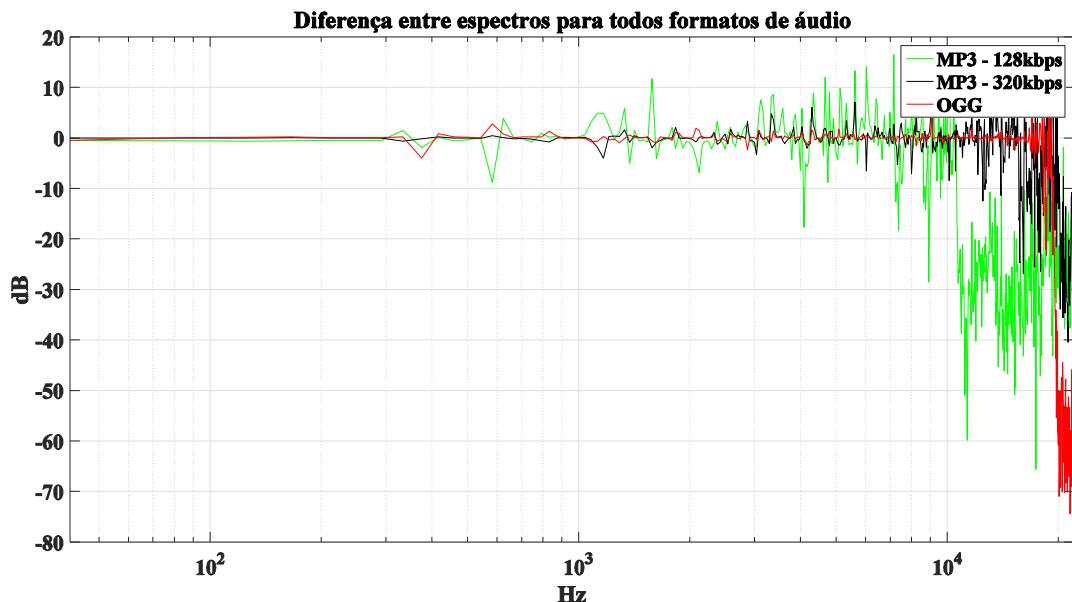


Figura 46 - Diferença dos espectros, do trecho de violão.
Fonte: Própria

4.2 ANÁLISE DOS TRECHOS ESPECÍFICOS

4.2.1 ANÁLISE DO TRECHO DE VIOLÃO

Os resultados para o trecho de violão já foram analisados na metodologia, sendo que o trecho foi retirado na marca de 0m0,702 s até 0m0,726 s da música Cirandeira de Elba Ramalho em formato WAV (32-bit float).

4.2.2 ANÁLISE DO TRECHO DE VOCAL

O trecho de vocal foi retirado na marca de 4m50,902 s até 4m50,926 s da música Cirandeira de Elba Ramalho em formato WAV (32-bit float).

4.2.2.1 ANÁLISE DA AMPLITUDE

Assim como observado para o trecho de violão, as maiores diferenças na amplitude ocorrem para o formato de MP3 - 128 kbps, como demonstrado na figura 47.

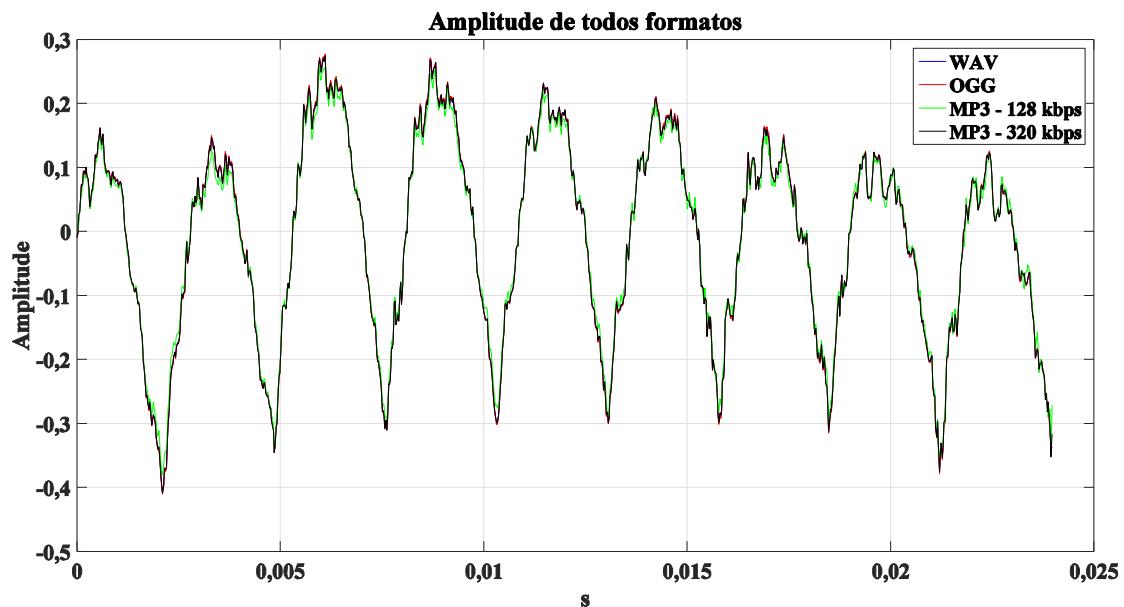


Figura 47 - Análise na amplitude com todos formatos analisados para o trecho de vocal da música Cirandeira.

Fonte: Própria.

Também foi observado que os atrasos provenientes da conversão

também foram percebidos para o trecho de vocal, desse jeito toda conversão para o formato MP3 de 128 kbps ou 320 kbps pelo software Audacity causa um atraso de 0,0512 segundos.

Pode-se demonstrar este fenômeno pelo Matlab também, pois quando criado o vetor pela função "audioread", eles resultaram em vetores de mesmo tamanho ao do trecho de violão, com 3451 amostras para o MP3 (128 kbps) e 3454

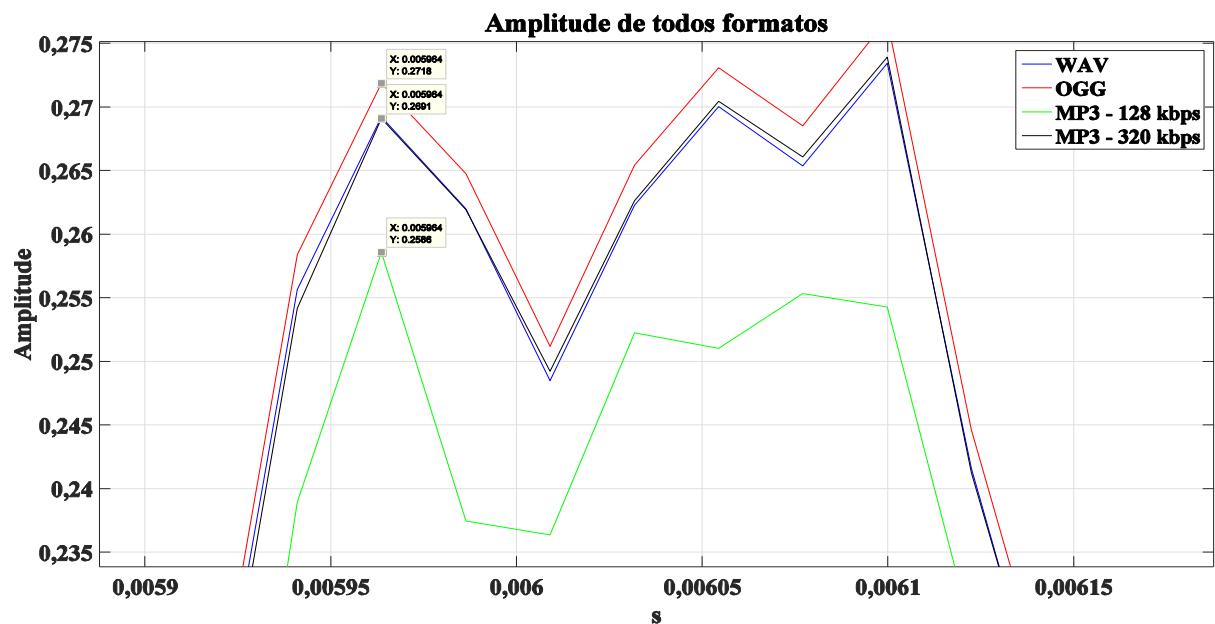


Figura 48 - Análise na amplitude para o formato MP3 em 128 kbps no trecho de vocal da música Cirandeira.

Fonte: Própria.

Pode-se mostrar este atraso também pelo instante em que cada amostra se encontra, pois para o mesmo tempo as amostras se encontram em seus respectivos picos, como mostrado pela figura 48.

4.2.2.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE

Assim como observado para o trecho de violão, para a amplitude do trecho de violão na figura 49, as maiores diferenças ocorrem justamente nos momentos de pico da forma de onda, sendo que o formato MP3 - 128 kbps apresenta a maior diferença, com valor de aproximadamente 2% da diferença de amplitude.

Em seguida o formato OGG, que também apresenta diferenças significativas para os instantes de pico, com valor de aproximadamente 0,25% da diferença de amplitude.

Já o formato em MP3 - 320 kbps não mostrou diferenças nos momentos de pico e seus valores máximos de diferença foram de 0,25%, porém em instantes menores.

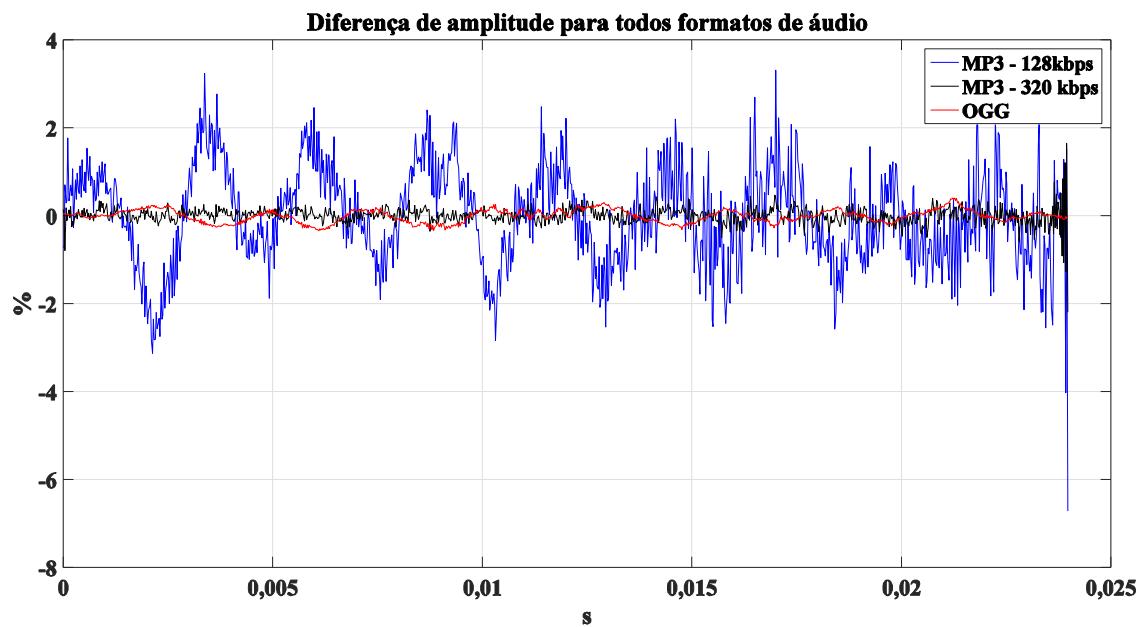


Figura 49 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de vocal da música Cirandeira.
Fonte: Própria.

4.2.2.3 ANÁLISE DOS ESPECTROS

Na análise dos espectros para o trecho de vocal, não foram observados os mesmos resultados do trecho de violão em relação aos formatos MP3 - 320kbps e OGG como mostrado na figura 50, pois a partir de 20 kHz o formato de MP3 - 320 kbps diminuiu aos mesmos valores do formato MP3 - 128 kbps. Já o formato OGG, ficou próximo do espectro do formato WAV.

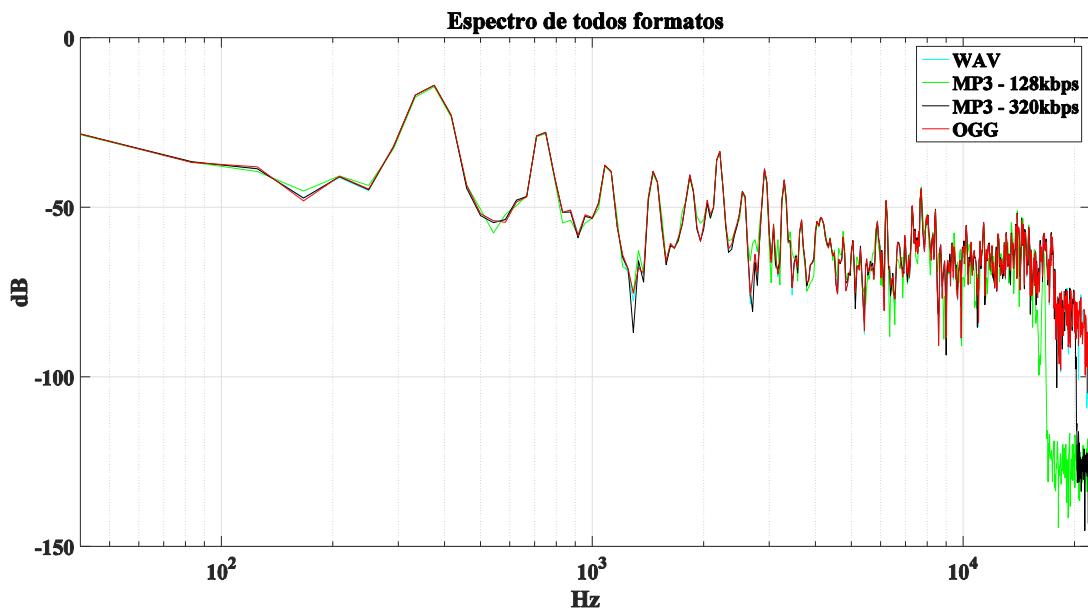


Figura 50 - Análise de todos os espectros em todos os formatos analisados para o trecho de vocal da música Cirandeira.

Fonte: Própria.

4.2.2.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS

Como constatado anteriormente na análise dos espectros, o formato MP3 - 320 kbps mostrou uma diferença maior em relação ao formato OGG, que para este trecho é o formato com menores diferenças no espectro da frequência como observado na figura 51.

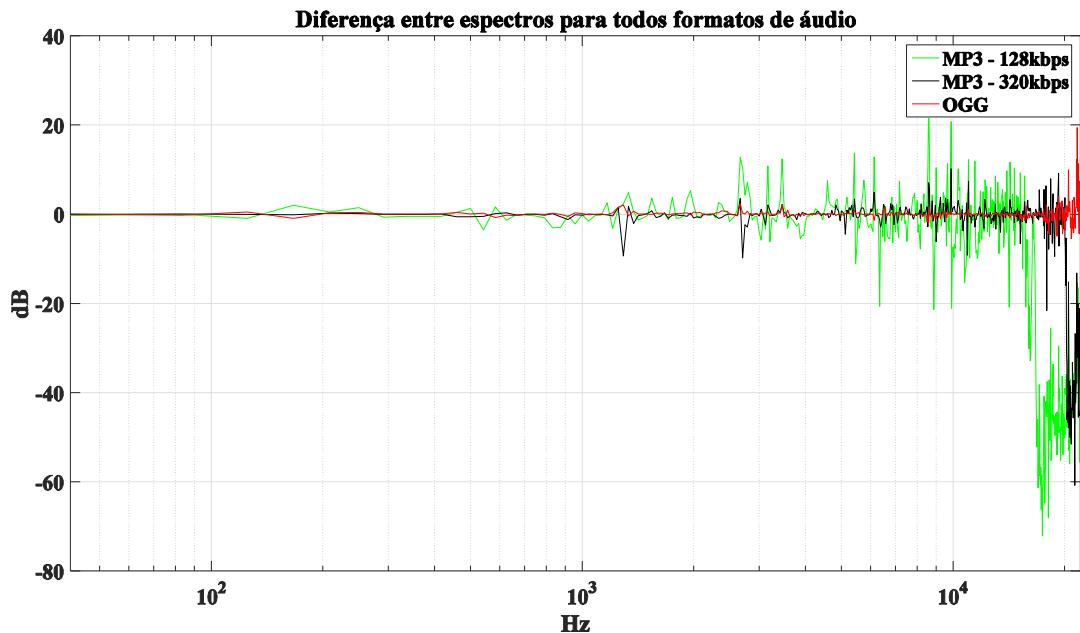


Figura 51 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de vocal da música Cirandeira.
Fonte: Própria.

4.2.3 ANÁLISE DO TRECHO DE FLAUTA

O trecho de flauta foi retirado na marca 4m25,951 s até 4m25,975 s da música Cirandeira de Elba Ramalho no formato WAV (32-bit float)

4.2.3.1 ANÁLISE DA AMPLITUDE

Novamente foi observado que o formato MP3 - 128 kbps apresentou a maior diferença de amplitude em relação ao formato WAV pela figura 52.

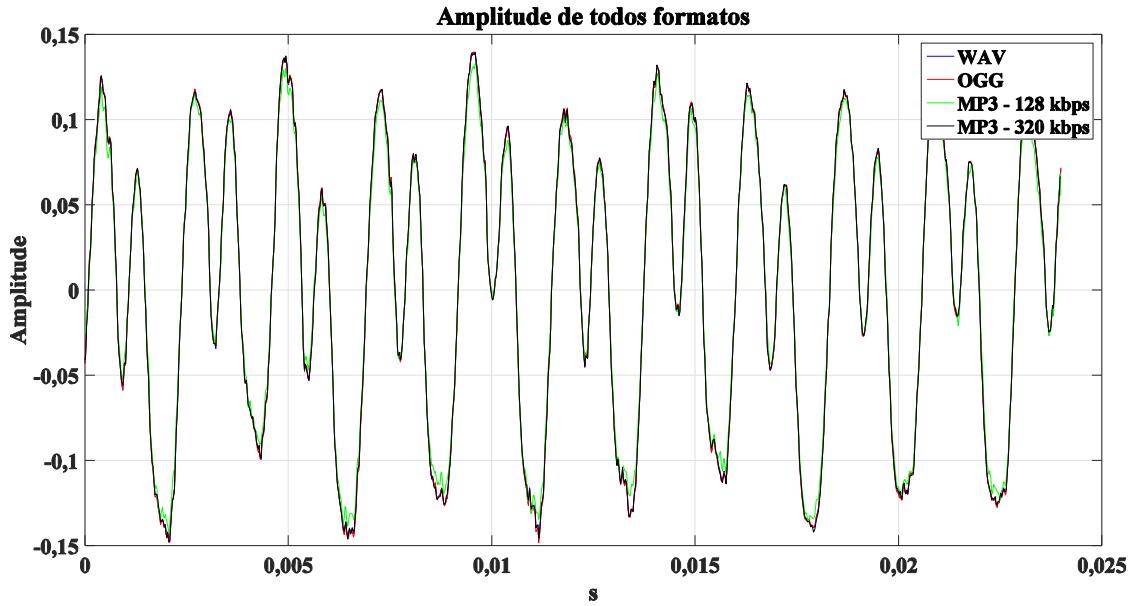


Figura 52 - Análise na amplitude com todos formatos analisados para o trecho de flauta da música Cirandeira.
Fonte: Própria.

4.2.3.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDE

Para o trecho de flauta foi observado menores diferenças de amplitude, em torno de 1,3% em relação ao formato WAV. Novamente ocorreu que os instantes de picos têm os maiores valores de diferença na amplitude para os formatos WAV - 128 kbps e OGG pela figura 52.

Os valores de diferença do formato MP3 - 320 kbps e OGG continuaram na mesma faixa de resultados do trecho anterior, de aproximadamente 0,2%.

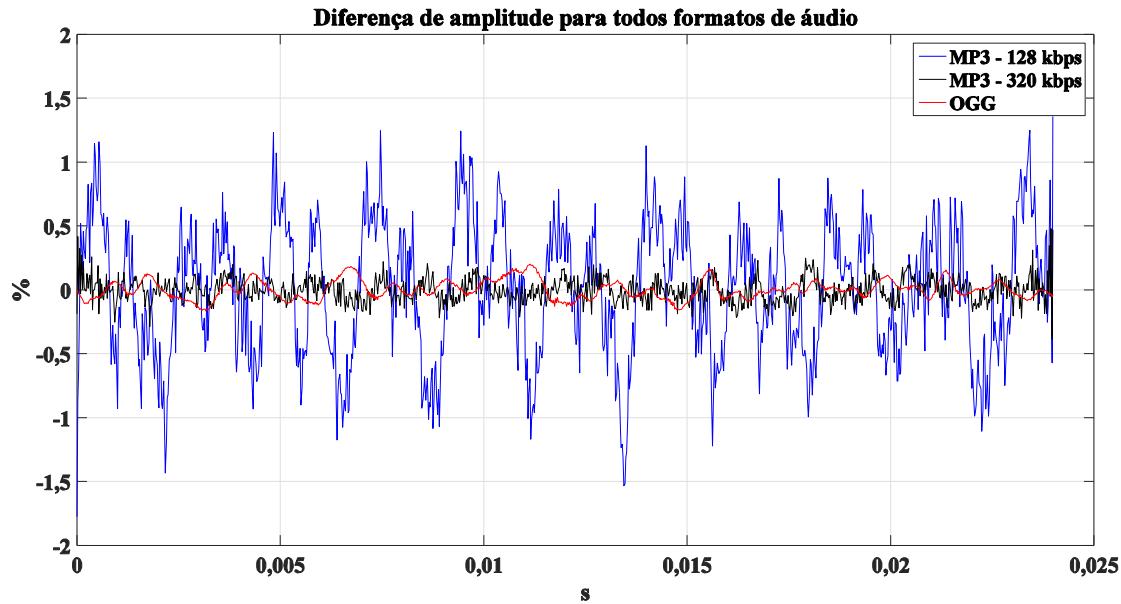


Figura 53 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de flauta da música Cirandeira.
Fonte: Própria.

4.2.3.3 ANÁLISE DOS ESPECTROS

Os resultados do espectro para todos formatos no trecho de flauta, foi similar aos resultados obtido no trecho de vocal pela figura 54. O formato MP3 - 128 kbps já apresenta diferenças significativas no espectro a partir de 10 kHz, enquanto que para o formato MP3 - 320 kbps as diferenças mais significantes aparecem a partir de 20 kHz. O formato OGG apresentou as menores diferenças no espectro em relação ao formato WAV.

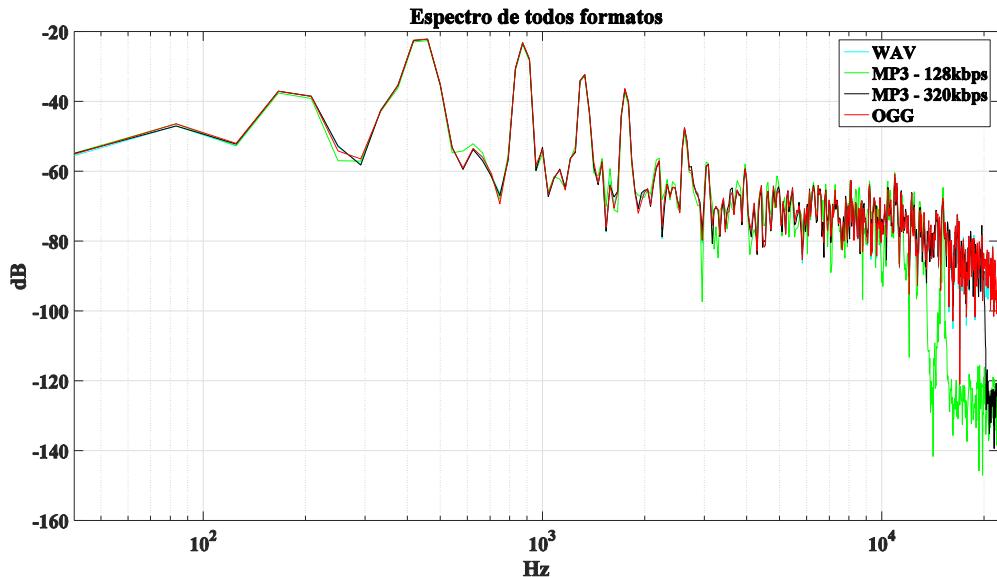


Figura 54 - Análise de todos espectros em todos formatos analisadas para o trecho de flauta da música Cirandeira.

Fonte: Própria.

4.2.3.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS

Observando a figura 55 temos que os formatos de MP3 mostraram as maiores diferenças entre o espectro WAV, o formato OGG não apresentou diferenças significativas no espectro.

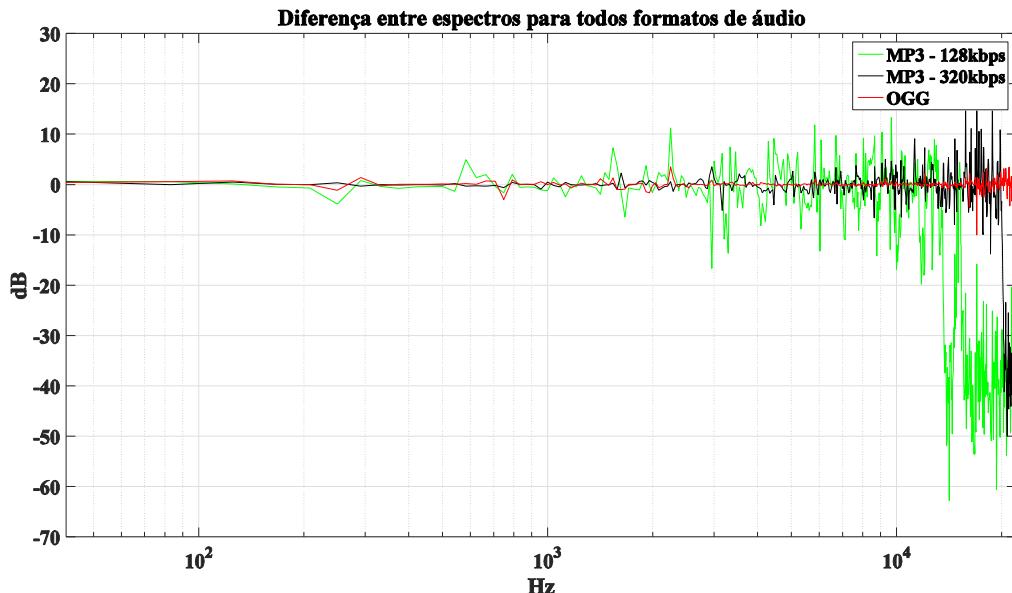


Figura 55 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de flauta da música Cirandeira.

Fonte: Própria.

4.2.4 ANÁLISE DO TRECHO DE VIOLINO

O trecho de violino foi retirado na marca de 4m45,221 s até 4m45,247 s da música Cirandeira de Elba Ramalho em formato WAV (32-bit float).

4.2.4.1 ANÁLISE DA AMPLITUDE

Como mostrado na figura 56, o sinal não chega a ser periódico como os outros trechos, talvez devido a frequência natural do instrumento. Novamente o formato MP3 - 128 kbps apresentou a maior diferença de amplitude.

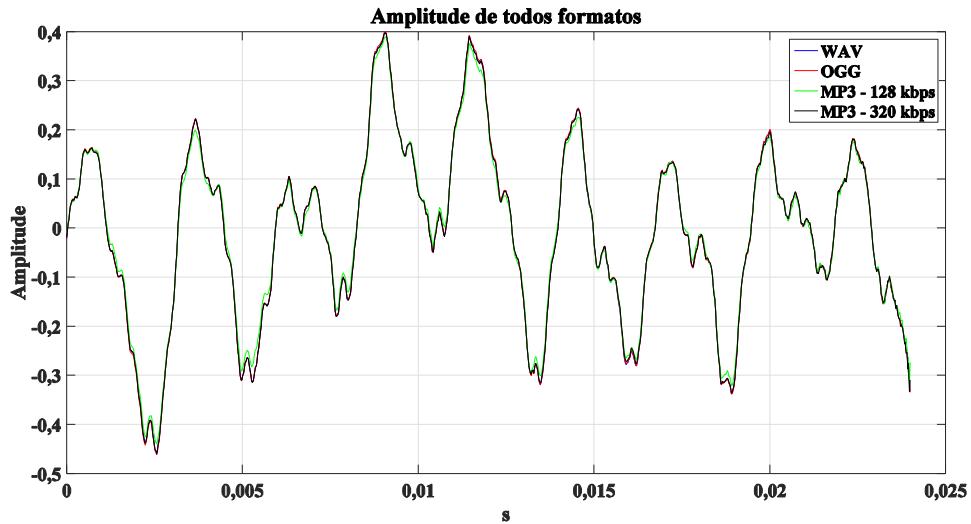


Figura 56 - Análise na amplitude com todos formatos analisados para o trecho de violino da música Cirandeira.

Fonte: Própria.

4.2.4.2 - ANÁLISE DAS DIFERENÇAS DE AMPLITUDE

Pela figura 57, notamos novamente que os formatos MP3 - 128 kbps e OGG mostraram as maiores diferenças nos instantes de pico. Enquanto que o formato MP3 - 320 kbps mostra diferenças em qualquer instante de tempo sem correlação se for algum pico de amplitude ou não.

Os valores de diferença aumentaram para o trecho de violino em relação aos outros trechos, sendo que o formato MP3 - 128 kbps apresentou 2.5% de

diferença enquanto que o formato OGG e MP3 - 320 kbps tiveram o maior valor de diferença em aproximadamente 0,5% que também é maior em relação aos outros trechos.

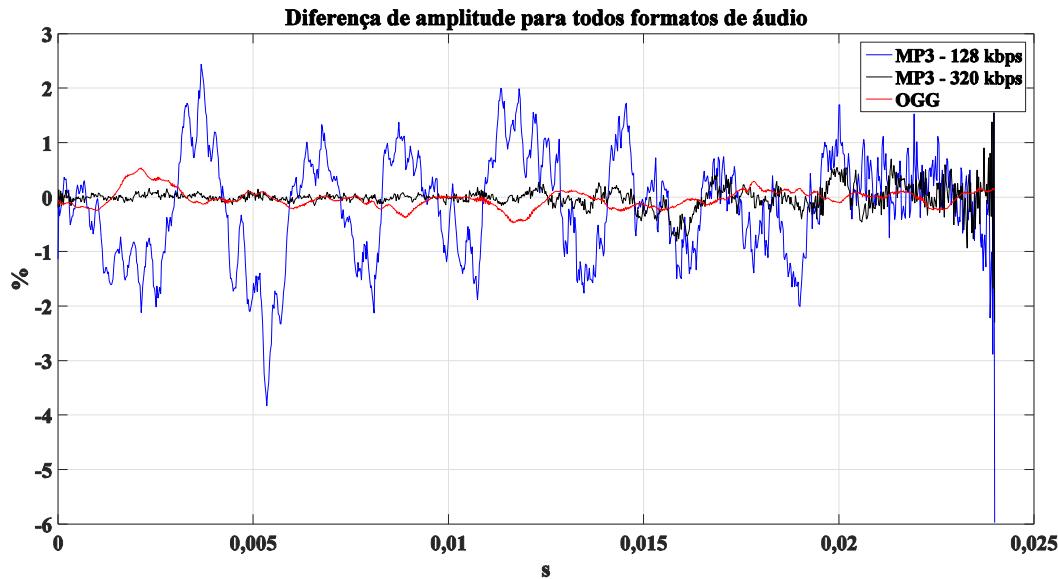


Figura 57 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de flauta da música Cirandeira.
Fonte: Própria.

4.2.4.3 ANÁLISE DOS ESPECTROS

Como mostrado na figura 58, os resultados obtidos foram similares aos trechos anteriores (exceto o trecho de violão), aonde o formato MP3 - 128 kbps mostrou diferenças a partir de 16 kHz e o formato MP3 - 320 kbps mostrou diferenças a partir de 20 kHz.

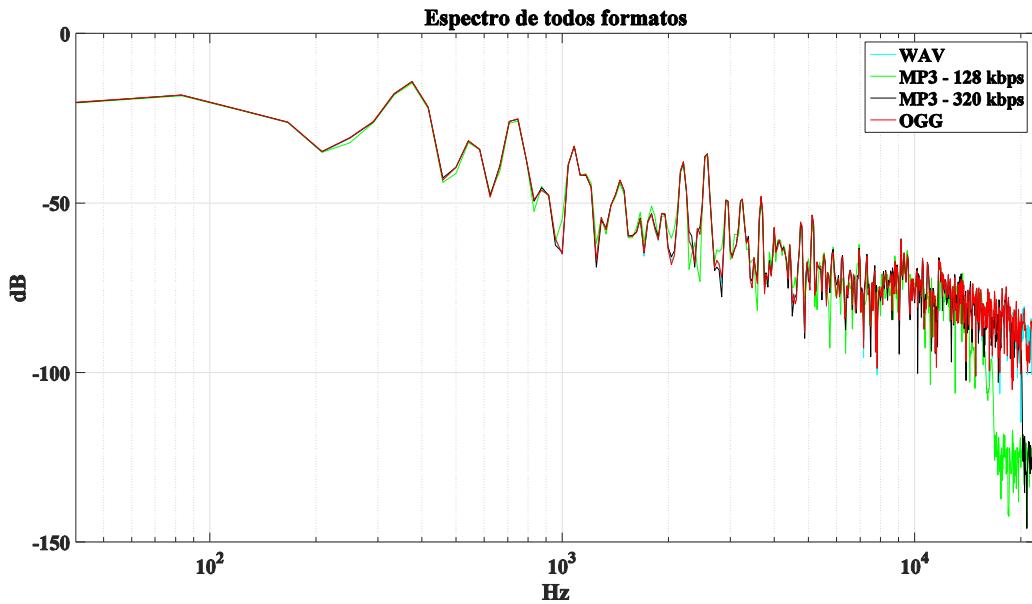


Figura 58 - Análise de todos espectros em todos formatos analisadas para o trecho de violino da música Cirandeira.
Fonte: Própria.

4.2.4.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS

Pela figura 59, o formato MP3 - 128 kbps mostrou as maiores diferenças em relação ao espectro WAV novamente, seguido pelo formato MP3 - 320 kbps. O formato OGG apesar de ter maiores diferenças de amplitude do que o formato MP3 - 320 kbps teve menores diferenças no espectro.

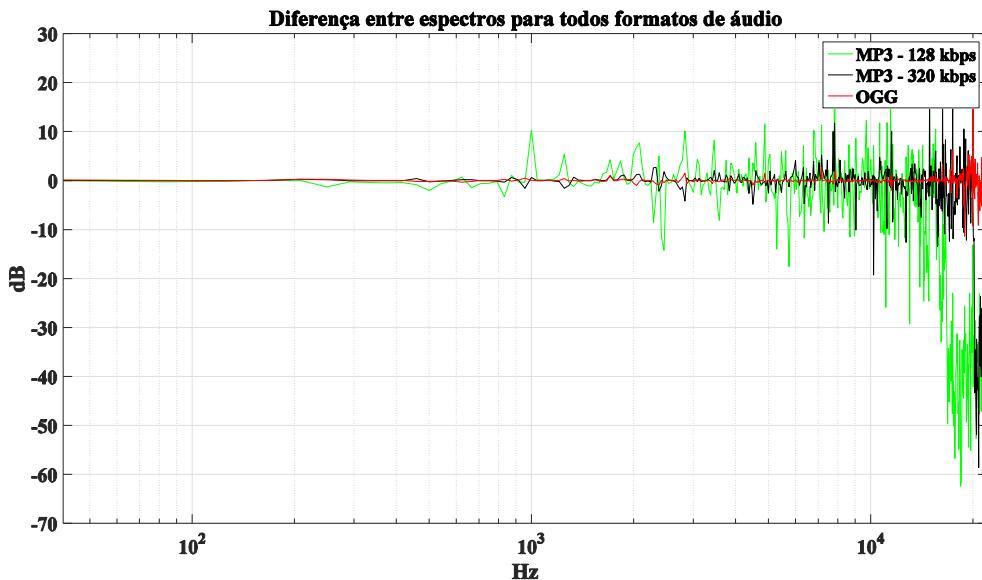


Figura 59 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de flauta da música Cirandeira.
Fonte: Própria.

4.2.5 ANÁLISE DO TRECHO DE PERCUSSÃO

O trecho da percussão foi retirado na marca de 5m02,079 s até 5m02,103 s da música Cirandeira de Elba Ramalho em formato WAV (32-bit float).

4.2.5.1 ANÁLISE DE AMPLITUDE

A amplitude de todos formatos para o trecho de percussão é mostrada na figura 60. Podemos observar pela forma de onda que há um número maior de harmônicos para o trecho de percussão em comparação com os outros trechos analisados, aumentando o número de picos na amplitude.

Pelo fato de houver maior picos, podemos esperar que haverá mais diferenças deste trecho em relação aos outros trechos analisados.

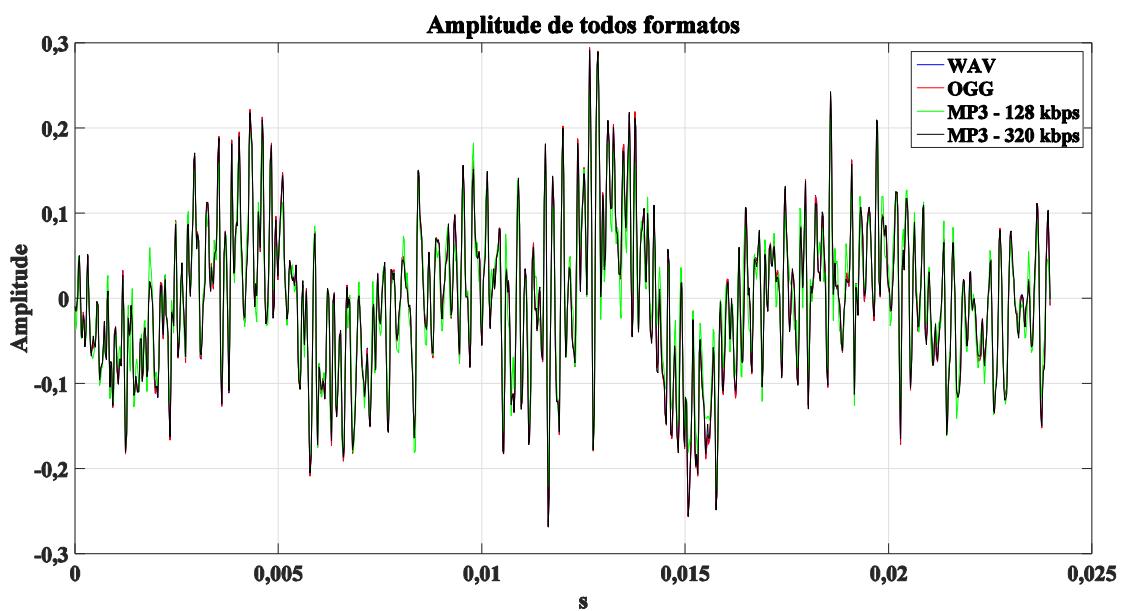


Figura 60 - Análise na amplitude com todos formatos analisados para o trecho de percussão da música Cirandeira.
Fonte: Própria.

4.2.5.2 ANÁLISE DAS DIFERENÇAS DE AMPLITUDES

Como era esperado, o formato MP3 - 128 kbit mostrou um número maior em quantidade e magnitude de diferença, devido ao número de picos que é maior. De todos trechos analisados o trecho de percussão mostrou os maiores

valores de diferença, de até 7% em relação ao formato WAV.

Curiosamente os resultados para a análise do trecho de percussão no formato OGG apresentou as menores diferenças. Amplificou-se o gráfico de diferenças na amplitude, como mostrado na figura 62, para ter-se uma análise mais precisa dos resultados é concluído que o formato em OGG apresentou as menores diferenças de amplitude para este trecho.

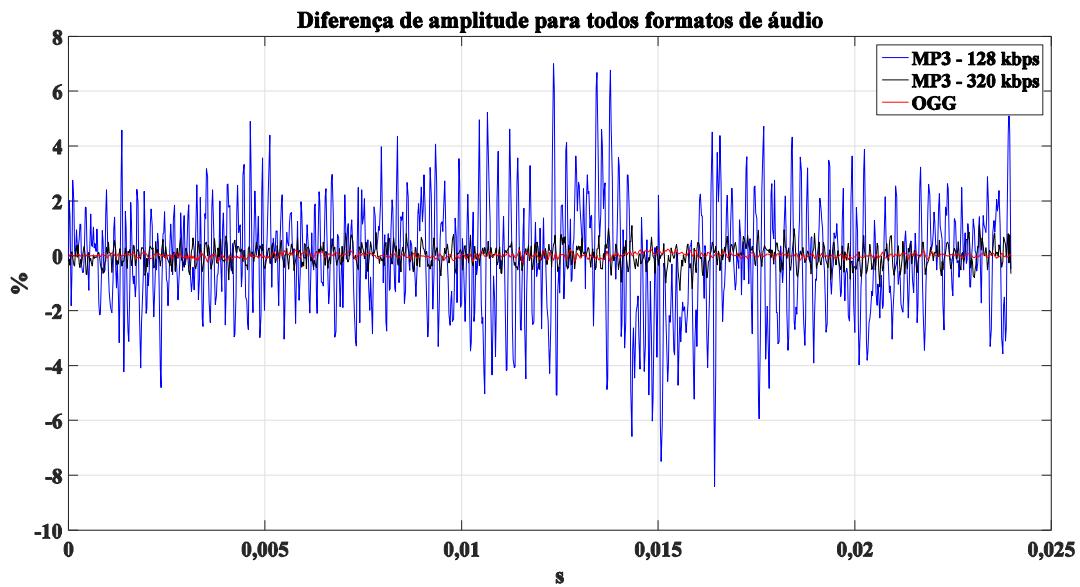


Figura 61 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de percussão da música Cirandeira.

Fonte: Própria.

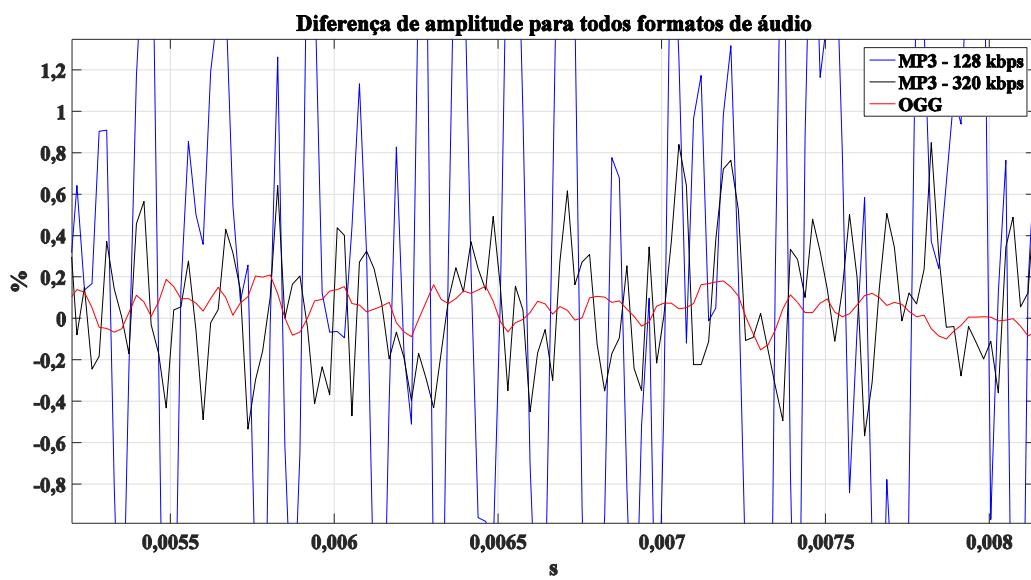


Figura 62 - Análise na amplitude com todas diferenças na amplitude analisadas para o trecho de percussão da música Cirandeira.

Fonte: Própria.

4.2.5.3 ANÁLISE DOS ESPECTROS

Foi observado que os gráficos dos espectros para o trecho de percussão apresentaram fenômenos similares aos observados nos trechos anteriores (exceto do trecho de violão). Aonde há uma queda dos valores em dB a partir de 16 kHz para os formatos em MP3.

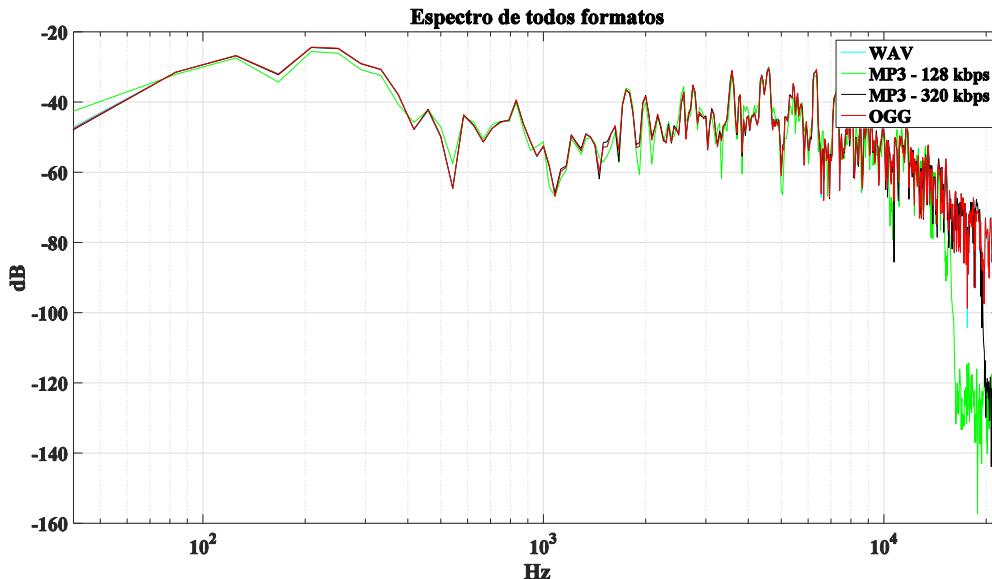


Figura 63 - Análise de todos os formatos analisadas para o trecho de percussão da música Cirandeira.

Fonte: Própria.

4.2.5.4 ANÁLISE DAS DIFERENÇAS DE ESPECTROS

Pela figura 64, foi novamente observado que os formatos de MP3 apresentaram as maiores diferenças. Podemos concluir por este gráfico que não há correlação da amplitude dos formatos de áudio com o espetro deles.

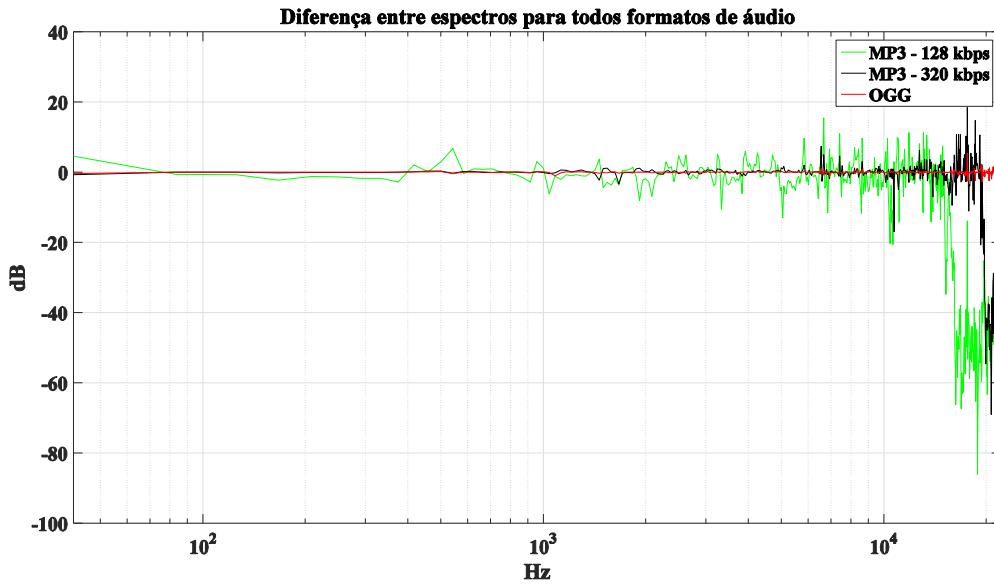


Figura 64 - Análise das diferenças dos espectros em todos formatos de áudio analisados para o trecho de repercussão da música Cirandeira.
Fonte: Própria.

4.6 ANÁLISE DO ESPECTROGRAMA

Para a análise dos spectrogramas, é definido o valor da janela ("window") para $0.5 * F_s$ para todas análises, assim obteve-se a figura 65, rotacionando o spectrograma é obtida a figura 66. É feito a análise do spectrograma para os outros formatos de áudio além também dos vetores de diferença da amplitude dos formatos de áudio em relação ao WAV.

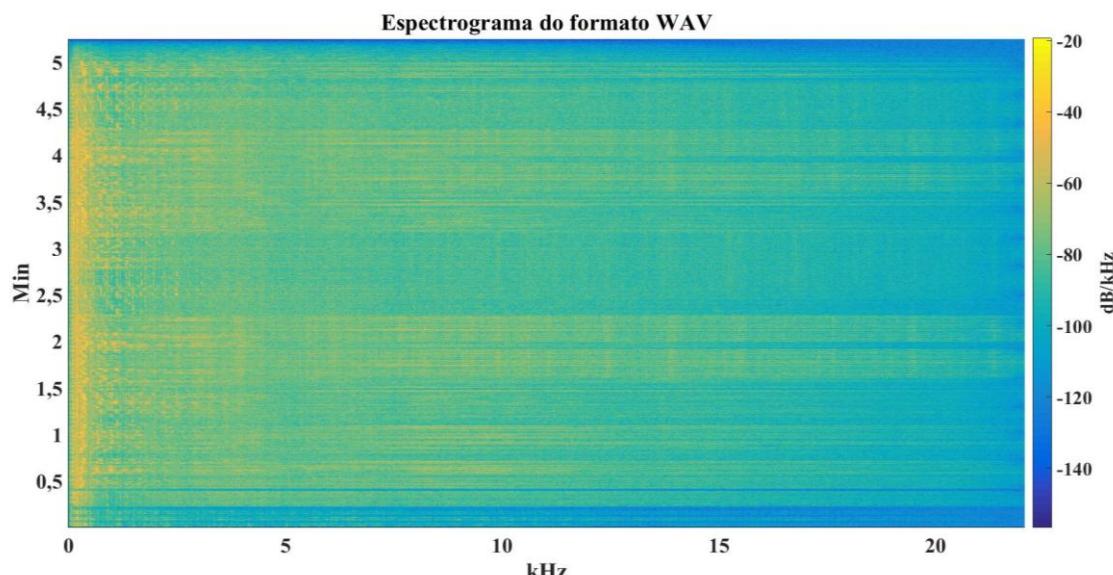


Figura 65 - Espectrograma do formato WAV.
Fonte: Própria.

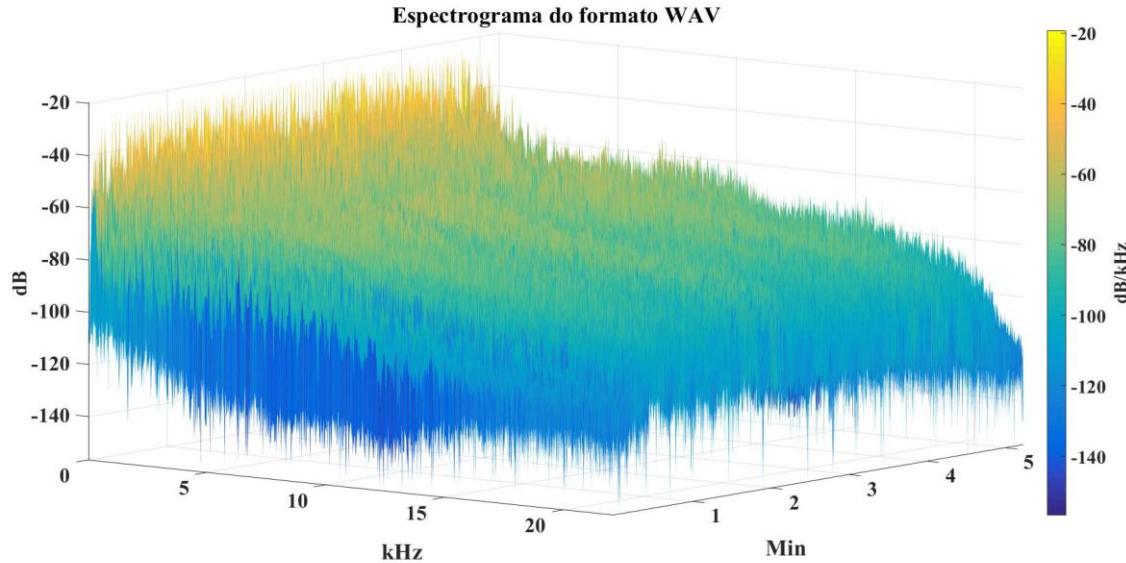


Figura 66 - Espectrograma do formato WAV, com as três dimensões.
Fonte: Própria.

Na análise do espectrograma para o formato MP3 - 128 kbps mostrado na figura 67 e figura 68, foi observado a queda na frequência a partir de 16 kHz como era esperado pelos resultados do spectro.

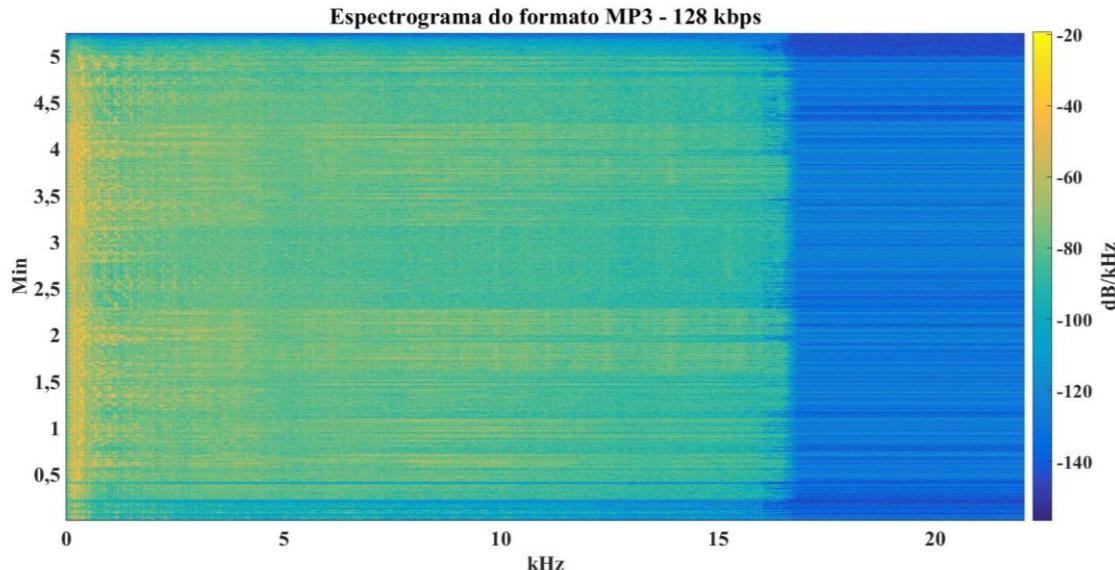


Figura 67 - Espectrograma do formato MP3 - 128 kbps.
Fonte: Própria.

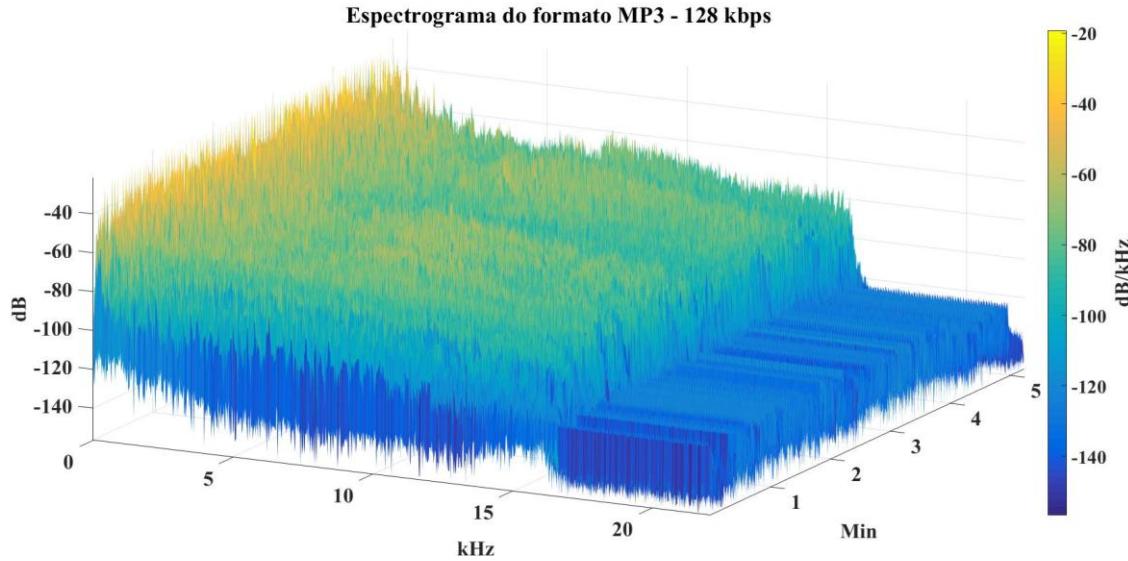


Figura 68 - Espectrograma do formato MP3 - 128 kbps, com as três dimensões.
Fonte: Própria.

Em seguida para a análise do formato MP3 - 320 kbps mostrado nas figuras 69 e 70, ocorre a mesma queda de frequência, porém a partir de 20kHz, ou seja, fora do limite audível humano.

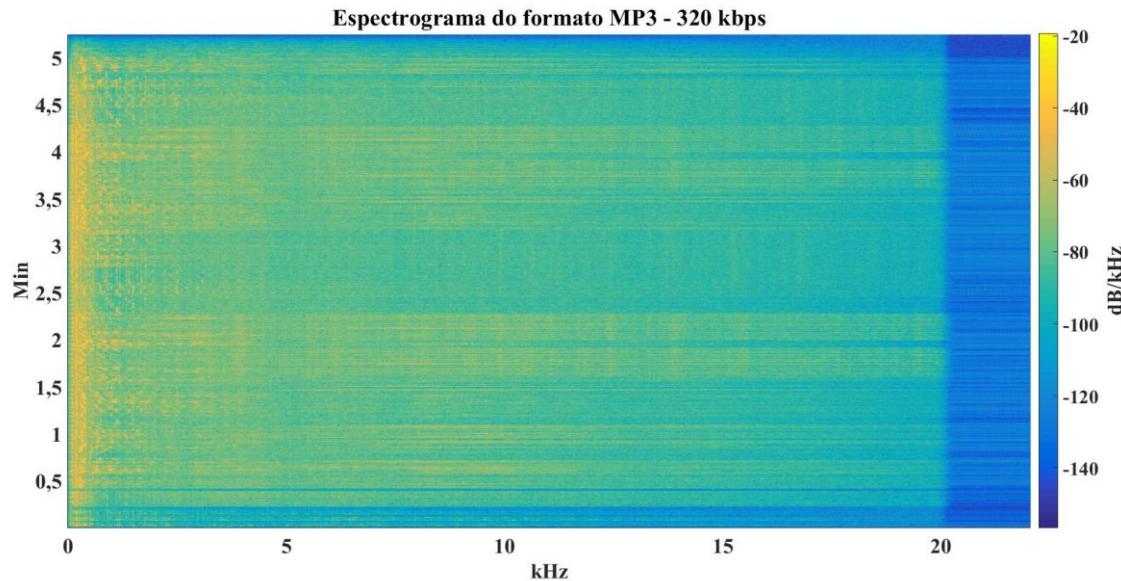


Figura 69 - Espectrograma do formato MP3 - 320 kbps.
Fonte: Própria.

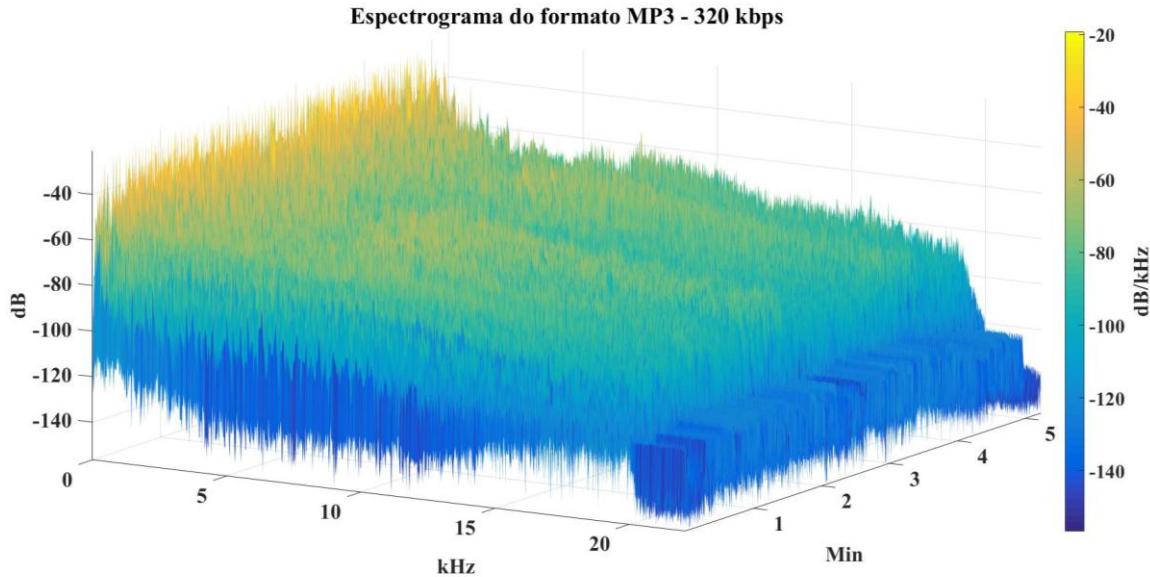


Figura 70 - Espectrograma do formato MP3 - 320 kbps, com as três dimensões.
Fonte: Própria.

O formato OGG mostrado na figura 71 e figura 72, mostrou um espectrograma com comportamento similar ao formato WAV, sem quedas de frequências.

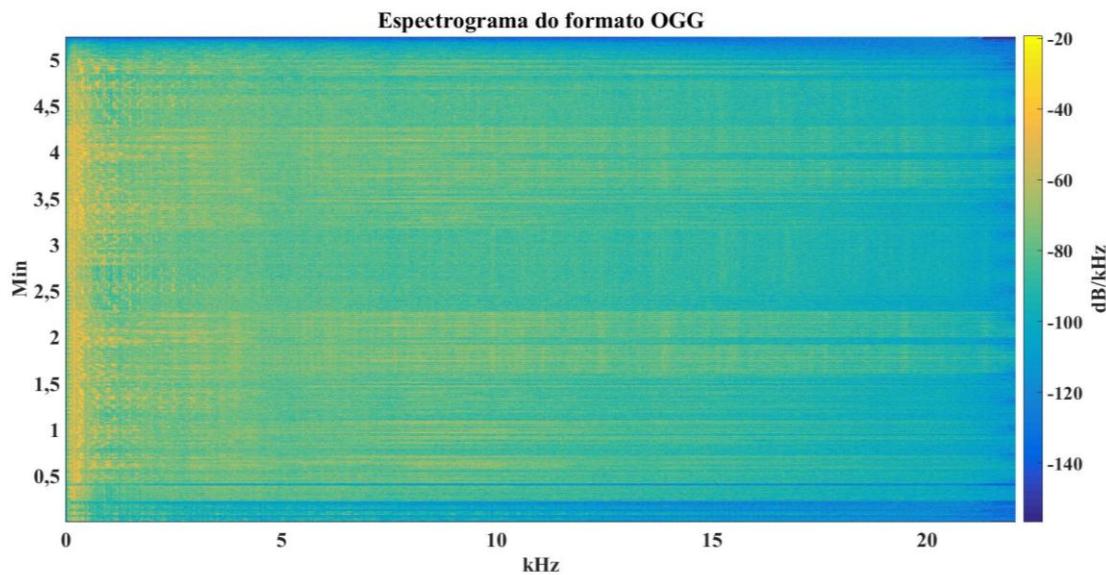


Figura 71 - Espectrograma do formato OGG.
Fonte: Própria.

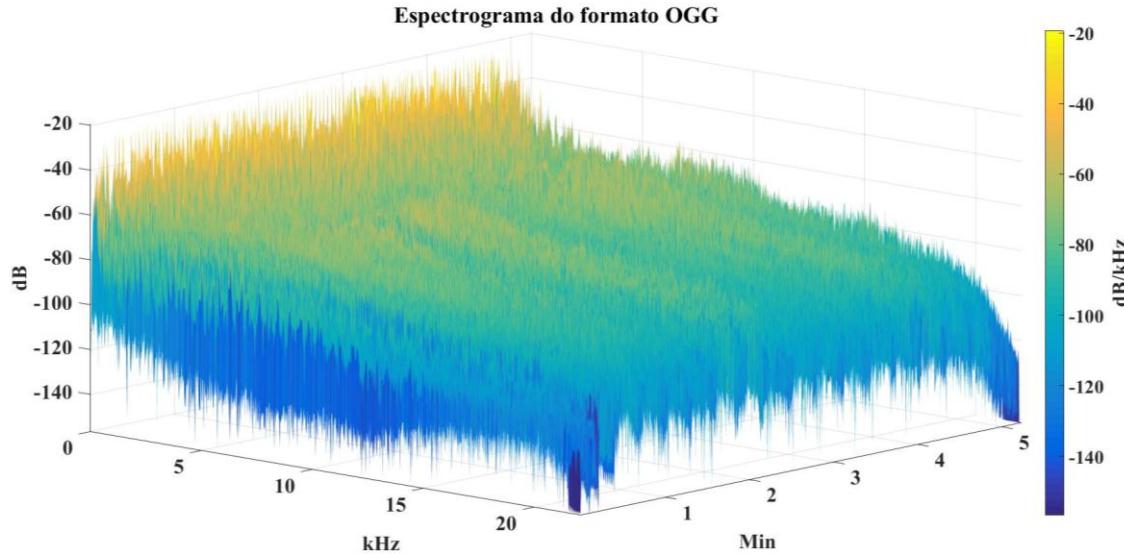


Figura 72 - Espectrograma do formato OGG, com as três dimensões.
Fonte: Própria

Fez-se em seguida o espectrograma da diferença entre o vetor da amplitude do formato MP3 e WAV, assim é obtido as figuras 73 e 74. Observa-se que a intensidade da relação dB/kHz diminuiu em comparação com o formato WAV.

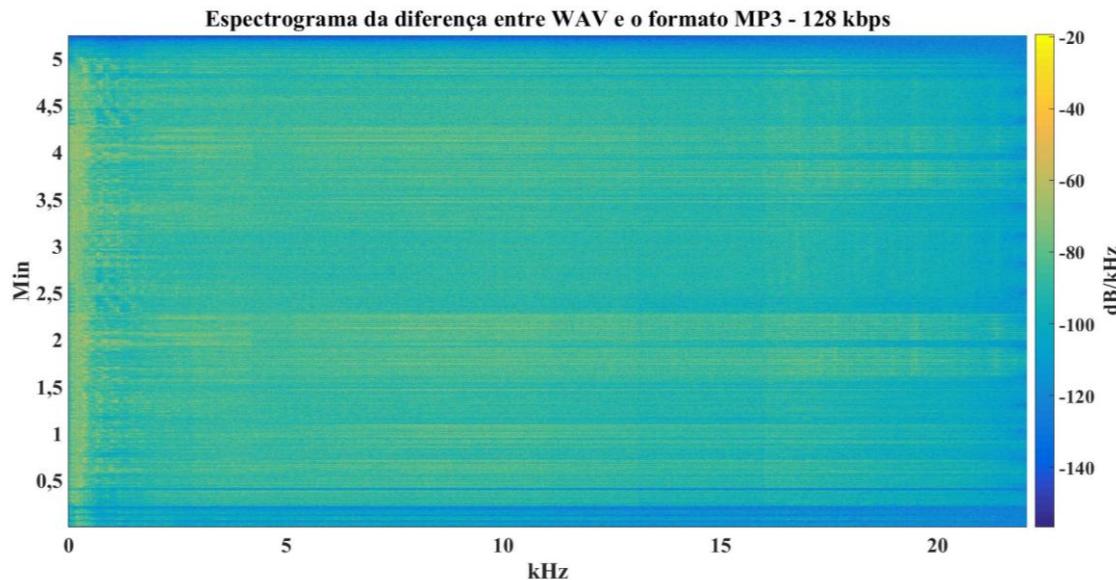


Figura 73 - Espectrograma da diferença entre WAV e MP3 - 128 kbps.
Fonte: Própria.

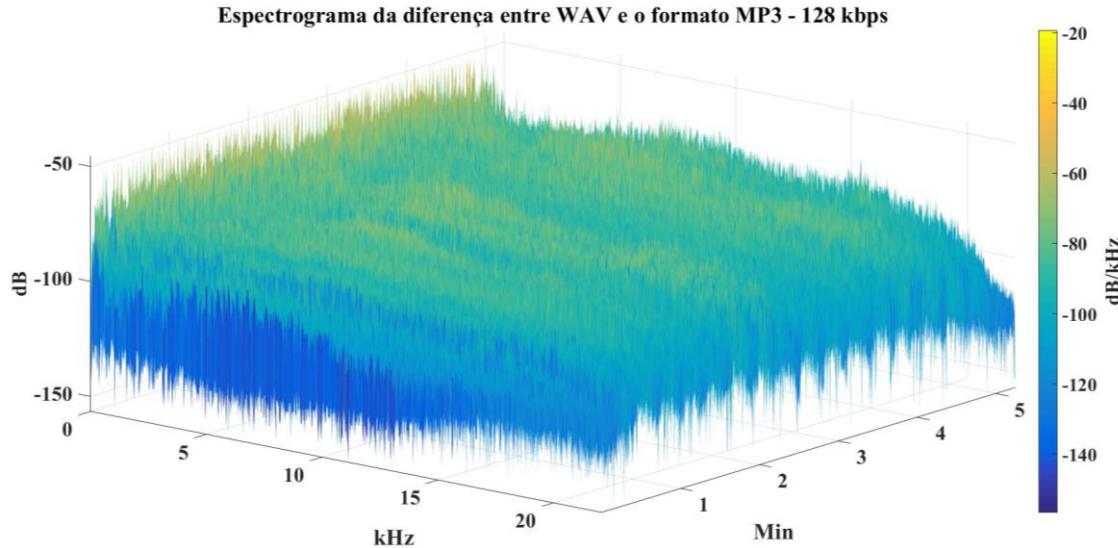


Figura 74 - Espectrograma da diferença entre WAV e MP3 - 128 kbps com as três dimensões.
Fonte: Própria.

Para a diferença entre MP3 - 320 kbps e WAV, foi notado que a intensidade da relação dB/kHz diminuiu mais em relação ao formato MP3 - 128 kbps como observado nas figuras 75 e 76.

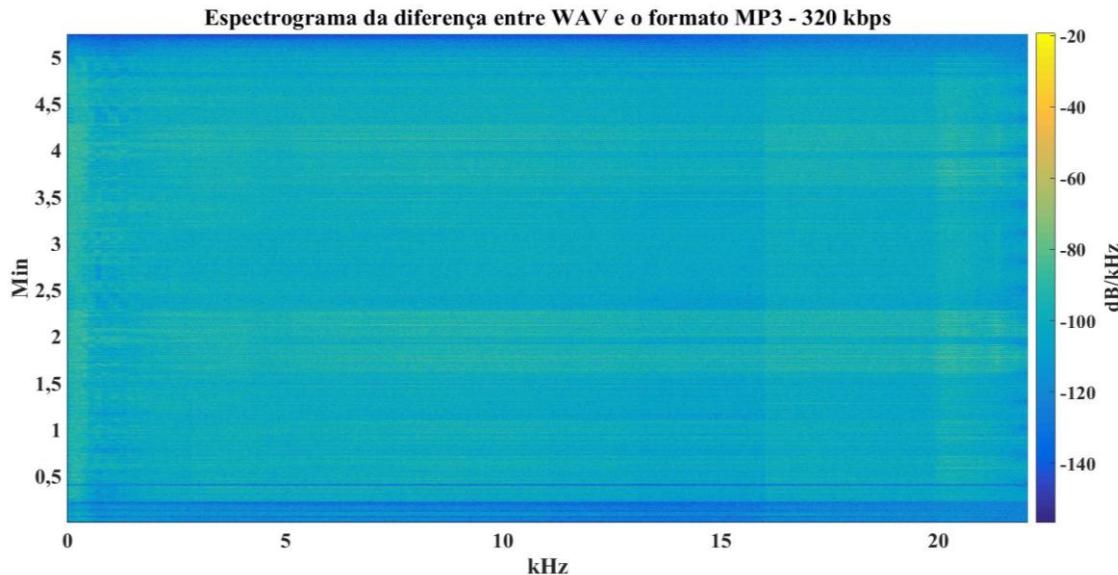


Figura 75 - Espectrograma da diferença entre WAV e MP3 - 320 kbps.
Fonte: Própria.

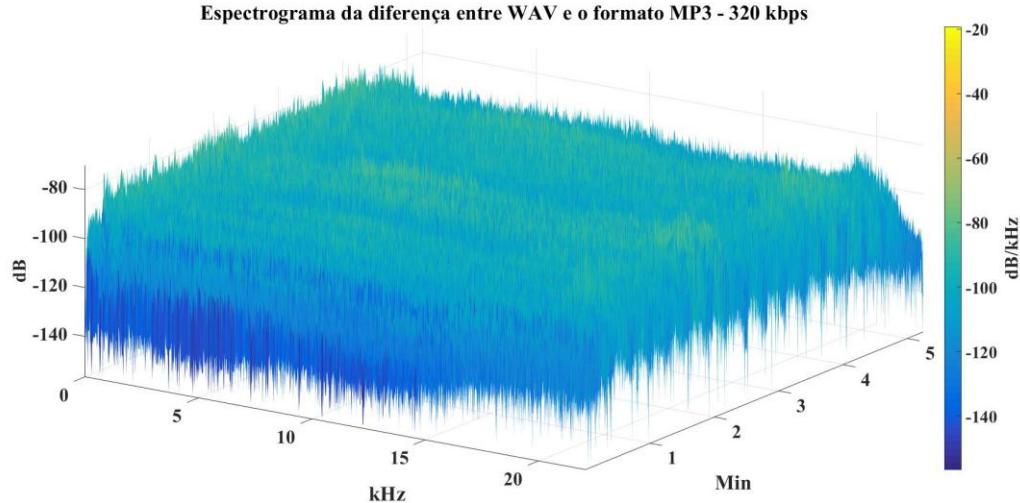


Figura 76 - Espectrograma da diferença entre WAV e MP3 - 320 kbps com as três dimensões.

Fonte: Própria.

Finalmente para a diferença do formato OGG com o WAV, mostrado nas figuras 77 e 78, foi observado que a intensidade de dB/kHz foi a menor, sendo que o espectrograma não apresenta muitas cores amarelas.

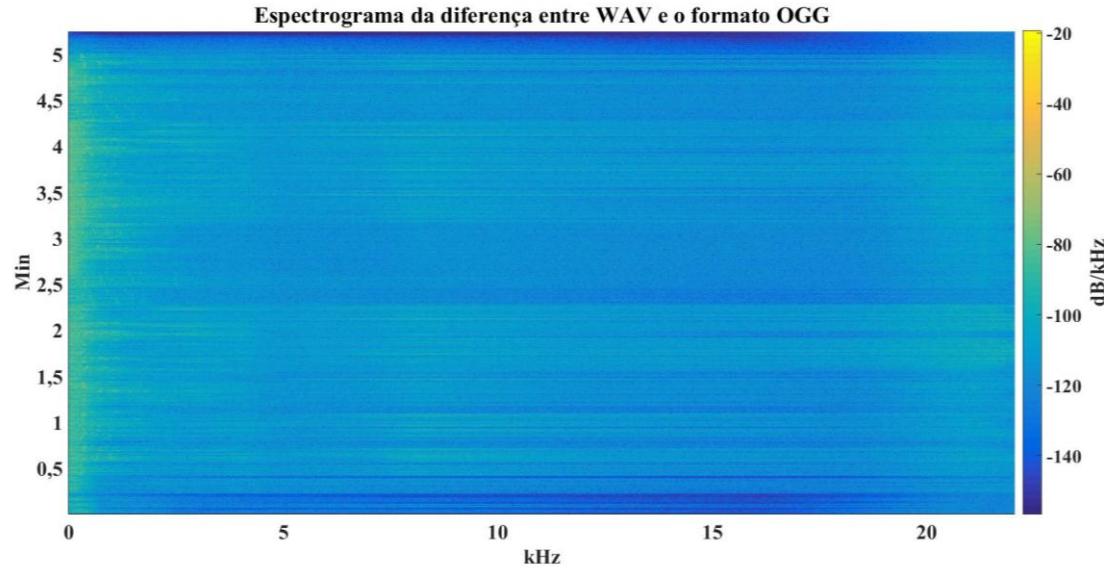


Figura 77 - Espectrograma da diferença entre WAV e OGG.

Fonte: Própria.

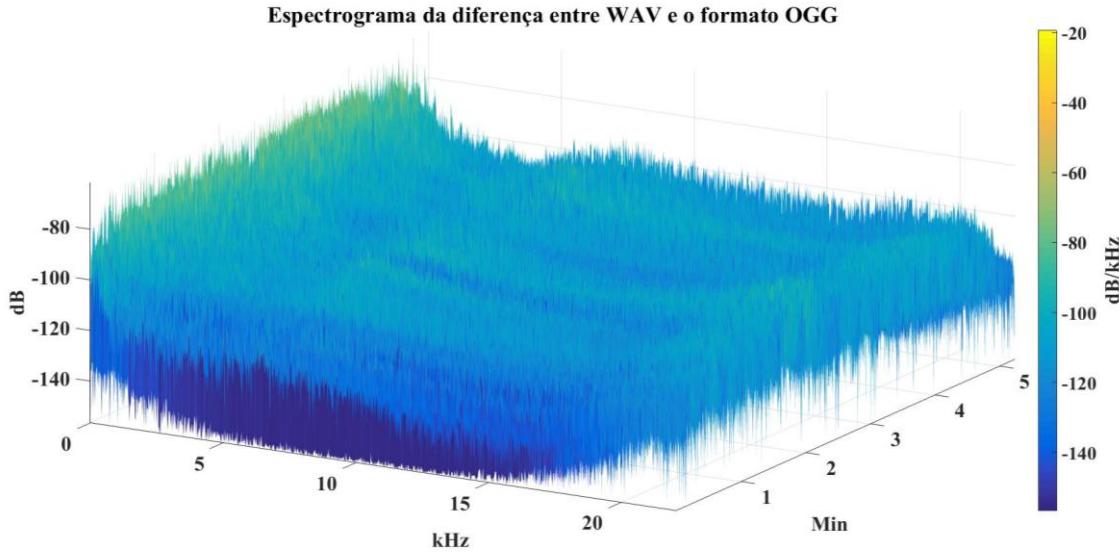


Figura 78 - Espectrograma da diferença entre WAV e OGG com as três dimensões.
Fonte: Própria.

5. CONCLUSÃO

O desenvolvimento do presente estudo possibilitou uma análise objetiva das diferenças entre os formatos de áudio compactado quando, comparando com o formato sem compressão, e também com outros formatos de compressão e se a diferença encontrada é percebida por ouvintes subjetivamente.

A fim de se cumprir os objetivos propostos desenvolveu-se uma fundamentação teórica para o tema, análise objetiva entre áudio comprimido e sem compressão, através de software, e a análise subjetiva da qualidade de amostras de áudio com compressão e sem compressão através da aplicação de formulários que foram submetidos a procedimentos estatísticos.

Na parte subjetiva do trabalho, os resultados obtidos a partir dos questionários não são estatisticamente significativos. Ou seja, estatisticamente é como se não houvesse diferença percebida entre o arquivo original e os arquivos com compressão. A partir dos resultados obtidos conclui-se que apesar de qualidade ter parâmetros possíveis de mensurar, cada indivíduo tem critérios pessoais e por vezes podem até gostar das distorções causadas pela conversão, conforme pode-se analisar através dos comentários presentes nos formulários.

Apesar de o resultado não ser o desejado, ao ler cada questionário individualmente observa-se que todos os outros participantes sentiram diferenças entre o áudio sem compressão e o áudio com compressão, porém não conseguem ordenar com exatidão qual faixa tem maior qualidade, por exemplo, houveram participantes que preferiram o som da faixa de mais baixa qualidade (MP3 a 128kbps) ao arquivo WAV original.

A diferença entre o arquivo de áudio original e o arquivo com compressão

é clara objetivamente. Subjetivamente podemos dizer que os participantes conseguiram sentir diferenças, mas a qualificação dos arquivos não é consoante com a semelhança deles ao original. Em síntese, a classificação da qualidade das amostras de áudio acaba dando-se a partir do gosto pessoal de cada indivíduo.

A análise objetiva foi subdividida e realizada em etapas que envolviam a análise de diferenças de amplitude, análise de espectro, análise de trechos específicos com instrumentos em destaque e por fim análise de spectrograma comparando os arquivos de áudio comprimido com o arquivo fonte sem compressão.

Na análise de amplitude observou-se uma situação onde a diferença entre os formatos chama a atenção é nos instantes de picos máximos e mínimos, notavelmente nos formatos de MP3 128 kbps e OGG.

Quando comparamos os espectros de cada formato, observamos diferenças principalmente em alta frequência, acima de 10.000 Hz, com o MP3 128 kbps tendo o pior desempenho e o OGG mais próximo do áudio sem compressão.

Através da análise dos trechos específicos pode-se observar que o formato MP3 128 kbps apresenta bastante diferença da onda original e que o formato MP3 tem sua diferença acentuada em músicas ou trechos que são ricos em dinâmica. A análise específica do trecho de percussão mostra muito claramente a diferença de desempenho entre os formatos. Apesar do OGG apresentar maior diferença nos picos de máximo e mínimos é o formato que mais se assemelhou ao original nesse quesito.

A criação e comparação de spectrogramas é extremamente útil e facilita

muito a visualização e localização das diferenças entre cada formato. Pode-se observar mais uma vez que o formato MP3 128 kbps tem o pior desempenho e que o OGG Vorbis fica mais próximo do arquivo WAV original.

Este trabalho buscou avaliar diferentes de formatos de áudio, escolhidos com base na popularidade de uso, de maneira a identificar/qualificar a qualidade de áudio de cada um deles. O MP3 é um formato de grande importância não só para esse estudo, mas também para a história da música, internet e inclusive o mercado fonográfico. Mas pode-se demonstrar que o OGG Vorbis é um formato superior quanto à fidelidade ao arquivo original.

REFERÊNCIAS

AES. An Audio Timeline. 1999. Disponível em <<http://www.aes.org/aeshc/docs/audio.history.timeline.html>>

Alberti, Peter W. "The anatomy and physiology of the ear and hearing." *Occupational exposure to noise: Evaluation, prevention, and control* (2001): 53-62.

Alec H. Reeves. The Past, Present and Future of PCM. 1965

Arthur Alexandre Hackbarth Neto, Carlos Efrain Stein. "UMA ABORDAGEM DOS TESTES NÃO-PARAMÉTRICOS COM UTILIZAÇÃO DO EXCEL". 2003.

AT&T "A telephone achievement ranking with the opening of the transcontinental line". 1923

B. Grill, S. Quackenbush. "MPEG-1 Audio". Disponível em <<http://mpeg.chiariglione.org/standards/mpeg-1/audio>>. Acessado em 05 de maio de 2017

Ballou, Glen. *"Handbook for sound engineers"*. 2008.

Bearinger, Clayton R., et al. "Flip chip silicone pressure sensitive conductive adhesive." U.S. Patent No. 5,611,884. 18 Mar. 1997.

Castells, M. "The Information Age", Volumes 1-3: Economy, Society and Culture. Editora John Wiley & Sons, 2011.

Castells, Manuel. *"Information technology, globalization and social development"*. Vol. 114. Geneva: United Nations Research Institute for Social Development, 1999.

Dan Waters. "Intro to Audio Programming, Part 2: Demystifying the WAV Format". Disponível em <https://blogs.msdn.microsoft.com/dawate/2009/06/23/intro-to-audio-programming-part-2-demystifying-the-wav-format/>. Acessado em 13 de Maio de 2017

David Solomon. Data Compression: The Complete Reference. Editora Springer-Verlag London, 2006.

Deloraine, E. Maurice, and Alec H. Reeves. "The 25th anniversary of pulse code modulation." *IEEE spectrum* 2.5 (1965): 56-63.

Engel, Friedrich, Peter Hammar, and R. Hess. "A selected history of magnetic recording." 2006. Disponível em: <<http://www.richardhess.com>

com/tape/history/Engel_Hammar--Magnetic_Tape_History.pdf> Acessado em 02 de novembro de 2017

Flávio Adami, "O Formante". Disponível em <<http://www.bynirsch.com.br/artigos-11-08-timbres55.shtml>>. Acessado em 15 de junho de 2017.

Flávio Adami, "O Timbre e seu Envelope". Disponível em <<http://www.bynirsch.com.br/artigos-11-08-audionews52.shtml>>. Acessado em 15 de junho de 2017.

Flávio Adami, "Os Instrumentos e seus Timbres". Disponível em <<http://www.bynirsch.com.br/artigos-11-07-audionews50.shtml>>. Acessado em 15 de junho de 2017.

Francis Rumsey. Sound and Recording. Editora Taylor & Francis, 2012.

FRAUNHOFER IIS. "MP3 HISTORY DEVELOPMENT". 2017. Disponível em <<https://www.mp3-history.com/en/development.html>>. Acessado em 15 de maio de 2017

Gregory F. Maxwell em "The History of Audio: The Engineering of Sound". Disponível em: <http://www.prosoundweb.com/channels/livesound/the_history_of_pa_part_1/2/>. Acessado em 2 de novembro de 2017.

H. Helmholtz. "On the Sensations of Tone" (Dover, Nova Iorque, 1954).

Harley, Reeves Alec. "Electric signaling system." U.S. Patent Número. 2,272,070. 3 Feb. 1942.

Howard, David M. (David Martin). "Acoustics and psychoacoustics". – 4th ed. 1956.

John E. Freund. "Estatística Aplicada". 2006

John Vardalas, Ph.D. "Pulse Code Modulation: It all Started 75 Years Ago with Alec Reeves". 2012

Join, A. E. S. "History of Magnetic Recording". 1949.

Kahn, David. "Cryptology and the origins of spread spectrum: Engineers during World War II developed an unbreakable scrambler to guarantee secure communications between Allied leaders; actress Hedy Lamarr played a role in the technology." *IEEE spectrum* 21.9 (1984): 70-80.

Karlheinz Brandenburg. "MP3 AND AAC EXPLAINED" Fraunhofer Institute for Integrated Circuits FhG-IIS A, Erlangen, Alemanha. 1999

L. J. Sivian, "Absolute calibration of condenser transmitters," em *The Bell System*

Technical Journal, vol. 10, no. 1, pp. 96-115, 1931.

L. Viali. "Análise estatística Não paramétricas. Apostila de teste de Hipóteses Não Paramétricos". 2008.

Lang, Gordon. "Lossless and lossy audio formats for music". 2004

Alvarenga, Luiz Gonzaga de. "Breve tratado sobre som e música". 2008

Marcos Moecke. "Curso Técnico em Telecomunicações PCM - Modulação por Código de Pulso: Detalhes da aplicação para telefonia digital". 2006

Martin Maldovan. "Sound and heat revolutions in phononics". *Nature* 503, 209–217 (14 Novembro 2013)

S. Hochheiser. "What makes the picture talk: AT&T and the development of sound motion picture technology". *IEEE Transactions on Education*. 1992.

Museum of Magnetic Sound Recording. Disponível em <<http://museumofmagneticsoundrecording.org/>> Acessado em 02 de novembro de 2017

Nuno Fonseca. "Introdução à Engenharia de Som". 2007

Paulo Alessio. "Notas de Aula". 2013

PEARSON E.S e HARTLEY H.O. "Biometrika Tables for Statisticians", Vol.1 (1966), Tábua 8, págs. 137 e 138.

Pohlmann, Kenneth C. "Principles of digital audio". Butterworth-Heinemann, 1985

Robert Haley. The Complete Guide to High-End Audio. 2010

S. Hochheiser. "What makes the picture talk: AT&T and the development of sound motion picture technology". *IEEE Transactions on Education*. 1992.

Shaw, Thomas. *The conquest of distance by wire telephony*. 1944.

Steven Schoenherr. "History of Magnetic Recording" apresentado em IEEE Magnetics Society Seminar, UCSD, Nov. 5, 2002.

T. Shaw. "The conquest of distance by wire telephony", *Bell Syst. Tech. J.*, vol. 23, pp. 337-407, 1944.

Thomas Edison "Improvement in phonograph or speaking machines". 1878. Disponível em: <<https://www.google.com/patents/US200521>>. Acessado em 2 de novembro de 2017.

Timetoast. Disponível em: <<https://www.timetoast.com/timelines/evolution-of-microphone>>. Acessado em 2 de novembro de 2017.

Trivedi, Yatri. "What Are the Differences Between MP3, FLAC, and Other Audio Formats?". 2011.

Western Electric Company. "Fundamentals of Telephone Communication Systems". 1969.

XIPH.ORG FOUNDATION. "ABOUT XIPH". 2017. Disponível em <<https://www.xiph.org/about/>>. Acessado em 15 de Junho de 2017

Yoeri Geutskens. "SA-CD.net - Super Audio CD – FAQ". Disponível em <<http://www.sa-cd.net/faq>> Acessado em 10 de Junho 2017.

ANEXO A1 – Formulários do primeiro dia de pesquisa

Nome: Cirandeira - Elba RamalhoIdade: 23; Profissão: Música / Produtor musical

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

1- Audiá no Geral
Média → Som mais desfocada.

2- Voz, agudos da violão, flauta mais acentuada, assim como as notas de Cello.

3- Corica um gênero mais acentuada → Freqüências médias mais definidas.

4- Faixa com duração clara de compreensão, com freqüências mais definidas e com maior volume, mais equilibradas entre os elementos da música.

Idade: 34; Profissão: Designer

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(2)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(3)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(4)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

1º sei tocar + violão + é bom fado no primeiro,
 mas definido no segundo,
 Pnebi melhor a caixa no terceiro.
 E a voz percbi melhor no quarto.

1 - ruídos

2 - é definido violão +
 sentiu o batido

3 - +colado +muito ritmo bom glo +2
 +plano

4 - guitarra + definição

Idade: 30; Profissão: FARMACÊUTICA

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(5)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(1)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(2)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(4)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

A primeira faixa achei o som mais nítido e mais profundo (dá pra diferenciar muito bem os instrumentos/voz) ^{entre} de todas as faixas.

A segunda faixa parece que os sons se misturam mais. Fica mais difícil diferenciar. A terceira faixa eu não consigo diferenciar tanto da segunda faixa mas ainda achei a qualidade melhor que a segunda

A quarta faixa eu achei mais pareida com a primeira os sons são mais definidos e os graves também são bem definidos mas ainda são menos nítidos que a primeira

1 Bem nítido + profundo

2 Parece que os sons se misturam mais

3 Não consigo diferenciar da 2^a

4 Sons mais graves só são mais definidos (- do que é?)

Sons mais profundos têm como a 1^a

Idade: 26; Profissão: estudante

1. CIRANDEIRA – ELBA RAMALHO

- 1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(3)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(3)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

Xave de sintonia de frequência com o dia

Idade: 23; Profissão: Estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

(SIM

(NÃO)

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(2)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(5)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(~~4~~) 4

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

A primeira faixa possuia pouca definição

A segunda faixa possuía uma ótta definição, mais vida. Os baixos se mostraram mais presentes e os vibratos na voz do canto foram mais nítidos.

A quarta faixa também foi boa, mas os baixos apesar de presentes, não estavam com tanta definição.

Idade: 27; Profissão: Professora de Língua Portuguesa

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

SIM

NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

consegui sentir a planta doce
bem nitida

(5)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

consegui perceber um som que
preciso uma respiração profunda (5) não senti a planta da faia interna
ou as idas e vindas das vagas do mar

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

Os agudos apareceram mais

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(4)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

Idade: 23; Profissão: Internacionalista

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

(SIM) (NÃO)

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

Audiôs agudos
(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

+ grave do que ^o primeira
(mais alto)
(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

- Audiôs da primeira intona mais agudos
- Áudio da última era melhor

Idade: 26; Profissão: estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

(SIM) (NÃO)

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(5)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(3)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

Tive diferenças na melodia,

Idade: 25; Profissão: Fotógrafa

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(3)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(2)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

O quarta faixa tocada pareceu mais "abafada".
Os instrumentos pareciam mais "separados" dentro da música.

Idade: 20; Profissão: estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

SIM

NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(4)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(5)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(5)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

- 1 - mais petulante mas
- 2 - melhor queleada que 1,
mais clara na voz
- 3 - não percebi diferenças com a 2
- 4 - voz mais seca,
mais seca

ANEXO A2 – Formulários do Segundo dia de pesquisa

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: _____

Idade: 43; Profissão: 6. estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(5)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?
A clareza dos instrumentos, e de voz

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: _____

Idade: 22; Profissão: músico / estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(4)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(5)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(4)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

na segunda as instrumentos estavam mais claros que na primeira.
 principalmente as flautas

na terceira a voz está mais centrada e mais preciso com uma

performance ao vivo.

na terceira os arpejos do violão no começo está um pouco obtuso

na quinta os arpejos estão mais claros que antes e antes

na quinta os arpejos estão mais claros que antes e antes

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: Elba RamalhoIdade: 23; Profissão: Estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(2)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(3)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

A percepção dos graves e agudos.

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: _____

Idade: 24; Profissão: Programador

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?
 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(0)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(5)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(4)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

- A ultima tem frequências de graves mais notáveis que as outras
- A principal diferença tem algumas frequências de uma para as outras.

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: EduardoIdade: 26; Profissão: Eng. Eletricista

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(3)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(3)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(4)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(5)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

Segunda faixa - Dinâmica bem menor, menor sensação de distribuição espacial. Graves menos profundos emparedados à faixa 1. Som com menos brilho em geral

Terceira faixa - dinâmica e graves similares à primeira faixa e efeitos cosselos parece menos pronunciado

Quarta faixa - muito similar à terceira faixa. Diferença menor. Graves um pouco menos pronunciados

Primeira faixa - parece o mais ~~mais~~ ^{melhor}!

REPRODUÇÃO SONORA EM EQUIPAMENTOS DE ALTA QUALIDADE.

Nome: Elba RamalhoIdade: 23; Profissão: Estudante

1. CIRANDEIRA – ELBA RAMALHO

1.1 Você notou alguma diferença de qualidade entre as faixas tocadas?

 SIM NÃO

1.2 Em uma escala de zero (baixa qualidade) a cinco (alta qualidade), qual classificação você daria para a primeira faixa tocada?

(4)

1.3 Usando a mesma escala da questão anterior, qual classificação você daria para a SEGUNDA faixa tocada?

(3)

1.4 Usando a mesma escala da questão anterior, qual classificação você daria para a TERCEIRA faixa tocada?

(3)

1.5 Usando a mesma escala da questão anterior, qual classificação você daria para a QUARTA faixa tocada?

(3)

1.6 Se você notou alguma diferença de qualidade, com suas palavras, quais aspectos foram os de maior relevância?

- A maior diferença observada foi o volume entre as faixas

ANEXO A3 – CÓDIGO 1

```

canal = 1;

[original, original_fs] = audioread('ElbaRamalho.wav');
t_original = (0:size(original, 1)-1)/original_fs;
x = original(:,1);
N1 = length(x);

[mp3, mp3_fs] = audioread('ElbaRamalho128.mp3');
t_mp3 = (0:size(mp3, 1)-1)/mp3_fs;
y = mp3(:,1);
N2 = length(y);

[mp320, mp320_fs] = audioread('ElbaRamalho320.mp3');
t_mp320 = (0:size(mp320, 1)-1)/mp320_fs;
z = mp320(:,1);
N3 = length(z);

[ogg, ogg_fs] = audioread('ElbaRamalhoOGG.ogg');
t_ogg = (0:size(ogg, 1)-1)/ogg_fs;
w = ogg(:,1);
N4 = length(w);

%%%%%%%%%%%%%
%%%%%%%%%%%%%
atraso = 0.0512*44100;
diferenca_mp3_128 = (original(round(1:N1), canal) -
mp3(round(atraso:atraso+N1-1), canal));
diferenca_mp3_320 = (original(round(1:N1), canal) -
mp320(round(atraso:atraso+N1-1), canal));
diferenca_ogg = (original(round(1:N1), canal) - ogg(round(1:N1),
canal));
%%%%%%%%%%%%%
%%%%%%%%%%%%%

% figure
% spectrogram(original(:,1), round(0.5*original_fs),
round(0.5*original_fs/2), 2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontWeight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma do formato WAV')
% decimal_comma(gca, 'XY');
% xlabel('kHz')
% ylabel('Min')
% zlabel('dB')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';

% figure
% spectrogram(diferenca_mp3_128,
round(0.5*mp3_fs),

```

```

round(0.5*mp3_fs/2), 2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato MP3 - 128
kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz')
% ylabel('Min')
% zlabel('dB')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(diferenca_mp3_320, round(0.5*mp320_fs/2), 2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato MP3 - 320
kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz')
% ylabel('Min')
% zlabel('dB')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(diferenca_ogg, round(0.5*ogg_fs), round(0.5*ogg_fs/2),
2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato OGG')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(mp3(:,1), round(0.5*mp3_fs), round(0.5*mp3_fs/2), 2048,
44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')

```

```
% title('Espectrograma do formato MP3 - 128 kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;

% figure
% spectrogram(mp320(:,1), round(0.5*mp320_fs), round(0.5*mp320_fs/2),
2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma do formato MP3 - 320 kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;

% figure
% spectrogram(ogg(:,1), round(0.5*ogg_fs), round(0.5*ogg_fs/2), 2048,
44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma do formato OGG')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
```

ANEXO A4 – CÓDIGO 2

```

round(0.5*mp3_fs/2), 2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato MP3 - 128
kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz')
% ylabel('Min')
% zlabel('dB')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(diferenca_mp3_320, round(0.5*mp320_fs/2), 2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato MP3 - 320
kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz')
% ylabel('Min')
% zlabel('dB')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(diferenca_ogg, round(0.5*ogg_fs), round(0.5*ogg_fs/2),
2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma da diferença entre WAV e o formato OGG')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
%
% figure
% spectrogram(mp3(:,1), round(0.5*mp3_fs), round(0.5*mp3_fs/2), 2048,
44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')

```

```
% title('Espectrograma do formato MP3 - 128 kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;

% figure
% spectrogram(mp320(:,1), round(0.5*mp320_fs), round(0.5*mp320_fs/2),
2048, 44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma do formato MP3 - 320 kbps')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;

% figure
% spectrogram(ogg(:,1), round(0.5*ogg_fs), round(0.5*ogg_fs/2), 2048,
44100)
% grid on
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 24, 'fontweight',
'bold')
% set(gcf,'color','white')
% title('Espectrograma do formato OGG')
% decimal_comma(gca, 'XY');
% xlabel('kHz', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% ylabel('Min', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% zlabel('dB', 'FontName', 'Times New Roman', 'FontSize', 24,
'fontweight', 'bold')
% caxis([-156.54 -19.42])
% c = colorbar;
% c.Label.String = 'dB/kHz';
% c.Label.FontSize = 24;
```