

INSTITUTO FEDERAL DE SANTA CATARINA

LUÍSA MACHADO

Estudo de codificadores de áudio

São José - SC

junho/2019

ESTUDO DE CODIFICADORES DE ÁUDIO

Pré-Projeto de trabalho de conclusão de curso apresentado à Coordenadoria do Curso de Engenharia de Telecomunicações do campus São José do Instituto Federal de Santa Catarina para a aprovação do tema perante banca na disciplina de TCC1.

Orientador: Marcos Moecke

Coorientador: Roberto de Matos

São José - SC

junho/2019

RESUMO

Uma das etapas mais importantes antes de realizar uma transmissão é a compressão do arquivo, pois com essa etapa é possível otimizar o tempo de realização do processo. Todavia um áudio é uma informação muito sensível, na qual a perda de parte dela pode comprometer totalmente a qualidade do arquivo. Sabendo disso, o objetivo deste trabalho é realizar e apresentar um estudo sobre alguns codificadores de áudio específicos e as suas principais características, simulando testes visando identificar qual o melhor codificador a ser utilizado em determinados cenários.

Palavras-chave: Codificação de áudio. Codec. FLAC. Vorbis. Opus.

LISTA DE ILUSTRAÇÕES

Figura 1 – Sistema auditivo	11
Figura 2 – Limiar de sensibilidade da audição humana	12
Figura 3 – Curvas de ponderação.	16

LISTA DE TABELAS

Tabela 1 – Cronograma das atividades previstas	18
----------------------------------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

AAC <i>Advanced Audio Coding</i>	18
AOSV <i>Avaliação Objetiva de Sinais de Voz</i>	10
bps <i>bits por segundo</i>	18
CBR <i>constant bitrate</i>	14
CELT <i>Constrained Energy Lapped Transform</i>	10
codec <i>codificador/decodificador</i>	9
dB <i>decibéis</i>	11
FFT <i>fast Fourier transform</i>	15
FLAC <i>Free Lossless Audio Codec</i>	9
ITU-R <i>International Telecommunication Union - Radiocommunication</i>	15
LP <i>Linear Prediction</i>	14
MDCT <i>Modified Discrete Cosine Transform</i>	14
MSE <i>Mean-squared Error</i>	15
MOQA <i>Medida Objetiva da Qualidade de Áudio</i>	10
MOS <i>mean opinion score</i>	15
PEAQ <i>Perceptual Evaluation of Audio Quality</i>	15
PSNR <i>Peak Signal-to-Noise Ratio</i>	7

SNR <i>Signal-to-Noise Ratio</i>	15
SPL <i>sound pressure level</i>	11
TCC Trabalho de Conclusão de Curso	10
THD <i>total harmonic distortion</i>	15
VBR <i>variable bit rate</i>	14
VoIP <i>Voice over Internet Protocol</i>	14

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Objetivo geral	10
1.2	Objetivos específicos	10
2	FUNDAMENTAÇÃO TEÓRICA	11
2.1	Sistema auditivo	11
2.1.1	Características psicoacústicas	11
2.2	Sinal de áudio	13
2.3	Codificador de áudio	13
2.3.1	Codificadores selecionados	13
2.4	Métodos de avaliação da qualidade do áudio	15
2.5	<i>Peak Signal-to-Noise Ratio</i> (PSNR)	15
2.6	Curva psofométrica	15
3	PROPOSTA	17
3.1	Estudo teórico	17
3.2	Ambiente de testes	17
3.3	Simulação	18
3.4	Análise de resultados	18
3.5	Cronograma	18
	REFERÊNCIAS	19

1 INTRODUÇÃO

No mundo das tecnologias há uma crescente busca por redução de custos, por exemplo, em uma transmissão de dados quanto maior for o tamanho desse dado maior será o tempo gasto para transmiti-lo, resultando em maior custo, por esta razão foram criados os codificadores de dados. De modo geral, um codificador tem como objetivo diminuir o tamanho de um arquivo, reduzindo o tempo de sua transmissão e, também, a quantidade de memória necessária para fazer o seu armazenamento, o que implica na redução de custos.

Com esse mesmo objetivo surgiram também os codificadores ou compressores de áudio, que de acordo com Spanias (2007), são algoritmos criados com o intuito de obter representações digitais mais compactas dos sinais de áudio com alta fidelidade para realizar a sua transmissão ou o seu armazenamento de forma mais eficiente. Sendo assim, esses compressores tem como principal objetivo reduzir ao máximo o número de bits utilizados na representação do sinal visando manter a sua qualidade, de modo que seja praticamente impossível distinguir o áudio de entrada original da saída codificada.

Codificadores de áudio podem ser classificados conforme o seu tipo de codificação, que pode ser *lossless* (sem perdas) ou *lossy* (com perdas) (SPANIAS, 2007). Os codificadores *lossless*, segundo a tradução do nome indica, não possui perdas geradas no momento da compressão, ou seja, o áudio codificado é totalmente fiel ao original. Este tipo de codificador/decodificador (*codec*) trabalha em cima das redundâncias dos sinais no áudio, então, o tamanho do arquivo no final da compressão vai depender apenas da quantidade de informação redundante existente no áudio, sendo assim, quanto mais redundância possuir, menor será o tamanho do áudio codificado.

No entanto, nos codificadores *lossy* existem perdas causadas pela compressão, porém seu principal objetivo é que essas perdas devem alterar o mínimo possível de informação, de modo que o ouvido humano seja incapaz de detectar as diferenças entre o original e o comprimido. Por esta razão, para produzir um bom codificador *lossy* é necessário conhecer primeiro o comportamento da audição humana, principalmente as suas características psicoacústicas, que auxiliam a detectar informações naturalmente ignoradas pelo ouvido humano.

Para a realização desse estudo foram selecionados três codificadores de áudio, o **FLAC**, o **Vorbis** e o **Opus**. Todos esses codificadores foram implementados pela **Xiph.Org Foundation**, que é um grupo sem fins lucrativos com o objetivo de desenvolver *softwares* de código aberto e gratuitos para diversas aplicações.

O *Free Lossless Audio Codec (FLAC)*, de acordo com a tradução do nome, é um codificador sem perdas (XIPH.ORG, 2014). A decisão de usar este *codec* se deve ao fato de possuir um rápido processamento, ser suportado por diversos equipamentos e para ser utilizado como uma base de comparação de qualidade, em virtude da sua saída não sofrer alterações perceptíveis ao ouvido humano, sendo assim o áudio codificado é considerado idêntico ao original.

O **Vorbis**, também conhecido como **Ogg Vorbis**, é um *codec* com perdas que é otimizado para ser usado para comprimir tanto a fala quanto a música (XIPH.ORG, 2015). Uma curiosidade sobre o **Vorbis**, é que o serviço de *streaming* do **Spotify** usa na compressão de suas músicas e, além dele, alguns jogos de *videogame* também o utilizam.

O último codificador a ser trabalhado é o **Opus** que, assim como o **Vorbis**, é um *codec* com perdas. Este é um híbrido de outros dois codificadores mais antigos, o **SILK** que é do **Skype** e é otimizado para

codificar a fala, e o *Constrained Energy Lapped Transform* (CELT) que é da Xiph.Org Foundation e foi desenvolvido para tratar arquivos de música (XIPH.ORG, 2017). Portanto, o Opus é, também, um compressor que trata fala e música obtendo uma boa qualidade em ambos.

Existem alguns trabalhos que segue uma linha de estudos semelhante a deste projeto. Um desses trabalhos é a tese de doutorado do Barbedo (2004), “Avaliação objetiva de qualidade de sinais de áudio e voz”. Nela o autor aborda a dificuldade em avaliar objetivamente a qualidade de um áudio e, a partir disso, desenvolve dois métodos objetivos de avaliação. O primeiro método é a Medida Objetiva da Qualidade de Áudio (MOQA), desenvolvido para avaliar a qualidade de música e voz entre as frequências de 20 Hz a 20 kHz, este método foi patenteado pelo autor. Enquanto o segundo método, Avaliação Objetiva de Sinais de Voz (AOSV), foi implementado para analisar um sinal de voz na mesma faixa de frequência da telefonia, entre 300 Hz a 3400 Hz, e, diferentemente do primeiro método, este não foi patenteado pelo autor.

Há também o trabalho “Análise de Qualidade de Áudio Objetiva e Subjetiva em vários formatos digitais”, que é um Trabalho de Conclusão de Curso (TCC) dos autores Tanaka, Barbosa e Kimura (2017), que busca uma forma objetiva para avaliar a qualidade de um áudio. Nesse TCC, os autores fizeram testes subjetivos, selecionando um grupo de pessoas para avaliar os áudios, e, posteriormente, analisaram matematicamente os sinais de áudio procurando obter uma métrica objetiva.

O presente documento está organizado em mais duas partes além deste capítulo. No próximo capítulo são apresentados os principais conceitos que serão utilizados no decorrer deste projeto. Enquanto, o Capítulo 3 descreve a proposta deste trabalho e apresenta o cronograma previsto das atividades.

1.1 Objetivo geral

Estudar e testar o funcionamento dos codificadores de áudio, procurando identificar métricas objetivas para avaliar o desempenho de cada um em relação a qualidade do áudio ajustando alguns de seus parâmetros de codificação, por exemplo: nível de codificação e taxa de bits.

1.2 Objetivos específicos

- Estudar os codificadores de áudio FLAC, Vorbis e Opus;
- Estudar os algoritmos e os parâmetros de ajuste de cada codificador;
- Encontrar cada codec pronto para realizar os testes de funcionamento;
- Selecionar uma base de dados com áudios para utilizar nos experimentos;
- Avaliar o desempenho da qualidade do áudio para os diferentes parâmetros de codificação;
- Conhecer as métricas de avaliação da qualidade do áudio existente;
- Desenvolver uma nova métrica de avaliação da qualidade do áudio utilizando a curva psfométrica em conjunto com a PSNR.

2 FUNDAMENTAÇÃO TEÓRICA

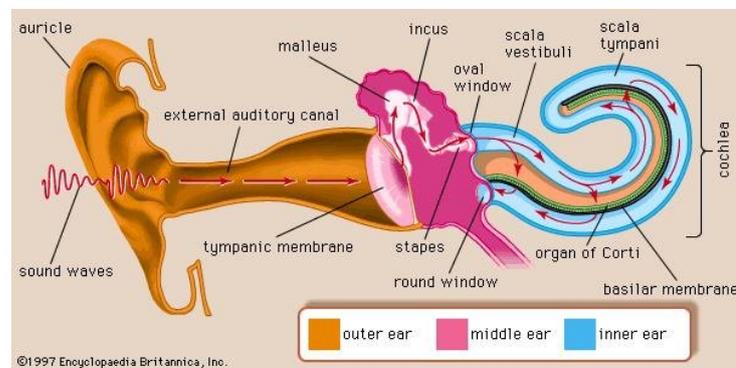
2.1 Sistema auditivo

O sistema auditivo humano é responsável por transformar as vibrações sonoras em impulsos nervosos que são enviados para o cérebro, o qual irá interpretar como um som (HAWKINS, 2018). Todo esse processo de tratamento do som é conhecido como audição.

A Figura 1 apresenta o sistema auditivo humano e o seu funcionamento. Observando a direção das setas vermelhas, a audição inicia quando as ondas sonoras (*sound waves*) entram no ouvido externo (*outer ear*) através do canal auditivo externo chegando na membrana timpânica, (*tympanic membrane*), mais conhecida como tímpano. O ouvido externo representa toda a parte que está em amarelo na imagem.

Continuando o processo de audição, o tímpano vibra na mesma frequência da onda sonora propagando essa vibração para os ossículos do ouvido médio (*middle ear*). O ouvido médio é toda a parte rosa da Figura 1. Esses ossículos são três pequenos ossos localizados enfileirados no ouvido médio, eles são chamados de martelo (*malleus*), bigorna (*incus*) e estribo (*stapes*). O som se propaga através desses ossos em direção ao ouvido interno (*inner ear*), a parte em azul da imagem, criando ondas nos fluidos da cóclea vibrando a membrana basilar estimulando as células sensoriais do órgão de Corti (*organ of Corti*), que, por sua vez, transmite a informação obtida para o cérebro.

Figura 1 – Sistema auditivo



Fonte: Hawkins (2018)

2.1.1 Características psicoacústicas

As características psicoacústicas da audição humana definem a percepção auditiva (SPANIAS, 2007). Essas características são exploradas pela maioria dos codificadores de áudio para obter uma maior compressão da informação sem afetar a sua qualidade. Estas características quando usadas por um codificador possibilitam que sejam eliminadas apenas as informações que não são detectadas pelo ouvido humano.

Um conceito necessário para avançar nesta subseção é o *sound pressure level* (SPL), que pode ser traduzido como nível de pressão sonora. O SPL é uma métrica padrão utilizada para mensurar a intensidade em decibéis (dB) de um som em relação ao nível de referência internacional. Este nível é

representado pela constante P_0 mostrada na [Equação 2.1](#) (SPANIAS, 2007).

$$P_0 = 2 \cdot 10^5 \text{ N/m}^2 \quad (2.1)$$

A seguir serão descritas as seguintes características psicoacústicas: limiar absoluto de audição, banda crítica e mascaramento simultâneo.

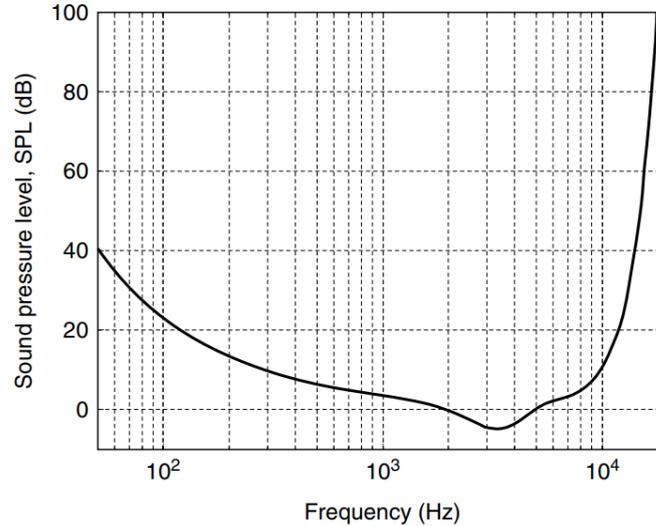
Limiar absoluto de audição

O limiar absoluto de audição define a quantidade de energia necessária para poder detectar um tom puro em um ambiente sem ruído (SPANIAS, 2007). A unidade de medida utilizada para representar esse limite é o **dB SPL**. Este limiar é calculado com base no nível de referência P_0 , portanto o limiar em silêncio mede aproximadamente 0 **dB SPL** e o limiar de dor é igual ou superior 140 **dB SPL**.

A dependência entre o limiar absoluto de audição e a frequência foi descoberta por Fletcher, em 1940, após estudos sobre a percepção auditiva (SPANIAS, 2007). A [Equação 2.2](#) apresenta uma fórmula que se aproxima do limiar de audição de um jovem e a [Figura 2](#) exibe o limiar de sensibilidade da audição calculado com base nessa equação no domínio da frequência.

$$T_q(f) = 3,64(f/1000)^{-0,8} - 6,5e^{-0,6(f/1000-3,3)^2} + 10^{-3}(f/1000)^4 \quad (\text{dB SPL}) \quad (2.2)$$

Figura 2 – Limiar de sensibilidade da audição humana



Fonte: Spanias (2007), pag. 114.

Bandas críticas

Como visto no início desta seção ([seção 2.1](#)), é na cóclea que acontece a conversão da frequência da onda sonora em impulsos nervosos. Segundo Spanias (2007), a cóclea pode ser considerada um banco de filtros passa banda, por isso foi criada a função largura de banda crítica, [Equação 2.3](#), para quantificar as bandas passantes da cóclea.

$$BW_c(f) = 25 + 75[1 + 1.4(f/1000)^2]^{0.69} \quad (\text{Hz}) \quad (2.3)$$

Mascaramento simultâneo

De acordo com Spanias (2007), o mascaramento simultâneo é um processo em que não é possível ouvir um som devido a inserção de um outro som simultaneamente. O mascaramento ocorre no sistema auditivo quando dois ou mais sons são captados, porém um deles possui características que anulam o outro som, por exemplo, quando dois sons com frequências muito próximas chegam ao ouvido, aquele com maior amplitude vai se sobrepor ao de menor amplitude, dando a sensação de que existe apenas um som.

2.2 Sinal de áudio

Um som ocorre a partir da vibração de algum objeto gerando ondas sonoras, por exemplo, a vibração das cordas vocais no momento da fala ou a vibração de algum instrumento musical. Essas ondas sonoras são conhecidas também como um sinal de áudio. O sinal de áudio pode ser gravado com o auxílio de um microfone, no qual as ondas sonoras são convertidas em sinais elétricos para serem armazenadas ou transmitidas (BRITANNICA, 2019).

2.3 Codificador de áudio

Os codificadores de áudio, também chamados de compressores de áudio, surgiram com a necessidade de reduzir ao máximo a quantidade de bits na representação de um áudio sem prejudicar a sua qualidade facilitando a sua transmissão ou o seu armazenamento (SPANIAS, 2007). Considera-se um bom codificador, aquele em que não há diferença audível entre o áudio original e o codificado.

Segundo Spanias (2007), uma forma de classificar os codificadores de áudio é de acordo com a sua técnica de tratamento do áudio. Essas técnicas de codificação podem ser do tipo que usa predição linear, transformada, sub-banda ou senoidal.

Outra forma de classificar os *codecs* é se existem perdas na codificação ou não (SPANIAS, 2007). Esse é o modo de classificação mais utilizado. Os codificadores sem perdas são conhecidos pelo termo em inglês *lossless*, enquanto os com perdas são chamados de *codecs lossy*.

Codificadores *lossless* não possuem perdas de compressão mantendo a qualidade do áudio original. Apesar de não alterem a qualidade do áudio, compressores deste tipo tem baixa taxa de compactação, ou seja, reduzem pouco o tamanho do áudio, pois eles apenas suprimem seus sinais redundantes, de modo que haverá maior compressão nos áudios que possuírem mais redundâncias.

No caso dos compressores *lossy* ocorrem perdas causadas pela codificação, porém estas perdas devem alterar o mínimo possível da qualidade do áudio. Os melhores *codecs* são aqueles que mais exploram as características psicoacústicas, sendo assim, as perdas ocorridas no processo são de informações que o ouvido humano não consegue detectar.

2.3.1 Codificadores selecionados

Com a finalidade de otimizar os estudos, foram selecionados apenas codificadores de código aberto e disponíveis gratuitamente para a instalação. Os codificadores escolhidos foram o **FLAC**, o **Vorbis** e o **Opus**, todos foram produzidos pela **Xiph.Org Foundation**. Essa fundação é grupo sem fins lucrativos focado em desenvolver projetos de código aberto e gratuito.

FLAC

O *Free Lossless Audio Codec* (**FLAC**) é um codificador otimizado para comprimir áudio sem perdas de qualidade, sendo classificado como um *codec lossless* (XIPH.ORG, 2014). Para obter uma

melhor compressão, este codificador aproveita o alto grau de correlação entre as amostras de um áudio, ou seja, trabalha suprimindo suas amostras redundantes (XIPH.ORG, 2014).

O **FLAC** é um codificador amplamente suportado, sendo suportado pela maioria dos sistemas operacionais, como **Windows** e **Linux** (XIPH.ORG, 2014). Além disso, os dois métodos principais do seu algoritmo, a codificação e a decodificação, e toda sua especificação são totalmente abertos para uso de qualquer público.

Considerado um **codec** assimétrico, tem seu método de codificação dentre os mais complexos codificadores sem perdas, enquanto a sua decodificação é bem mais simples, obtendo um dos menores tempo de processamento em relação aos **codecs** de mesmo tipo (XIPH.ORG, 2014).

Vorbis

O **Vorbis** é classificado como um **codec** do tipo com perdas de uso geral. Este codificador é projetado para trabalhar tanto com altas quanto com baixas taxas de amostragem, ou seja, consegue tratar a voz e a música de forma otimizada (XIPH.ORG, 2015). Sua codificação utiliza características psicoacústicas que, por consequência, aumentam o nível de complexidade do seu algoritmo, enquanto isso, sua decodificação possui baixa complexidade.

O **Vorbis** trabalha com *variable bit rate* (**VBR**), que significa que a taxa de bits é variável se ajustando conforme a necessidade, e seus pacotes também são ajustáveis, pois não possuem tamanho definido (XIPH.ORG, 2015). Este **codec** possui dois tipos, o **Vorbis I**, que usa a *Modified Discrete Cosine Transform* (**MDCT**) para preparar o áudio para a codificação, e o **Vorbis II**, que implementa um banco de filtros *wavelet* híbrido substituindo a **MDCT** (XIPH.ORG, 2015). Este projeto vai trabalhar apenas com o **Vorbis I**.

Este codificador foi projetado para ser utilizado em conjunto com algumas ferramenta de transmissão, isto significa que tem o objetivo de preparar os áudios para serem transmitidos. O serviço de *streaming* de música do **Spotify**, por exemplo, utiliza o **Vorbis** para preparar as músicas antes de transmitir. Outro exemplo, são alguns jogos de *videogame* que também utilizam este compressor.

Opus

O **Opus** é um codificador de áudio com perdas definido como um **codec** interativo, que pode ser utilizado em aplicações *Voice over Internet Protocol* (**VoIP**), videoconferências e bate-papo em jogos (XIPH.ORG, 2017). Este compressor é baseado em outros dois existentes: o **SILK**, do **Skype**, e o *Constrained Energy Lapped Transform* (**CELT**), da **Xiph.Org Foundation**.

O **SILK** é um codificador baseado em *Linear Prediction* (**LP**), que o torna mais eficiente no tratamento de baixas frequências, ou seja, é um ótimo codificador de voz (XIPH.ORG, 2017). Assim como o **Vorbis**, este também trabalha com taxas variáveis (**VBR**).

Ao contrário do **SILK**, o **CELT** é um **codec** que utiliza a *Modified Discrete Cosine Transform* (**MDCT**) e, este é um codificador muito mais eficiente no tratamento de altas frequências, por exemplo: uma música (XIPH.ORG, 2017). O **CELT** trabalha com *constant bitrate* (**CBR**), ou seja, este **codec** opera com taxa de bits constante em toda a sua codificação.

O **Opus** possui uma camada que utiliza uma versão modificada do **SILK** e outra que se baseia no **CELT**, ambas as camadas conseguem operar tanto com **VBR** quanto com **CBR** (XIPH.ORG, 2017). Por este motivo, o **Opus** consegue obter uma boa qualidade de codificação de áudios de voz e música.

2.4 Métodos de avaliação da qualidade do áudio

Avaliação da qualidade de um áudio é uma tarefa bem complicada, porém muito importante para validar o funcionamento de um codificador de áudio. Essa avaliação pode ser feita de duas formas: objetiva ou subjetiva.

Medidas objetivas, como a *Signal-to-Noise Ratio* (SNR) e a *total harmonic distortion* (THD), comumente utilizadas na análise de sinais em geral, não obtêm bons resultados quando aplicados na avaliação de um áudio. Os métodos objetivos que funcionam para avaliar a qualidade de um áudio são aqueles que melhor reproduzem a percepção auditiva humana.

A *International Telecommunication Union - Radiocommunication* (ITU-R) define na recomendação BS.1387-1 o modelo objetivo *Perceptual Evaluation of Audio Quality* (PEAQ), esse possui duas versões, uma baseada em *fast Fourier transform* (FFT) e a outra baseada em FFT em conjunto com banco de filtros.

Barbedo (2004) propôs em sua tese de doutorado dois métodos objetivos de avaliação, a Medida Objetiva da Qualidade de Áudio (MOQA) e Avaliação Objetiva de Sinais de Voz (AOSV), usando como base o modelo PEAQ. O primeiro analisa a qualidade do áudio entre as frequências de 20 Hz a 20 kHz. Enquanto, o segundo avalia a qualidade do áudio com as frequências na faixa da telefonia, sendo ela entre 300 Hz a 3400 Hz.

Segundo Palomar et al. (2008), no caso da avaliação da qualidade de um áudio, os métodos subjetivos são os mais confiáveis, porém possuem um alto custo de realização e precisam de muito tempo de execução. A *mean opinion score* (MOS) é o principal método de avaliação subjetiva, ela consiste em realizar testes de audição subjetivos. Para sua realização é necessário organizar um grupo considerável de pessoas treinadas para ouvir e avaliar a qualidade do áudio com uma nota. Os resultados obtidos no final dos testes subjetivos são utilizados para formar um valor médio para quantificar a qualidade do áudio.

2.5 PSNR

A *Peak Signal-to-Noise Ratio* (PSNR) é uma métrica geralmente utilizada para analisar a qualidade visual de imagens comprimidas. A PSNR depende matematicamente da *Mean-squared Error* (MSE), essa é calculada através da Equação 2.4, enquanto a PSNR é calculada pela Equação 2.5 (CHANDLER; HEMAMI, 2007).

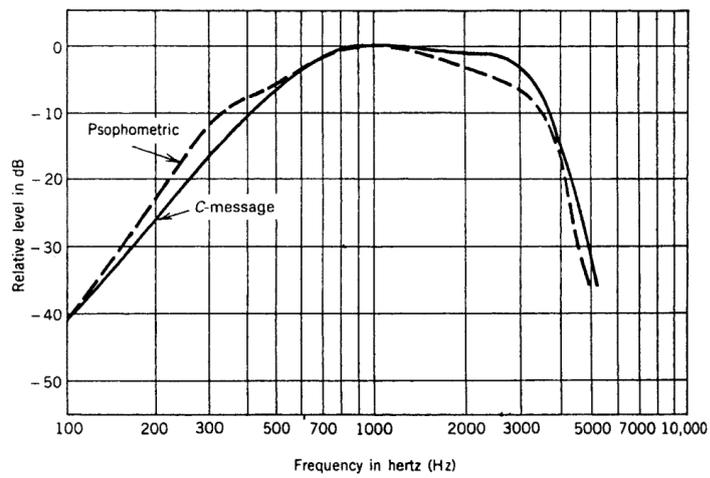
$$MSE = \frac{1}{N} \sum_{i=1}^N E_i^2 \quad (2.4)$$

$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \quad (\text{dB}) \quad (2.5)$$

2.6 Curva psfométrica

Existem várias curvas de ponderação, algumas dessas são utilizadas na telefonia para determinar o nível de ruído presente na fala. A curva psfométrica é uma dessas curvas. Além da curva psfométrica, existem a curva de ponderação *C-message* que também é usada na telefonia. A Figura 3 apresenta essas duas curvas, sendo que a tracejada representa a curva psfométrica, enquanto a contínua mostra a curva *C-message* (FREEMAN, 2004).

Figura 3 – Curvas de ponderação.



Fonte: Freeman (2004), pag. 153.

3 PROPOSTA

Para este projeto é proposto a realização de um estudo sobre o funcionamento de codificadores de áudio. Também serão feitas simulações práticas para analisar o desempenho dos [codecs](#) através de seus resultados. Para o cumprimento desta proposta, o presente documento divide o projeto em quatro etapas: o estudo teórico, o ambiente de testes, a simulação e a análise dos resultados.

3.1 Estudo teórico

O principal objetivo deste projeto é o estudo de codificadores de áudio. Para isto foram selecionados três codificadores, sendo um deles classificado como [codec](#) sem perdas, enquanto os outros dois como [codecs](#) com perdas. Portanto, é fundamental entender as diferenças nos princípios gerais de funcionamento de cada tipo de codificador.

Antes de trabalhar manipulando áudios, é necessário conhecer o comportamento da audição humana, ou seja, como o ouvido capta os sons a sua volta. Também será realizado um estudo sobre as características psicoacústicas humanas para compreender o funcionamento dos codificadores de áudio *lossy* escolhidos para este trabalho. Alguns exemplos dessas características são: o mascaramento simultâneo, limiar absoluto de audição e análise de frequência de banda crítica ([SPANIAS, 2007](#)).

A maior parte desses estudos será concentrada em três codificadores: o [FLAC](#), o [Vorbis](#) e o [Opus](#), visando identificar os principais parâmetros de codificação destes. Cada codificador será avaliado através de experimentos nos quais esses parâmetros serão alterados.

Algumas métricas existentes de avaliação da qualidade do áudio serão analisadas para serem utilizadas na etapa de análise dos resultados obtidos. Também será proposto uma métrica que considere a curva psfométrica em conjunto com a [PSNR](#).

3.2 Ambiente de testes

Esta fase consiste em preparar o ambiente de realização dos testes dos codificadores de áudio. No final desta etapa é necessário ter os codificadores funcionais e uma base de arquivos de áudio não codificados para utilizar nos testes. As principais ferramentas de *hardware* e *software* necessárias são:

- computador com sistema operacional [Linux](#);
- pacote [FLAC](#);
- pacote [Vorbis](#);
- pacote [Opus](#);
- [MatLab 2015](#);
- base de dados com áudios.

3.3 Simulação

Esta fase consiste em executar os três **codecs** de áudio escolhidos alterando seus parâmetros de codificação, tais como: o nível de codificação e a taxa de bits, que é medida em bits por segundo (**bps**), procurando identificar as principais mudanças ocorridas no áudio em determinado cenário.

Os testes dos codificadores serão realizados utilizando a base de arquivos, que possui 40 amostras de áudio, do [Kamedo \(2014\)](#), ele utilizou essa base na realização de testes subjetivos de audição. Esses testes foram feitos às cegas, ou seja, dividiram as amostras em dois grupos e distribuíram aleatoriamente entre os ouvintes sem que saibam qual o **codec** foi utilizado no processamento de cada áudio. Os codificadores avaliados neste experimento foram o **Opus**, o *Advanced Audio Coding (AAC)* e o **Ogg Vorbis** com taxa de bits igual a 96 *kbps*, e o **MP3** a um taxa de 128 *kbps*.

3.4 Análise de resultados

Para analisar a qualidade dos áudios resultantes da compressão serão utilizadas algumas das métricas anteriormente descritas, visto que as métricas do [Barbedo \(2004\)](#) não estão disponíveis para uso. Também será validada a métrica proposta comparando-a com os resultados obtidos com outras já consolidadas como a *mean opinion score (MOS)* e a *Perceptual Evaluation of Audio Quality (PEAQ)*.

Outro ponto que deve ser analisado nesta etapa, é o funcionamento dos codificadores de áudio a partir dos resultados obtidos na execução da fase anterior, visando identificar qual obteve o melhor desempenho em cada cenário específico.

3.5 Cronograma

O cronograma das atividades previstas para o desenvolvimento deste projeto está apresentado na [Tabela 1](#). A lista exibida em seguida mostra o nome de cada etapa citada no cronograma.

Tabela 1 – Cronograma das atividades previstas

Etapa	Mês							
	mai	jun	jul	ago	set	out	nov	dez
1	√	√						
2		√	√	√	√	√		
3				√	√	√		
4					√	√	√	
5						√	√	√

1. Definição do tema do projeto
2. Estudo teórico
3. Ambiente de testes
4. Simulação
5. Análise de resultados

REFERÊNCIAS

- BARBEDO, J. G. A. *Avaliação objetiva de qualidade de sinais de áudio e voz*. Tese (Doutorado) — Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, jul 2004. Citado 3 vezes nas páginas 10, 15 e 18.
- BRITANNICA, E. *Microphone*: Electroacoustic device. Encyclopædia Britannica, Inc, 2019. Disponível em: <<https://www.britannica.com/technology/microphone-electroacoustic-device>>. Acesso em: 29 jun. 2019. Citado na página 13.
- CHANDLER, D. M.; HEMAMI, S. S. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE, IEEE Transactions on Image Processing*, v. 16, p. 2284–2298, 2007. Citado na página 15.
- FREEMAN, R. L. *Telecommunication system engineering*. Hoboken, N.J: Wiley-Interscience, 2004. ISBN 0471451339. Citado 2 vezes nas páginas 15 e 16.
- HAWKINS, J. E. *Human ear*: Anatomy. Encyclopædia Britannica, Inc, 2018. Disponível em: <<https://www.britannica.com/science/ear>>. Acesso em: 24 jun. 2019. Citado na página 11.
- KAMEDO. *Results of the public multiformat listening test (July 2014)*. 2014. Disponível em: <<https://listening-test.coresv.net/results.htm>>. Acesso em: 27 jun. 2019. Citado na página 18.
- PALOMAR, D. et al. Objective assessment of audio quality. In: *IET Irish Signals and Systems Conference (ISSC 2008)*. [S.l.: s.n.], 2008. p. 37–42. ISSN 0537-9989. Citado na página 15.
- SPANIAS, A. *Audio signal processing and coding*. Hoboken, N.J: Wiley-Interscience, 2007. ISBN 9780471791478. Citado 5 vezes nas páginas 9, 11, 12, 13 e 17.
- TANAKA, A. Y.; BARBOSA, E. R.; KIMURA, R. S. Y. *Análise de Qualidade de Áudio Objetiva e Subjetiva em vários formatos digitais*. Monografia (TCC) — Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, 2017. Citado na página 10.
- XIPH.ORG, F. *FLAC - Free Lossless Audio Codec*. 2014. Disponível em: <<https://xiph.org/flac/index.html>>. Acesso em: 27 jun. 2019. Citado 3 vezes nas páginas 9, 13 e 14.
- XIPH.ORG, F. *Vorbis I specification*. 2015. Disponível em: <https://xiph.org/vorbis/doc/Vorbis_I_spec.html>. Acesso em: 27 jun. 2019. Citado 2 vezes nas páginas 9 e 14.
- XIPH.ORG, F. *Opus Codec*. 2017. Disponível em: <<http://opus-codec.org/>>. Acesso em: 27 jun. 2019. Citado 2 vezes nas páginas 10 e 14.