

Universidade Estadual Paulista
Instituto de Biociências, Letras e Ciências Exatas
Departamento de Ciência da Computação e Estatística

Gabriel Andrey Perego Nunes

Avaliação de sinais acústicos para o pré-diagnóstico de
covid-19

São José do Rio Preto - SP

2021

Gabriel Andrey Perego Nunes

Avaliação de sinais acústicos para o pré-diagnóstico
de covid-19

Monografia apresentada ao Programa de
graduação em Ciência da Computação da
UNESP para obtenção do título de Bacharel.

Orientador: Prof. Dr. Rodrigo Capobi-
anco Guido

São José do Rio Preto - SP

2021

Ficha catalográfica elaborada pelo Serviço de Biblioteca do IBILCE/UNESP

Gabriel Andrey Perego Nunes

Avaliação de sinais acústicos para o pré-diagnóstico de covid-19
Avaliação de sinais acústicos para o pré-diagnóstico de covid-19. /
Gabriel Andrey Perego Nunes. São José do Rio Preto, 2021
45 p.

Trabalho de conclusão de curso Bacharelado - Ciência da
Computação - Universidade Estadual Paulista - UNESP, Instituto de
Biociências Letras e Ciências Exatas, São José do RIO Preto

Orientador: Rodrigo Capobianco Guido.

1. Processamento de sinais. 2. Transformada de Wavelets. 3. Operador
de Teager 4. Covid-19

I. Avaliação de sinais acústicos para o pré-diagnóstico de covid-19.

Gabriel Andrey Perego Nunes

Avaliação de sinais acústicos para o pré-diagnóstico
de covid-19

Monografia apresentada ao Programa de
graduação em Ciência da Computação da
UNESP para obtenção do título de Bacharel.

Comissão Examinadora:

Prof. Dr. Rodrigo Capobianco Guido
UNESP – Câmpus de São José do Rio Preto

Orientador

Profa. Dra. Renata Spolon Lobado

UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Geraldo Francisco Donega Zafalon

UNESP – Câmpus de São José do Rio Preto

São José do Rio Preto - SP

06 de Janeiro de 2022

Dedico a minha família e amigos.

Agradecimentos

A Deus;

À Universidade Estadual Paulista por tornar possível a existência desta monografia;

Ao meu orientador Prof. Dr. Rodrigo Capobianco Guido por toda orientação e dedicação durante o desenvolvimento do trabalho;

Aos professores doutores Renata Spolon Lobado e Geraldo Francisco Donega Zafalon por aceitarem fazer parte da banca e contribuírem no desenvolvimento do trabalho;

Aos meus pais, João Batista Nunes e Alexandra Maria Perego, pelo constante apoio incondicional, sem os quais a minha formação acadêmica não seria possível;

À minha irmã, Andressa Paola Perego Nunes, pelo suporte e apoio nos estudos;

À Bianca Lobo pelo suporte, apoio e carinho durante todo o desenvolvimento do trabalho.

“Sonhos determinam o que você quer. Ação determina o que você conquista.”

Aldo Novak

Resumo

NUNES, G.A.P. *Avaliação de sinais acústicos para o pré-diagnóstico de covid-19*. 2021. 45p. TCC UNESP 2021.

Os sinais produzidos pelo corpo humano têm sido objeto de pesquisa na busca de indicadores relacionados às mais diversas doenças. Com a explosão da pandemia do vírus Sars-Cov-2 em 2020, surge a oportunidade ímpar do processamento de sinais contribuir na identificação da patologia. Os primeiros trabalhos relacionados ao novo vírus se basearam em estudos recentes de outras doenças que dispõem de sintomas similares aos causados pelo Coronavírus. Dado que a doença apresenta sintomas comuns em pacientes, e que esses sintomas podem causar algum grau de comprometimento no sistema fonador e gerar vibrações irregulares, o presente trabalho teve por objetivo identificar assinaturas da covid-19 em áudios de vogais sustentadas por meio da análise da energia presente nas ondas sonoras. Utilizando Transformada de Wavelets e o operador de Teager, foi possível, com métodos estatísticos, identificar características únicas de indivíduos saudáveis e de infectados que podem ser utilizadas na criação de classificadores para identificação da patologia. Foi descoberto que pessoas infectadas apresentam um valor médio no vetor resultante de 0,00034 enquanto pessoas saudáveis obtiveram um valor de 0,01665. Foi possível também analisar a diferença entre os valores mínimos e máximos encontrados nos vetores.

Palavras-chave: Processamento de sinais. Transformada de Wavelets. Identificação de patologia. Operador de Teager.

Abstract

NUNES, G.A.P. *Avaliação de sinais acústicos para o pré-diagnóstico de covid-19*. 2021. 45p. TCC UNESP 2020.

The signals produced by the human body have been the subject of researches on indicators related to different diseases. With the outbreak of the Sars-Cov-2 pandemic in 2020, there is a unique opportunity for signal processing to contribute to the identification of the pathology. The first papers on the new virus were based on recent studies of other diseases that have similar symptoms to those caused by the coronavirus. Since the disease has common symptoms in patients and these symptoms can cause some degree of impairment in the vocal system and generate irregular vibrations, this study aims to identify Covid-19 signatures in audios with sustained vowels through analysis of the energy in the sound waves. By using Wavelet transform and Teager operator, it was possible to identify, with statistical methods, unique characteristics of healthy and infected individuals, which can be used in the creation of classifiers to identify the pathology. It was found that infected people seem to have a vector with an average value of 0.00034 while healthy people produce a value of 0.01665. Also, it was possible to analyze the difference between the minimum and maximum values found in the vectors.

Keywords: Signal processing. Wavelet transform. Pathology identification. Teager Operator

Lista de Figuras

Figura 2.1 - Aparelho fonador	19
Figura 2.2 - Frequência fundamental	20
Figura 2.3 - Codificação da voz	22
Figura 2.4 - Filtro de áudio	23
Figura 3.1 - Wavelet de Haar	32
Figura 4.1 - Informações dos metadados	35
Figura 4.2 - Energia mínima e máxima	37
Figura 4.3 - Espectrograma das amostras	39

Lista de Tabelas

Tabela 2.1 - Trabalhos relacionados com sintomas da covid	26
Tabela 4.1 - Informações estatísticas da amostra	38

Lista de Abreviaturas

CWT	<i>Continuous Wavelet Transformation</i>
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transform</i>
DTWT	<i>Discrete Time Wavelet Transform</i>
EER	<i>Equal Error Rate</i>
FIR	<i>Finite Impulse Response</i>
IEEE	<i>Institute of Electrical and Eletronic Engineers</i>
IISC	<i>Indian Institute of Science</i>
LPCM	<i>Linear Pulse Code Modulation</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
MLP	<i>Multiple Layer Perceptron</i>
NIST	<i>National Institute of Standards and Technology</i>
PCM	<i>Pulse Code Modulation</i>
RNA	Rede Neural Artificial
SVM	<i>Support Vector Machine</i>
WAVE	<i>Waveform</i>
WPT	<i>Wavelet Packet Transform</i>
ZCR	<i>Zero Crossing Rate</i>

Sumário

1	Introdução	15
1.1	Considerações iniciais	15
1.2	Objetivos	17
1.3	Organização do trabalho	17
2	Revisão Bibliográfica	19
2.1	Produção da fala	19
2.2	Codificação da voz	21
2.3	Processamento de sinais digitais	22
2.4	Estado da arte	23
2.4.1	Classificação de sinais de vozes patológicas	23
2.4.2	Classificação de sinais em casos de Covid-19	25
2.5	Dados	27
3	Detalhamento do Trabalho Proposto	29
3.1	Considerações iniciais	29
3.2	Metodologia	30
3.2.1	Pré-processamento	30
3.2.2	Processamento usando Wavelets	31
3.2.3	Energia do Operador de Teager	33
3.3	Resultados esperados	34
4	Testes e Resultados	35
4.1	Apresentação descritiva dos metadados	35
4.2	Energia presente nos áudios	36

5 Conclusões e Trabalhos Futuros	41
Referências	42

Capítulo 1

Introdução

1.1 Considerações iniciais

Os sinais produzidos pelo corpo humano têm sido objeto de pesquisa na busca de indicadores relacionados às mais diversas doenças. Sinais produzidos a partir da tosse, vibrações vocais e respiração já são utilizados no pré diagnóstico de patologias. Com o auxílio da inteligência artificial, esses sinais servem como fonte de dados para alimentar modelos matemáticos que são capazes de identificar vozes doentes com uma acurácia superior a 90 % (MUHAMMAD et al., 2017)(AL-NASHERI et al., 2017).

Para produzir som, o corpo humano dispõe de um conjunto de órgãos chamado aparelho fonador, que é composto por pulmões, brônquios, traqueia, laringe, faringe, pregas vocais e boca. Com o ar que sai dos pulmões, as membranas que compõem as pregas vocais abrem e fecham a região glótica e vibram em resposta à passagem de ar pela glote (SINGH, 2019). A mobilidade, funcionalidade e forma de vibração das pregas vocais podem ser afetadas devido a alguma anormalidade no trato vocal, o que resulta em vibrações irregulares e aumento do ruído acústico e que pode sinalizar a presença de alguma doença.

A análise computacional de sinais de áudio tem como objetivo principal a detecção, de forma automática e não invasiva, de anomalias ligada à voz. Atualmente o processamento de sinais desempenha um papel importante não apenas no diagnóstico precoce de doenças como também no acompanhamento da evolução clínica de pacientes que estão em tratamento.

A dinâmica de mudança de comportamento da sociedade está diretamente relacionada à evolução tecnológica dos últimos anos. O crescimento de dispositivos inteligentes e o aumento constante do poder computacional fazem com que dispositivos, como smartphones, atuem auxiliando o controle da saúde de seus usuários. Por meio dos wearables, nome dado às tecnologias vestíveis, pessoas diabéticas ou com doenças que necessitam de um controle mais rigoroso das condições físicas podem contar com as possibilidades do monitoramento em tempo real. Esses avanços também tornaram a captação e o processamento de sinais cada vez mais acessível. Nos anos 2000, já surgiram estudos (JO et al., 2001) que utilizam sons de linhas telefônicas de hospitais, para auxiliar na identificação de doenças ligadas a voz. Atualmente há estudos que extraem dados de plataformas como youtube para ajudar na identificação de tuberculose (MIRANDA; DIACON; NIESLER, 2019).

Com a disseminação do coronavírus e a eclosão da pandemia em todo o mundo, houve a necessidade cada vez maior de rápido diagnóstico da doença em busca de conter o vírus. Para isso, surgiram os testes de reação em cadeia da polimerase de transcrição reversa (RT-PCR) que são o padrão ouro para detecção de COVID-19, devido à sua alta sensibilidade e especificidade (PASCARELLA et al., 2020). No entanto, esses testes são compostos por reagentes caros, o que pode atrapalhar na abrangência de pessoas testadas. Outro fator importante é a necessidade de locomoção até o local de testagem, o que acaba expondo o indivíduo a uma potencial infecção.

Com a proposta de não apenas auxiliar no controle da pandemia como também garantir um acesso mais igualitário, abrangente e eficiente aos teste de covid, em meados de 2020 surgiram os primeiros estudos (BROWN et al., 2020) (SCHULLER et al., 2020) buscando encontrar características que pudessem ajudar no diagnóstico ágil de pessoas infectadas por meio da análise acústica de áudio. Dado que a maioria dos sintomáticos enfrentam algum impacto na função respiratória (ISMAIL; DESHMUKH; SINGH, 2021) surge a possibilidade de analisar perturbações na vibração da voz. A análise de vogais sustentadas e de tosse se mostrou bastante eficiente na identificação de pessoas com covid-19, com áudios coletados em ambientes controlados (FAKHRY et al., 2021) (ISMAIL; DESHMUKH; SINGH, 2021).

Para auxiliar no controle da pandemia e também fomentar a pesquisa no processamento de sinais de pessoas infectadas com covid surgem plataformas colaborativas, onde qualquer

pessoa pode contribuir enviando áudios, e providenciar informação sobre seu atual estado em relação a covid (BROWN et al., 2020). Com a disponibilização desses dados surge a oportunidade de estender os estudos para além dos hospitais e clínicas, com o objetivo de encontrar características que possam auxiliar na criação de modelos com maior acurácia, possivelmente ajudando a identificar pessoas contaminadas.

1.2 Objetivos

Dado o estado da arte atual, é notório a importância da criação de modelos com máxima acurácia para o pré-diagnóstico de covid-19 por meio da análise de sinais acústicos. Para tal, é necessário a identificação de características discrepantes entre áudios de sujeitos infectados e não infectados.

A fim de ampliar o estudo na área, esse trabalho tem por objetivo a análise variações da energia presente na frequência, e medir se essas variações são discrepantes o suficiente entre indivíduos saudáveis e com a patologia. Os dados utilizados são disponibilizados pelo Instituto de Ciência Indiano (IISC), com a participação de voluntários que enviaram seus dados por meio da internet.

Com o atual cenário da pandemia, vê-se a grande oportunidade da área de processamento de sinais contribuir ainda mais na identificação de patologias, e também auxiliar no controle da pandemia.

1.3 Organização do trabalho

O texto vindouro do presente trabalho está organizado da seguinte forma:

- No Capítulo 2 apresenta-se uma série de trabalhos publicados envolvendo a área de identificação e classificação de vozes patológicas, mostrando como são inúmeras as pos-

sibilidades de se realizar essa tarefa. Expõe-se, também, os principais conceitos e teorias que estão relacionados com o trabalho que foi desenvolvido.

- No Capítulo 3 apresenta-se, com detalhes, todo o desenvolvimento do trabalho proposto e de que forma os conceitos discutidos no capítulo anterior foram utilizados.
- No Capítulo 4 relatam-se todos os resultados obtidos no trabalho, a partir dos testes de classificação de vozes patológicas que foram realizados.
- No Capítulo 5 apresentam-se as conclusões sobre o trabalho, bem como propostas para pesquisas futuras.

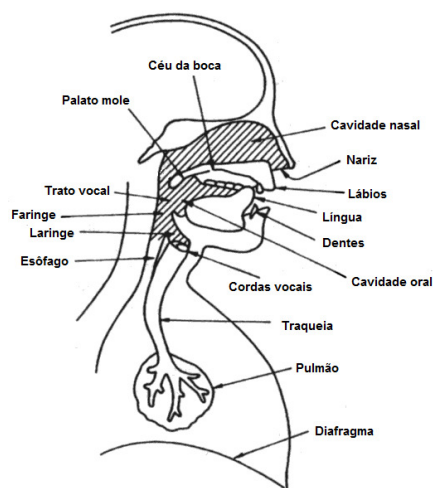
Capítulo 2

Revisão Bibliográfica

2.1 Produção da fala

O ser humano é capaz de produzir som e se comunicar graças a um conjunto de órgãos que formam o aparelho fonador, mostrado na figura 2.1. A produção da fala é feita por meio da conversão do ar que sai dos pulmões em vibrações sonoras. No interior da laringe existe um tecido que forma duas pregas musculares conhecidas como cordas vocais, que vibram com a passagem do ar vindo dos pulmões. Esse som é perturbado por articuladores no trato vocal que nos permitem produzir som distintos, ou seja, com frequências diferentes (PASLEY; FLINKER; KNIGHT, 2015).

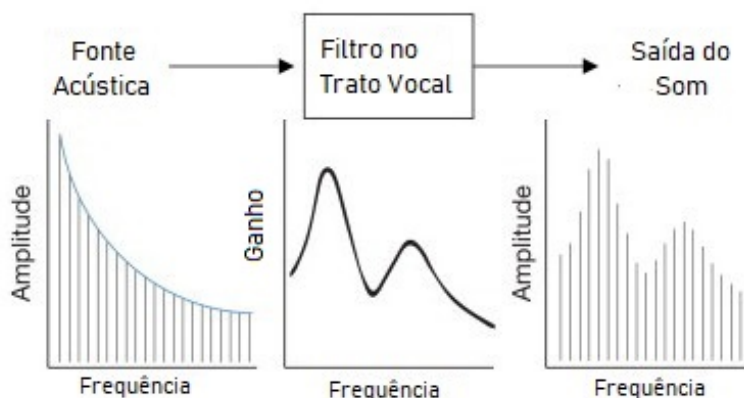
Figura 2.1 – Aparelho fonador



Fonte: Adaptado de (FURUI, 2000)

Ao vibrar, as cordas vocais produzem ondas sonoras em determinada frequência. Essa frequência, chamada de frequência fundamental, é definida como a média das oscilações por segundo das cordas vocais, e é expressa em hertz (YEGNANARAYANA; MURTY, 2009). Diferentes frequências são produzidas por meio da variação da forma do trato vocal, ou seja, a onda sonora é transformada em fala pelos movimentos de várias estruturas, como língua, boca e lábios, como mostrado na figura 2.2. Essa variação na onda sonora ao longo do tempo pode ser visualizada por meio de uma representação visual chamada espectrograma, no qual o eixo vertical representa a frequência da onda sonora e o eixo horizontal representa o tempo.

Figura 2.2 – Frequência fundamental



Fonte: Adaptado de (PASLEY; FLINKER; KNIGHT, 2015)

Em indivíduos, a frequência da vibração do ar supra-glótico está relacionada com fatores como comprimento e espessura das cordas vocais, pressão da corrente de ar, aspectos culturais, etc. Quanto mais longas e espessas as cordas vocais, mais lenta as vibrações, ocasiona vozes mais graves. (DELLER; PROAKIS; HANSEN, 1993)

A variação da frequência fundamental contribuiu para muitos estudos na área de análise de sinais. No artigo de Yegnanarayana e Murty (2009) ela é usada no reconhecimento dos falantes, e em Vergyri et al. (2003) é usada no reconhecimento da fala. Os autores Silva, Oliveira e Andrea (2009) usam essa estratégia na identificação de vozes patológicas.

Como a oscilação se origina de uma estrutura orgânica, ela não é exatamente periódica, mas contém flutuações significativas. A variabilidade ou perturbação da frequência fundamental é chamada de jitter, e a variação na amplitude do período é conhecida como shimmer.

Segundo o autor Rabiner e Schafer (1978) os sons produzidos na fala podem ser classifica-

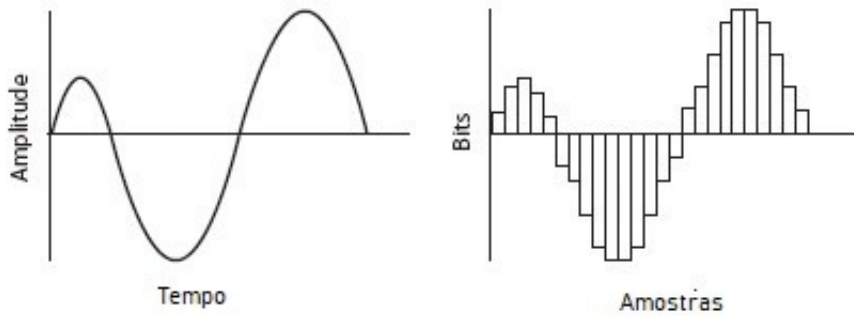
dos em 3 categorias. Ele define como “*Voiced sounds*” os sons que são produzidos forçando o ar por meio da glote com a tensão das cordas vocais ajustada para que vibrem em uma oscilação de relaxamento e como *Unvoiced sounds* os sons que são gerados formando uma constrição em algum ponto do trato vocal e forçando o ar por meio da constrição a uma velocidade alta o suficiente para produzir turbulência. Isso cria uma fonte de ruído de amplo espectro para excitar o trato vocal.

Geralmente, vogais possuem *voiced sounds*, ou seja, os fonemas sonoros são considerados pseudo-periódicos. Em contrapartida, consoantes podem não ter uma estrutura harmônica tão clara, e podem ser confundidos com ruído. Dada essa característica das vogais, elas são amplamente utilizadas no processamento de sinais para análise de periodicidade e de perturbação na frequência fundamental.

2.2 Codificação da voz

A voz pode ser representada como uma sequência de valores amostrais retirados periodicamente. Uma amostra de áudio é um número que representa o valor da onda acústica medida em um ponto específico no tempo, e está representado na figura 2.3. A taxa de amostragem refere-se ao número de amostras de áudio captadas a cada segundo e é medido em hertz. Um áudio, por exemplo, que foi gravado em 16 kHz tem 16000 amostras por segundo. Essa amostragem é armazenada ou representada por uma quantidade de bits, e a quantidade de bits representa a resolução das amostras de áudios. Como a resolução define a quantidade de bits, ela determina a precisão de cada amostra, pois com mais bits é possível ter maior precisão na representação daquele valor.

O áudio pode ser armazenado em arquivos de tipo MP3, WMA, WAV, entre outros. O formato WAV não aplica nenhuma compressão ao fluxo de bits e armazena as gravações de áudio com diferentes taxas de amostragem e taxas de bits. Foi e é um dos formatos padrão para CDs de áudio. O formato de áudio WAV mais comum é o áudio não compactado no formato de modulação por código de pulso linear (LPCM) que armazena áudio amostrado a 44.100 Hz

Figura 2.3 – Codificação da voz

Fonte: Adaptado de (TAN; JIANG, 2018)

com 16 bits por amostra.

2.3 Processamento de sinais digitais

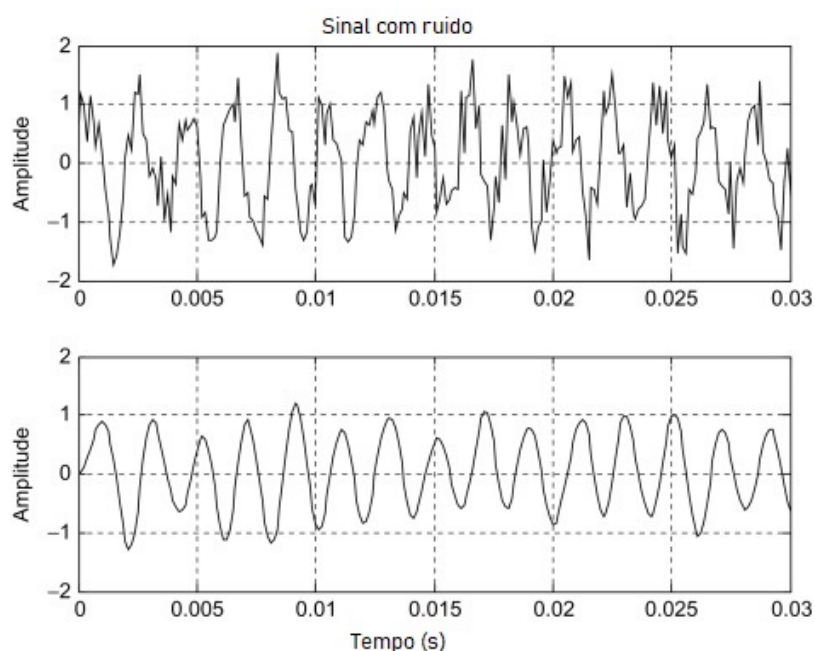
O principal propósito da fala é a comunicação. Pode-se caracterizar a fala como uma onda acústica que carrega uma mensagem. No entanto, essas ondas sonoras não carregam apenas informações relevantes para comunicação. Os sinais que constituem a voz humana carregam informações semânticas que podem ser usadas na identificação de locutor, humor e até na identificação de patologias.

A representação digital da voz é dada por uma sequência de bits. Isso permite que essas ondas sonoras possam ser processadas por computadores. O processamento de sinais tem como objetivo transformar a onda sonora em uma forma mais apropriada para aplicações específicas (TAN; JIANG, 2018). Para isso, são usadas diversas abordagens como transformação no domínio do sinal, mudanças na taxa de amostragem, aplicação de filtros, etc.

Os filtros de frequência são amplamente utilizados no pré-processamento do sinal. Pode-se usar um filtro para remover frequências não desejadas em uma onda, como mostrado na figura 2.4. Exemplos de filtro de frequência são os filtros passa-baixa e o passa-alta. O filtro passa-baixa oferece baixa reatância aos sinais com frequências mais baixas do que a frequência de corte, para que as frequências baixas possam passar, mas fornece alta reatância ao sinal de alta frequência e, assim, bloqueia-o. O filtro passa-alta atenua o sinal de baixa frequência e permite

que apenas o sinal de alta frequência passe por ele (RABINER; SCHAFER, 1978).

Figura 2.4 – Filtro de áudio



Fonte: Adaptado de (TAN; JIANG, 2018)

2.4 Estado da arte

Nessa sessão será apresentada o estado da arte atual relacionado ao processamento de sinais na identificação de patologias, como também os estudos existentes que buscam ajudar na identificação precoce da covid-19 por meio da análise de áudios.

2.4.1 Classificação de sinais de vozes patológicas

Surge na década de 60, os primeiros trabalhos que visavam procurar características no sinal de fala por meio de espectrogramas que permitissem identificar vozes patológicas e vozes não patológicas (LIEBERMAN, 1963), sendo ainda incapaz de identificar qual era a patologia rela-

cionada. Por meio da análise de espectrogramas, o autor House (1976) foram capazes de avaliar não apenas a situação clínica de reabilitação vocal dos pacientes, como também os resultados pós operatório de várias doenças nas pregas vocais.

Em 1980, os autores Deller e Anderson (1980) apresentaram uma técnica de identificação de disfunções na laringe por meio da aplicação de um filtro digital inverso, onde a informação necessária para classificação está contida nos pólos e zeros polinomiais do filtro inverso. Os autores dividem a análise em dois procedimentos. O primeiro está relacionado com a capacidade de identificar a doença e o segundo com a avaliação da evolução do tratamento. Para a classificação, os autores quantificam os níveis de jitter do período fundamental e shimmer para vozes saudáveis e não saudáveis, e também fazem a análise da quantidade de ruído presente no espectro do sinal.

Para a criação de modelos mais assertivos na identificação de vozes patológicas, nos anos 90 os autores Pinto e Titze (1990) tentaram definir medidas padrões de perturbações de jitter e shimmer para otimizar a estimação de parâmetros de modo a conseguir resultados de classificação mais fidedignos.

A distinção entre a alteração das pregas vocais que ocorrem de forma natural e as modificações associadas a patologias é de suma importância na distinção de vozes saudáveis de vozes patológicas. Visando ajudar nessa distinção, os autores Yumoto, Gould e Baer (1982) propuseram o HNR (harmonics-to-noise ratio) que estipula o nível de ruído no sinal e quantifica a parte harmônica e a parte do ruído, fornecendo assim uma indicação da frequência geral do sinal de fala. Os estudos nessa área avançam bastante, e no trabalho de Ferrer et al. (2009), foi analisado a influência do shimmer no cálculo do HNR e propuseram um modelo para remover essa influência.

Como mostrado pelos autores Niedzielska, Glijer e Niedzielski (2001) que realizaram um teste feito em 46 pacientes com idade entre 4 e 14 anos mostrou que havia grande diferença nos valores de jitter, shimmer, frequência fundamental e HNR quando comparado crianças com nódulos vocais com crianças saudáveis. Também foi observado que os valores em crianças com nódulos, durante o tratamento, se aproximavam de valores normais, demonstrando então que a análise dos sinais vocais serviu não apenas para identificar a patologia, mas também para acompanhar a evolução dos pacientes durante o tratamento.

Os autores Jo et al. (2001) visando aumentar a acessibilidade e a velocidade de diagnósticos de problemas ligados a voz, os autores usam a rede telefônica pública para extraírem medidas como jitter, shimmer, HNR e usam uma rede neural para conseguir identificar vozes patológicas. Para o trabalho é utilizado áudios de vozes patológicas gravados em ambiente controlado e armazenada em DAT (digital audio tape). Esses dados são utilizados para treinar e avaliar a rede neural. Após isso, foram usados os áudios gravados na rede telefônica como entrada para a rede neural, e se observou uma taxa de acerto muito similar na classificação de áudios clínicos com áudios gravados pelo telefone. A taxa de acerto ficou em 88.5%.

Com os trabalhos apresentados acima, fica notório a relevância do processamento de sinais para identificação de diversas patologias. Nota-se também que a extração e a análise da frequência fundamental, assim como suas perturbações (shimmer e jitter), ajudaram na identificação e no acompanhamento clínico de diversas doenças.

2.4.2 Classificação de sinais em casos de Covid-19

Com a explosão da pandemia em 2020, os primeiros trabalhos relacionados a Covid se basearam em estudos recentes de outras doenças que dispõem de sintomas similares aos causados pelo coronavírus. Na tabela 2.1 é mostrado alguns exemplos desses trabalhos.

Os primeiros artigos que analisam especificamente a covid-19 surgem na metade do ano de 2020, buscando detectar covid por meio de tosse, que é um dos sintomas mais comum do covid. No trabalho Bagad et al. (2020) identificou evidências estatísticas que era possível a identificação de áudios com tosse de pessoas patológicas.

No trabalho de Mouawad, Dubnov e Dubnov (2021) os autores usaram os dados coletados pelo projeto Corona Voice Detect em parceria com a Voca.ai e a universidade Carnegie Mellon. Os dados coletados são usados como entrada para um modelo de inteligência artificial. Primeiro cria-se um vetor de características usando Mel-Frequency Cepstral Coefficients (MFCC) e com essas informações os autores fazem a análise de recorrência, que é construída a partir de séries temporais e refletem a dinâmica do sistema que produziu essas séries. A partir disso,

Tabela 2.1 – Trabalhos relacionados com sintomas da covid

Aplicação	Tecnologia	Fonte de Dados	Validação
Detecção de tosse de tuberculose	STFT, MFCC, MFB com CNN	O conjunto de áudio do Google extraído de 1,8 milhões de vídeos do Youtube e banco de dados de áudio Freesound	.946 AUC
Medição automática da taxa de respiração da fala	Cepstrograma e SVM com função de base radial	Gravação de fala de 16 participantes da faixa etária de 21 anos	89 % F1 score and RMSE de 4.5 respiração/min para taxa de respiração da fala
Sensor de tosse preciso e com preservação da privacidade usando um microfone de baixo custo	PCA em espectrogramas de áudio, coeficientes FFT e Random Forest Classifier	Tosse atuada de 17 pacientes com tosse devido ao resfriado comum (n = 8), asma (n = 3), alergias (n = 1) e tosse crônica (n = 5)	Taxa de verdadeiro positivo de 92% e taxa de falso positivo de 0,5%

Fonte: Adaptado de (DESHPANDE; SCHULLER, 2020)

é feito uma redução dessa série temporal multivariada até uma sequência simbólica, mantendo as sub-sequências recorrentes mais informativas. O modelo atingiu uma precisão de 83% na identificação das tosses de pessoas que estavam infectadas.

Os autores Brown et al. (2020) fazem uma análise dos dados da Crossware-Data, uma plataforma gerenciada pelo Instituto de Ciência Indiano (IISC) que contém sons de vogais e de tosse de pessoas de todo o mundo que decidiram contribuir enviando seus áudios pela internet. Para a criação do modelo, os autores utilizaram sons de tosse e respiração. Para criação do vetor de característica são extraídas informações como duração, período, RMS Energy, Spectral Centroid, Roll-off, Zero-crossing, MFCC, entre outros. O melhor resultado atingiu AUC de 82% e foi alcançado usando apenas o som de tosse. Quando usado o som de respiração os resultados ficaram em 80%.

Os autores do trabalho Fakhry et al. (2021) propõem o uso de uma rede neural multi-branch com deep learning que recebe entradas heterogêneas como áudios com tosse, informações clínicas como histórico de doenças respiratórias, sintomas de covid, etc. Para extração de características dos dados, os autores usaram MFCCs e também espectrogramas mel-frequency. Os autores escolheram dividir os áudios entre sexo e idade e todos os áudios foram coletados em clínicas. Com o algoritmo, os autores conseguiram como resultado 0,99 AUC para detectar

COVID-19 com precisão de 98,4%

A fim de ajudar a criar modelos cada vez mais assertivos, os pesquisadores começam a estender sua análise para além dos áudios de tosses e começam a abranger também a pesquisa da frequência fundamental. No trabalho Ismail, Deshmukh e Singh (2021) os autores levantam que a maioria dos sintomas ligados a covid estão relacionados com órgãos presentes no trato vocal, e que dada essa característica, seria possível identificar a covid por meio das oscilações das cordas vocais. O método utilizado na pesquisa foi o uso do algoritmo ADLES, desenvolvido pelos próprios autores, e que tem por objetivo a identificação de padrões de oscilações em sinais de áudio. O trabalho foi possível com a utilização de áudios gravados em clínicas. Com o estudo, o autor conclui que padrões de oscilação podem indicar a presença de covid, mas ressalta que esse padrão é mais provável de acontecer em pacientes sintomáticos, e que não foi testada a hipótese de que aquelas oscilações vocais fossem causadas pelo covid e não por outra doença respiratória. O autor conclui que o covid causa assimetria no movimento das cordas vocais, o que pode produzir características descritivas que podem ser usadas em modelos que detecte covid.

Dado esses trabalhos, nota-se que as abordagens iniciais relacionadas ao processamento de sinais de pacientes com covid focaram na análise da tosse, por ser um dos sintomas mais marcantes da doença. Verifica-se também que a análise da frequência fundamental vem sendo abordada para a identificação de covid com o uso de áudios gravados em clínicas. No entanto, atualmente há poucos dados disponíveis para pesquisa. O autor de (XIA et al., 2021) cita problemas em conjunto de dados desbalanceados, o que pode levar modelos de inteligência artificial a ter um viés, possivelmente gerando falsos positivos.

2.5 Dados

Esse trabalho utiliza dados disponibilizados pela Project Coswara. O Projeto Coswara é liderado pelo Indian Institute of Science (IISC) e faz parte de um projeto que busca construir uma ferramenta de diagnóstico para a Covid-19 com base em registro de sons respiratórios,

sons de tosse, fonação sustentada de sons vocálicos e exercício de contagem. O projeto conta com áudios de uma grande variedade de idades, gêneros, condições respiratórias pré-existentes, e localizações geográficas, parâmetros estes que podem ser muito úteis para o processamento dos dados, assim como para o treinamento de modelos de machine learning.

Capítulo 3

Detalhamento do Trabalho Proposto

3.1 Considerações iniciais

Para o desenvolvimento do trabalho, será feita primeiramente uma análise dos áudios, a fim de selecionar aqueles que contenham baixo ruído e que mantenham a vogal `\a\` sustentada por mais de três segundos. Após a seleção, dá-se início à análise estatística quantitativa nos metadados. Como cada áudio possui um metadado correspondente, é possível extrair informações como proporção de homens e mulheres, quantidades de doentes e saudáveis, média de idade, etc.

A partir dos áudios selecionados, serão aplicadas diversas técnicas de pré-processamento para limpeza e normalização. Com os dados pré-processados, é realizado o cálculo da transformada Direta de Wavelet usando os filtros de Haar. Após isso, é extraída a energia de Teager e, com o vetor resultante, será feito o cálculo de métricas estatísticas quantitativas relacionadas à perturbação dessa energia nos áudios.

Como resultado, espera-se ser possível determinar se o processamento aplicado nos áudios possibilitou a identificação de presença de covid-19 por meio das variações na energia do operador de Teager.

3.2 Metodologia

Nesta sessão será apresentado o detalhamento da metodologia usada para o pré-processamento e para a extração da energia do operador de Teager, como também a exibição da base teórica relacionada às abordagens escolhidas para o desenvolvimento deste trabalho.

3.2.1 Pré-processamento

A base de dados utilizada para a pesquisa foi disponibilizada pelo Indian Institute of Science (IISC) e faz parte de um projeto que busca construir uma ferramenta de diagnóstico para a Covid-19 com base em registro de sons respiratórios, sons de tosse, fonação sustentada de sons vocálicos e exercício de contagem. Como os dados foram disponibilizados por voluntários ao redor do mundo que gravaram e enviaram seus dados, é necessário aplicar técnicas de pré-processamento para normalização dos dados, buscando remover ruídos e interferências que possam atrapalhar a análise posterior das ondas sonoras.

Como primeira etapa, foi realizada a normalização dos dados, a fim de transformar todos os valores na mesma ordem de grandeza utilizando a função 3.1. Após isso, teremos todos os valores entre 1 e -1, a fim de evitar viés e desbalanceamento na amostra dos dados.

$$f(x) = (x - \text{media})/x.\text{max}() \quad (3.1)$$

Uma ferramenta de pré-processamento comumente usada para a remoção de irradiação labial é o filtro de pré-ênfase, que enfatiza as frequências mais altas e nivela espectralmente o sinal da fala. Normalmente, a pré-ênfase é aplicada como um filtro de resposta ao impulso finita (FIR) no domínio do tempo. Os filtros são condicionadores de sinal e têm por objetivo separar os componentes de interesse de um sinal de seu sinal não filtrado. Existem vários tipos de filtros, como filtro passa-baixa (LPF) que aceita que apenas sinais de baixa frequência atraves-

sem, eliminando as altas frequências, sendo então conveniente para controlar a faixa mais alta de frequências em um sinal de áudio. Um filtro passa-alta (HPF) é totalmente oposto ao LPF, uma vez que ele rejeita apenas componentes de frequência abaixo de um determinado limite. O filtro passa-alta usado neste trabalho é descrito por 3.2

$$H(z) = 1 - 95z^{-1} \quad (3.2)$$

3.2.2 Processamento usando Wavelets

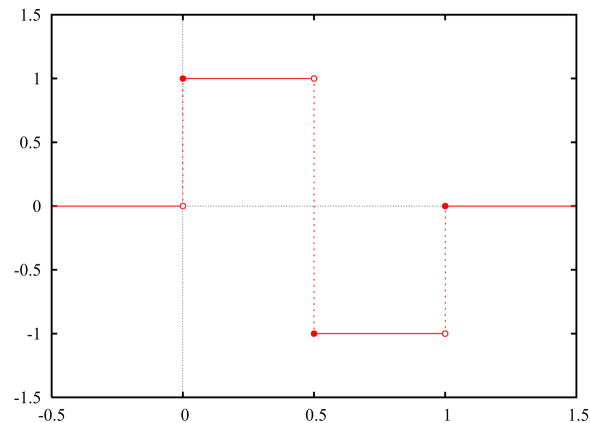
Wavelets são funções matemáticas que dividem os dados em diferentes componentes de frequência e, em seguida, analisa cada componente com uma resolução correspondente à sua escala (GRAPS, 1995). As wavelets vem sendo utilizadas em inúmeras áreas como compressão de imagem, radar, previsão de terremoto, processamento de sinais entre outros.

A transformada de Wavelets é uma alternativa à transformada de Fourier pois utiliza funções base de suporte diferindo de fourier que utiliza função base senoidais para a transformação do sinal (DONALD et al., 2009). A vantagem de Wavelets é que na transformada de Fourier há uma mudança do domínio do tempo para o domínio da frequência, tendo então como saída da transformada um espectro de frequência do sinal original, perdendo o tempo que aquela determinada frequência aconteceu. Já nas wavelets é possível ter a resolução no domínio da frequência e do tempo, devido a janela flutuante que percorre sobre o sinal.

A fórmula geral de wavelets, considerando um sinal contínuo, é mostrada em 3.3, onde o $x(t)$ é o sinal de entrada, ψ uma função wavelet mãe arbitrária e a e b definidos como escala e translação dessa função wavelet mãe.

$$X_{a,b} = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt \quad (3.3)$$

A função wavelet mãe é escolhida de acordo com os interesses da pesquisa, e também levando em conta as características de $x(t)$ (GUIDO et al., 2020). A onda quadrada de Haar, mostrada na figura 3.3 é uma das mais famosas funções walevets mãe.

Figura 3.1 – Wavelet de Haar

Fonte:(DONALD et al., 2009)

A análise de wavelet em um sinal no domínio do tempo é realizada com uma versão contraída e de maior frequência da wavelet mãe, enquanto a análise de frequência é realizada por meio de uma tradução da mesma wavelet. O resultado dessa decomposição é codificar os dados pelos coeficientes de escala da wavelet.

O termo wavelets é genérico, pois há diferentes tipos de wavelets, como a transformada contínua de wavelet (CWT), transformada discreta de wavelet (DWT), transformada de tempo discreto de wavelet (DTWT) e também a transformada estacionária de tempo discreto de wavelet (SDTWT). No desenvolvimento deste trabalho, foi utilizada a transformada discreta de wavelet Packet juntamente com o filtro Haar como função wavelet mãe.

O procedimento prático para aplicação da wavelet começa com a decomposição de $x(t)$ em uma árvore binária com bandas de tamanho $\frac{n}{2^j}$ onde j é o nível da árvore e n é o tamanho do vetor. Após a decomposição é aplicado sobre cada banda o filtro passa-alta $h(x)$ e o passa-baixa $g(x)$, ou seja, é feita a convolução do sinal com os filtros, como mostrado abaixo. O filtro $g(x)$ é definido como o inverso de $h(x)$.

```

for ( i = 0; i < n; i += 2)
{
    t [ j ] = 0;
    for ( k = 0; k < size_h; k++)
        t [ j ] += x [ ( i + k) % n ] * h [ k ];
    j++;
}

```

```

}
for (long i=0; i<n; i+=2)
{
    t[j]=0;
    for (k=0; k<size_g; k++)
        t[j]+=x[(i+k)%n]*g[k];
    j++;
}

```

Determinar qual será o valor de j é uma parte importante, pois essa escolha define o nível da árvore de decomposição da DWT, ou seja, quando maior o valor de j , mais refinado está $x(t)$ em relação à frequência, porém com menos informação relacionado ao domínio do tempo (GUIDO, 2017). Para o desenvolvimento do trabalho foi escolhido o nível médio da árvore, ou seja, dado que o tamanho máximo da árvore corresponde a $\log_2(M)$, tem-se como nível médio o valor $j = \lfloor (1 + \log_2(M))/2 \rfloor$ buscando o equilíbrio entre informação temporal e resolução de frequência.

3.2.3 Energia do Operador de Teager

Após o cálculo da DWT, será aplicado o operador de Teager para medir a variação de energia dos sinais de áudios analisados. Proposto por H. M. Teager (TEAGER, 1980), o cálculo de energia do operador de Teager vem sendo usado em diversas áreas como processamento de sinais (KAISER, 1990) (MARAGOS; KAISER; QUATIERI, 1993), processamento de imagens (MITRA et al., 1995) entre outros. O operador suprime sinais de baixa frequência e enriquece sinais transientes de alta frequência (SUBASI; YILMAZ; TUFAN, 2011). Como mostrado em 3.4, o operador utiliza o produto ao quadrado da amplitude pela instantânea frequência para aproximadamente modelar a energia do sinal de tempo discreto.

$$\Psi(x[n]) = x^2[n] - x[n-1]x[n+1] \quad (3.4)$$

3.3 Resultados esperados

Após as etapas de tratamento e processamentos dos dados, inicia-se a análise com parâmetros estatísticos para identificar variações relacionadas a energia do sinal em diferentes cenários. Como cada áudio usado nesse trabalho é acompanhado de um arquivo de metadado com informações sobre o autor do áudio, é possível estabelecer medidas estatísticas de energia baseado no estado clínico do autor, idade, sexo, etc.

Por meio de todo tratamento aplicado aos dados utilizados, espera-se ser possível estabelecer valores padrões que consigam diferir entre pessoas positivas para covid-19 e pessoas saudáveis. Com isso, será possível o avanço das pesquisas de inteligência artificial que buscam criar modelos para um diagnóstico mais cedo e preciso de pessoas infectadas com covid-19, facilitando assim o controle da pandemia. Acredita-se que a energia presente no vetor resultante de cada áudio poderá ser usada como característica de entrada nos mais diversos classificadores.

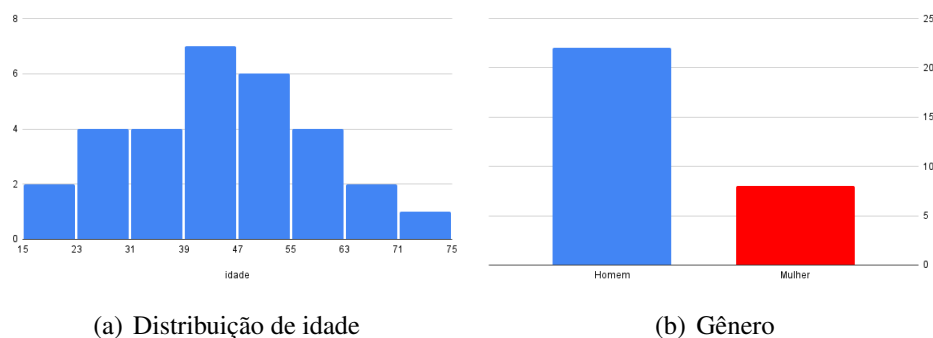
Capítulo 4

Testes e Resultados

4.1 Apresentação descritiva dos metadados

Com a finalização do desenvolvimento prático e a aplicação dos conceitos demonstrados no capítulo 3, inicia-se a análise dos resultados obtidos. Inicialmente, foi escolhida uma amostragem composta por trinta áudios. Estes foram escolhidos levando em consideração a qualidade, com baixo nível de ruído e que contêm vogais sustentadas por mais de dois segundos. Dos trinta selecionados, quinze áudios são de pessoas saudáveis, ou seja, que não estavam contaminadas pela covid-19, e o restante de pessoas contaminadas pela covid-19. Como cada amostra tem atrelada a ela um arquivo de metadado com informações como idade, sexo, país e presença de outras doenças sobre a pessoa que está enviando os dados, essa sessão será focada na apresentação e inspeção dessas informações.

Figura 4.1 – Informações dos metadados



Fonte: Elaborado pelo autor

Nesse trabalho foi escolhido uma amostra que contava com 23 homens e 7 mulheres, mostrada no gráfico 4.2(b). Dado que as ondas sonoras produzidas pelas mulheres vibram em frequências diferentes da dos homens devido a características do sistema fonador, a abordagem de ter ambos os sexos foi escolhida para garantir que o processamento aplicado resultaria em uma classificação acurada independente do sexo. Também percebe-se pela distribuição da idade, mostrada em 4.2(a), que houve uma pluralidade na escolha das idades, abrangendo idades de 15 a 74 anos, com uma concentração entre 39 e 55 anos, visando novamente ter uma amostragem diversa, que pudesse testar e garantir que o processamento aplicado seria relevante para qualquer idade e sexo. Nenhum dos indivíduos alegaram fumar, ter diabetes ou qualquer comorbidade. Isso foi escolhido buscando trazer o mínimo de impacto ao trabalho, que busca identificar a energia na vogal sustentada \a\ para detectar covid-19, e qualquer doença que pudesse interferir na vibração das ondas sonoras produzidas pelo aparelho fonador poderia atrapalhar o resultado.

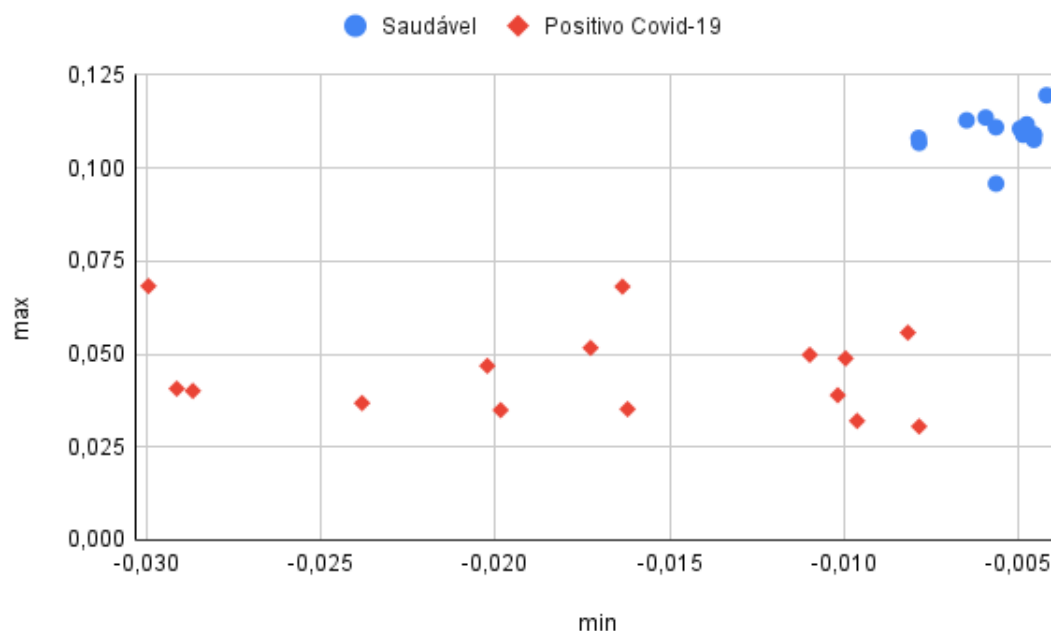
4.2 Energia presente nos áudios

O desenvolvimento do trabalho foi baseado na análise da energia presente nas amostras de áudios selecionadas. Após o pré-processamento, que buscou eliminar frequências e ruídos que poderiam interferir nos resultados, foi aplicado a Transformada Discreta de Wavelets Packet com o objetivo de atingir uma maior pureza nas frequências presentes. Em seguida, para medir a variação de energia, foi utilizado o operador de Teager. Com a finalização do processamento dos sinais, ocorreu a extração de medidas estatísticas que descrevessem os valores dos vetores resultantes.

A figura 4.2 ilustra os valores máximos e mínimos de energia encontrados nos vetores após todo o processamento. Nota-se que pessoas contaminadas com a covid-19 apresentaram valores mínimos e máximos de energia menores do que pessoas saudáveis. É importante ressaltar que em valores mínimos em algumas amostras tem-se um grau de similaridade, mas quando analisa-se juntamente com os valores máximos nota-se a grande discrepância entre pessoas saudáveis e

infectadas.

Figura 4.2 – Energia mínima e máxima



Fonte: Elaborado pelo autor

Dado que o coronavírus, na maioria dos casos sintomáticos, causa um certo grau de comprometimento das funções respiratórias, era esperado encontrar sinais da presença da doença por meio da análise de frequência, pois, como citado no capítulo 2, o aparelho fonador é composto por pulmões, brônquios, traqueia, laringe, faringe, pregas vocais e boca, e seu funcionamento é baseado no ar que sai dos pulmões e que faz as membranas que compõem as pregas vocais abrirem e fecharem a região glótica causando então uma vibração em resposta à passagem de ar pela glote. Dado esse sistema complexo, a mobilidade, funcionalidade e forma de vibração das pregas vocais podem ser afetadas devido a alguma anormalidade no trato vocal, que pode ser causado pela covid, e que resulta em vibrações irregulares e aumento do ruído acústico.

Afim de provar se a diferença entre as médias encontradas nos vetores de pessoas saudáveis e de infectados são significativas, foi aplicado o teste de teste de hipótese para diferença entre médias. Levando em consideração as informações apresentadas na tabela 4.1, considera-se as seguintes hipóteses:

$$\begin{aligned}
 H_0 : \mu_0 &= \mu_1 \\
 H_1 : \mu_0 &> \mu_1
 \end{aligned}
 \tag{4.1}$$

onde $H_0 : \mu_0 = \mu_1$ é a hipótese nula, ou seja, que os valores das médias não diferem em nada, e $H_1 : \mu_0 > \mu_1$ como hipótese alternativa, que diz que há diferença entre as médias de valores encontrados nos vetores resultados, foi utilizado o teste t de student para provar estatisticamente a diferença nas médias. Seguindo os conceitos da distribuição normal Z foi estabelecido um nível de significância de 5%.

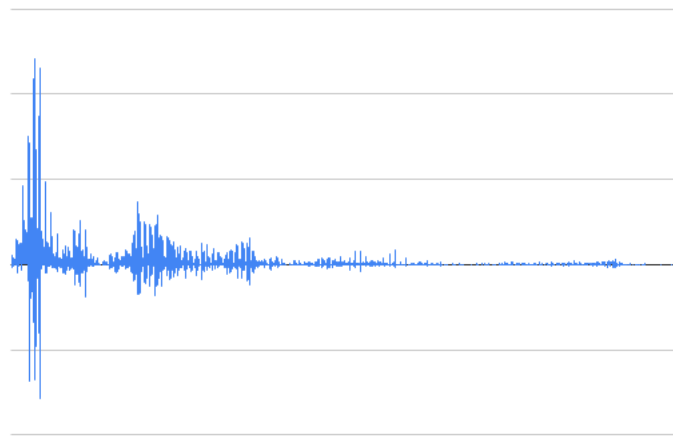
Tabela 4.1 – Informações estatísticas da amostra

	Saudável	Infectado
Média	0,01665	0,00034
Desvio Padrão	0,00210	0,00004
Número de amostras	15	15

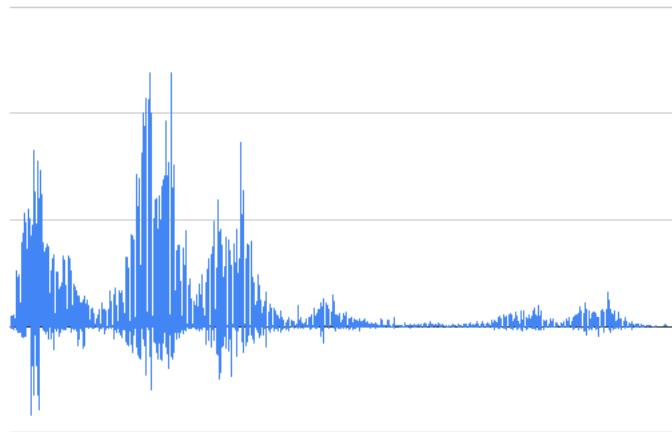
Fonte: Elaborado pelo autor

Como resultado obteve-se um valor $t = 2,0314E-14$. Esse valor é menor que a significância escolhida de 0,05 resultando então na rejeição da hipótese nula H_0 . Com a prova da rejeição da hipótese nula, pode-se afirmar que existe uma diferença significativas entre as médias nos vetores resultantes quando se aplica as técnicas de processamento mostrado nesse trabalho e se compara os valores encontrados no vetor resultado, e com isso é possível afirmar que pessoas saudáveis demonstram ter um valor médio de energia maior que pessoas infectadas com a covid-19, sendo então a média uma característica interessante para ser usada em classificadores de patologia referentes a covid-19.

Outra forma de análise é observar o espectrograma mostrado na figura 4.3. Após o processamento aplicado sobre os áudios da vogal \a\ de forma sustentada, é possível notar a diferença de valores encontrados na amostra de um voluntário saudável em comparação a um infectado com a covid-19. É possível observar que pessoas saudáveis conseguem manter uma faixa de energia maior por mais tempo, enquanto o espectrograma do indivíduo infectado mostra uma energia menor, afirmado também com os valores mínimos e máximos de energia mostrados em 4.2, como também na prova estatística de teste de hipótese de diferença entre medias que foi demonstrado anteriormente.

Figura 4.3 – Espectrograma das amostras

(a) Pessoa doente



(b) Pessoa saudável

Fonte: Elaborado pelo autor

Em suma, foi demonstrado que, com o tratamento aplicado aos dados, foi possível, com métodos estatísticos básicos, identificar características únicas de indivíduos saudáveis e de infectados. Pôde-se concluir que há uma tendência de valores menores de energia para casos positivos da covid-19, por meio da análise conjunta de valores mínimos, máximos, média aritmética e observação do espectrograma resultante. Também foi mostrado que essa diferença é discriminativa o suficiente para ser usada até nos mais simples classificadores. Tem-se então que a energia provida pelo operador de Teager é uma característica importante, que pode ser usada nas mais diversas aplicações de machine learning para rápida identificação da patologia.

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho foi efetuada uma avaliação de sinais acústicos para o pré-diagnóstico de covid-19. Para o desenvolvimento foi-se utilizado a base de dados disponibilizada pelo Indian Institute of Science (IISc) que faz parte de um projeto que busca construir uma ferramenta de diagnóstico para a Covid-19 com base em registro de sons respiratórios, sons de tosse, fonação sustentada de sons vocálicos e exercício de contagem. No presente trabalho, a amostragem escolhida continha áudios da vogal sustentada a que carregam a maior quantidade de energia dentre as vogais. Como os dados foram disponibilizados por voluntários ao redor do mundo que gravaram e enviaram seus dados, foi necessário aplicar técnicas de pré-processamento para normalização dos áudios, buscando remover ruídos e interferências que poderiam atrapalhar a posterior análise das ondas sonoras. Após a limpeza foi aplicada a transformada discreta de Wavelets para um refinamento dos áudios, e posteriormente, calculado a energia com o operador de Teager.

O trabalho foi desenvolvido focado em identificar características únicas em áudios de indivíduos saudáveis e de infectados que pudesse servir como entrada de classificadores para a identificação da presença da patologia por meio da inspeção das ondas sonoras produzidas pelo corpo humano.

Com o resultado do processamento aplicado aos dados foi possível identificar que pessoas que estavam com covid-19 apresentavam um valor menor de energia quando comparado a pessoas saudáveis. Essa análise foi efetuada por meio da observação dos valores mínimos, máximos e a média de energia presentes no vetor resultante após o processamento, como

também, a inspeção das ondas sonoras por meio da análise do espectrograma. Dado que o covid-19, em casos sintomáticos, compromete alguns órgãos do sistema fonador, que é responsável pela fala, era esperado encontrar assinaturas desse comprometimento por meio da verificação da energia presente no áudio após o devido tratamento para limpeza e normalização. Foi mostrado que essa assinatura de energia provida pelo operador de Teager é uma característica importante, que pode ser usada nas mais diversas aplicações de machine learning para rápida identificação da patologia.

Para trabalhos futuros, sugere-se investigar se essa perda de energia também ocorre devido a outras doenças, como gripe, tumor, ou alguma doença presente em órgãos do aparelho fonador, pois o objetivo do trabalho foi encontrar características discriminantes entre pessoas saudáveis e infectadas com o Sars-Cov-2, e não investigou-se se essas assinaturas encontradas na energia acontecem somente em casos de covid-19. Também sugere-se a continuidade dessa pesquisa na criação de modelos de machine learning que utilize como entrada o processamento aplicado aos dados desse trabalho, com uso não somente de vogais, mas também de tosse, respiração etc.

Referências

- AL-NASHERI, A.; MUHAMMAD, G.; ALSULAIMAN, M.; ALI, Z. Investigation of voice pathology detection and classification on different frequency regions using correlation functions. **Journal of Voice**, Elsevier, v. 31, n. 1, p. 3–15, 2017.
- BAGAD, P.; DALMIA, A.; DOSHI, J.; NAGRANI, A.; BHAMARE, P.; MAHALE, A.; RANE, S.; AGARWAL, N.; PANICKER, R. Cough against covid: Evidence of covid-19 signature in cough sounds. **arXiv preprint arXiv:2009.08790**, 2020.
- BROWN, C.; CHAUHAN, J.; GRAMMENOS, A.; HAN, J.; HASTHANASOMBAT, A.; SPATHIS, D.; XIA, T.; CICUTA, P.; MASCOLO, C. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In: **Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. [S.l.: s.n.], 2020. p. 3474–3484.
- DELLER, J. R.; ANDERSON, D. J. Automatic classification of laryngeal dysfunction using the roots of the digital inverse filter. **IEEE Transactions on Biomedical Engineering**, IEEE, n. 12, p. 714–721, 1980.
- DELLER, J. R.; PROAKIS, J. G.; HANSEN, J. H. L. **Discrete-time processing of speech signals**. [S.l.]: Institute of Electrical and Electronics Engineers, 1993.
- DESHPANDE, G.; SCHULLER, B. An overview on audio, signal, speech, & language processing for covid-19. **arXiv preprint arXiv:2005.08579**, 2020.
- DONALD, D.; EVERINGHAM, Y.; MCKINNA, L.; COOMANS, D. 3.23 - feature selection in the wavelet domain: Adaptive wavelets. In: BROWN, S. D.; TAULER, R.; WALCZAK, B. (Ed.). **Comprehensive Chemometrics**. Oxford: Elsevier, 2009. p. 647–679. ISBN 978-0-444-52701-1. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780444527011000338>.
- FAKHRY, A.; JIANG, X.; XIAO, J.; CHAUDHARI, G.; HAN, A.; KHANZADA, A. Virufy: A multi-branch deep learning network for automated detection of covid-19. **arXiv preprint arXiv:2103.01806**, 2021.
- FERRER, C.; GONZÁLEZ, E.; HERNÁNDEZ-DÍAZ, M. E.; TORRES, D.; TORO, A. del. Removing the influence of shimmer in the calculation of harmonics-to-noise ratios using ensemble-averages in voice signals. **EURASIP Journal on Advances in Signal Processing**, Springer, v. 2009, p. 1–7, 2009.
- FURUI, S. **Digital Speech Processing: Synthesis, and Recognition, Second Edition**. Taylor & Francis, 2000. (Signal Processing and Communications). ISBN 9781420002669. Disponível em: <https://books.google.com.br/books?id=X6mZGqZmcbgC>.

- GRAPS, A. An introduction to wavelets. **IEEE computational science and engineering**, IEEE, v. 2, n. 2, p. 50–61, 1995.
- GUIDO, R. C. Effectively interpreting discrete wavelet transformed signals [lecture notes]. **IEEE Signal Processing Magazine**, v. 34, n. 3, p. 89–100, 2017.
- GUIDO, R. C.; PEDROSO, F.; FURLAN, A.; CONTRERAS, R. C.; CAOBIANCO, L. G.; NETO, J. S. Cwt× dwt× dtwt× sdtwt: Clarifying terminologies and roles of different types of wavelet transforms. **International Journal of Wavelets, Multiresolution and Information Processing**, World Scientific, v. 18, n. 06, p. 2030001, 2020.
- HOUSE, W. F. Cochlear implants. **Annals of Otology, Rhinology & Laryngology**, SAGE Publications Sage CA: Los Angeles, CA, v. 85, n. 3_suppl, p. 3–3, 1976.
- ISMAIL, M. A.; DESHMUKH, S.; SINGH, R. Detection of covid-19 through the analysis of vocal fold oscillations. In: IEEE. **ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2021. p. 1035–1039.
- JO, C.; KIM, K.; KIM, D.; WANG, S. Screening of pathological voice from ars using neural networks. In: **Second International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications**. [S.l.: s.n.], 2001.
- KAISER, J. F. On a simple algorithm to calculate the 'energy' of a signal. In: IEEE. **International conference on acoustics, speech, and signal processing**. [S.l.], 1990. p. 381–384.
- LIEBERMAN, P. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 35, n. 3, p. 344–353, 1963.
- MARAGOS, P.; KAISER, J. F.; QUATIERI, T. F. Energy separation in signal modulations with application to speech analysis. **IEEE transactions on signal processing**, IEEE, v. 41, n. 10, p. 3024–3051, 1993.
- MIRANDA, I. D.; DIACON, A. H.; NIESLER, T. R. A comparative study of features for acoustic cough detection using deep architectures. In: IEEE. **2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.], 2019. p. 2601–2605.
- MITRA, S.; THURNHOFER, S.; LIGHTSTONE, M.; STROBEL, N. Two-dimensional teager operators and their image processing applications. In: **Proc. IEEE Workshop Nonlinear Signal Image Processing**. [S.l.: s.n.], 1995. p. 959–962.
- MOUAWAD, P.; DUBNOV, T.; DUBNOV, S. Robust detection of covid-19 in cough sounds. **SN Computer Science**, Springer, v. 2, n. 1, p. 1–13, 2021.
- MUHAMMAD, G.; ALHAMID, M. F.; HOSSAIN, M. S.; ALMOGREN, A. S.; VASILAKOS, A. V. Enhanced living by assessing voice pathology using a co-occurrence matrix. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 17, n. 2, p. 267, 2017.
- NIEDZIELSKA, G.; GLIJER, E.; NIEDZIELSKI, A. Acoustic analysis of voice in children with noduli vocales. **International journal of pediatric otorhinolaryngology**, Elsevier, v. 60, n. 2, p. 119–122, 2001.

PASCARELLA, G.; STRUMIA, A.; PILIEGO, C.; BRUNO, F.; BUONO, R. D.; COSTA, F.; SCARLATA, S.; AGRÒ, F. E. Covid-19 diagnosis and management: a comprehensive review. **Journal of internal medicine**, Wiley Online Library, v. 288, n. 2, p. 192–206, 2020.

PASLEY, B.; FLINKER, A.; KNIGHT, R. Speech sounds. In: TOGA, A. W. (Ed.). **Brain Mapping**. Waltham: Academic Press, 2015. p. 661–666. ISBN 978-0-12-397316-0. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780123970251000622>.

PINTO, N. B.; TITZE, I. R. Unification of perturbation measures in speech signals. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 87, n. 3, p. 1278–1289, 1990.

RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals**. [S.l.]: Prentice-Hall, 1978.

SCHULLER, B. W.; SCHULLER, D. M.; QIAN, K.; LIU, J.; ZHENG, H.; LI, X. Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. **arXiv preprint arXiv:2003.11117**, 2020.

SILVA, D. G.; OLIVEIRA, L. C.; ANDREA, M. Jitter estimation algorithms for detection of pathological voices. **EURASIP Journal on advances in signal processing**, Springer, v. 2009, p. 1–9, 2009.

SINGH, R. **Profiling humans from their voice**. [S.l.]: Springer, 2019.

SUBASI, A.; YILMAZ, A.; TUFAN, K. Detection of generated and measured transient power quality events using teager energy operator. **Energy Conversion and Management**, v. 52, p. 1959–1967, 04 2011.

TAN, L.; JIANG, J. **Digital signal processing: fundamentals and applications**. [S.l.]: Academic Press, 2018.

TEAGER, H. Some observations on oral air flow during phonation. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, IEEE, v. 28, n. 5, p. 599–601, 1980.

VERGYRI, D.; STOLCKE, A.; GADDE, V.; FERRER, L.; SHRIBERG, E. Prosodic knowledge sources for automatic speech recognition. In: **2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)**, Hong Kong, China: IEEE, 2003. v. 1, p. I–208–I–211. ISBN 9780780376632. Disponível em: <http://ieeexplore.ieee.org/document/1198753/>.

XIA, T.; HAN, J.; QENDRO, L.; DANG, T.; MASCOLO, C. Uncertainty-aware covid-19 detection from imbalanced sound data. **arXiv preprint arXiv:2104.02005**, 2021.

YEGNANARAYANA, B.; MURTY, K. S. R. Event-based instantaneous fundamental frequency estimation from speech signals. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 17, n. 4, p. 614–624, 2009.

YUMOTO, E.; GOULD, W. J.; BAER, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. **The journal of the Acoustical Society of America**, Acoustical Society of America, v. 71, n. 6, p. 1544–1550, 1982.