



**Instituto Federal de Educação, Ciência e Tecnologia da Paraíba**

**Campus João Pessoa**

**Programa de Pós-Graduação em Engenharia Elétrica**

**ELLIOTT VICTOR DE SOUSA CHAVES**

**ATHENA ACOUSTIC ANALYSER - SOLUÇÃO E-HEALTH  
BASEADA EM COMPUTAÇÃO EM NUVEM PARA  
AVALIAÇÃO DA QUALIDADE VOCAL**

**DISSERTAÇÃO DE MESTRADO**

**JOÃO PESSOA**

**2023**

**Elliott Victor de Sousa Chaves**

**Athena Acoustic Analyser - Solução e-Health baseada em  
computação em nuvem para avaliação da qualidade vocal**

Dissertação apresentada como requisito parcial  
para obtenção do título de Mestre em Engenharia  
Elétrica, pelo Programa de Pós-Graduação em  
Engenharia Elétrica do Instituto Federal de Edu-  
cação, Ciência e Tecnologia da Paraíba – IFPB.

Orientadora: Profa. Dra. Silvana Luciene do Nas-  
cimento Cunha Costa  
Coorientador: Prof. Dr. Vinícius Jefferson Dias  
Vieira

João Pessoa

2023

Dados Internacionais de Catalogação na Publicação (CIP)  
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa

C512a Chaves, Elliott Victor de Sousa.

*Athena acoustic analyser* – solução *e-health* baseada em computação em nuvem para avaliação da qualidade vocal / Elliott Victor de Sousa Chaves. – 2023.

97 f. : il.

Dissertação (Mestrado em Engenharia Elétrica) - Instituto Federal de Educação da Paraíba / Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), 2023.

Orientação : Prof<sup>a</sup>. D.ra Silvana Luciene do Nascimento Cunha Costa.

Coorientação : Prof<sup>o</sup> D.r Vinicius Jefferson Dias Vieira.

1. Telessaúde. 2. Processamento de sinais. 3. Análise acústica. 4. Gráficos de recorrência. 5. Computação em nuvens. I. Título.

CDU 61:004.775(043)



MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA

**PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU***

**MESTRADO EM ENGENHARIA ELÉTRICA**

**ELLIOTT VICTOR DE SOUSA CHAVES**

**ATHENA ACOUSTIC ANALYSER - SOLUÇÃO E-HEALTH BASEADA EM COMPUTAÇÃO EM NUVEM PARA  
AVALIAÇÃO DA QUALIDADE VOCAL**

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia Elétrica, pelo Programa de Pós- Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB - Campus João Pessoa.

Aprovado em 24 de fevereiro de 2023.

**Membros da Banca Examinadora:**

**Dra. Silvana Luciene do Nascimento Cunha Costa**

IFPB - PPGEE

**Dr. Vinícius Jefferson Dias Vieira**

IFPB - Membro Externo

**Dra. Suzete Élide Nóbrega Correia**

IFPB - PPGEE

**Dr. Ruan Delgado Gomes**

IFPB - PPGEE

**Dr. Leonardo Wanderley Lopes**

UFPB - Membro Externo

João Pessoa/2023

Documento assinado eletronicamente por:

- **Silvana Luciene do Nascimento Cunha Costa, PRO-REITOR(A) - CD2 - PRPIPG-RE**, em 10/04/2023 10:23:21.
- **Ruan Delgado Gomes, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 10/04/2023 10:52:33.
- **Leonardo Wanderley Lopes, PROFESSOR DE ENSINO SUPERIOR NA ÁREA DE ORIENTAÇÃO EDUCACIONAL**, em 10/04/2023 11:35:28.
- **Suzete Elida Nobrega Correia, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 10/04/2023 16:20:03.
- **Vinicius Jefferson Dias Vieira, PROFESSOR DE ENSINO SUPERIOR NA ÁREA DE ORIENTAÇÃO EDUCACIONAL**, em 10/04/2023 18:25:51.

Este documento foi emitido pelo SUAP em 24/02/2023. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código 393484  
Verificador: f98fb36b81  
Código de Autenticação:



Av. Primeiro de Maio, 720, Jaguaribe, JOÃO PESSOA / PB, CEP 58015-435  
<http://ifpb.edu.br> - (83) 3612-1200

*Este trabalho é dedicado a todos os marginalizados  
pelo sistema educacional brasileiro.*

# AGRADECIMENTOS

Agradeço primeiramente ao meu filho, Gael, instrumento da Força Maior no direcionamento da minha vida.

A Raimundo, Margarete e Yluska, pai, mãe e irmã, respectivamente, que estiveram comigo nos momentos mais sombrios e não desistiram de mim.

Ao meu tio, Euro da Cunha Chaves (PR7-EC) (*in memoriam*), que possibilitou o contato com o radioamadorismo e eletrônica, desenvolvendo o interesse na Engenharia Elétrica.

À profª. Dra. Silvana Cunha Costa e ao prof. Dr. Vinicius Vieira pelas valorosas orientações e paciência com as entregas.

À banca examinadora, composta pelos professores Dr. Leonardo Wanderley, Dr. Ruan Gomes e Dra. Suzete Correia, pelas valorosas contribuições ao trabalho e pela compreensão com o resultado obtido.

Ao Programa de Pós Graduação em Engenharia Elétrica do IFPB, por intermédio de todos os professores, pelas contribuições na formação acadêmica e humana.

Aos meus colegas e amigos pelo suporte e disponibilidade de tempo, nas pessoas de Daniel Pereira e Willames Diniz.

Ao IFPB pela experiência profissional, como professor substituto, e na formação continuada desde o ensino médio, passando pela graduação e agora na pós-graduação.

*"Se queres progredir, conforma-te em parecer insensato e tolo quanto às coisas exteriores. Não pretendas parecer saber coisa alguma."(Epicteto)*



## RESUMO

As Tecnologias da Informação e Comunicação (TICs) desempenham um papel propulsor nos mais variados segmentos, como financeiro, elétrico, agronegócio e educação. Na saúde não é diferente, espaço onde o conceito de telessaúde ou, do inglês, *e-Health* é desenvolvido. Por intermédio da Internet, é possível realizar teleatendimentos, na forma de conferência, entre o profissional de saúde e o paciente, de maneira prática e cômoda, facilitando, inclusive, o acompanhamento de pessoas com restrições de mobilidade. Em concomitância ao conceito de teleatendimento, o uso de ferramentas suplementares que auxiliem o diagnóstico tem se mostrado de relevante interesse pelos profissionais de saúde voltados para a voz, como fonoaudiólogos e otorrinolaringologistas. Uma dessas ferramentas é a análise acústica, que possibilita a caracterização do sistema de produção vocal por intermédio de diversas medidas, sejam estas oriundas da análise linear ou não linear. Dentro do contexto não linear, os gráficos e medidas de quantificação de recorrência viabilizam uma nova forma de caracterização do sistema de produção vocal, possibilitando uma melhor diferenciação entre vozes saudáveis e patológicas. Com o intuito de viabilizar o uso de uma solução de auxílio ao diagnóstico por intermédio da Internet, utilizando os gráficos e medidas de quantificação de recorrência, permitindo o seu escalonamento horizontal e vertical de acesso público, o *Athena Acoustic Analyser* é proposto. Para tanto, um ambiente baseado em Computação em Nuvem é utilizado para possibilitar a sua escalabilidade, permitindo o desenvolvimento de uma aplicação modular voltada para Web. Como resultados finais, foram desenvolvidas uma interface para aquisição de sinais de voz, que permite o envio de arquivos no formato WAV ou sua gravação direta pelo navegador da Internet, e a análise dos sinais, utilizando-se análise visual e quantitativa. Neste trabalho, é realizada a análise acústica de sinais de voz, empregando-se técnicas de processamento digital de sinais. São extraídas medidas dos sinais a partir da análise linear (frequência fundamental, formantes, análise preditiva linear e cepstral, espectro de Fourier e espectrograma) e da análise não linear (gráficos de recorrência e suas medidas de quantificação). O resultado alcançado pode ser verificado por intermédio de um navegador da Web.

**Palavras-chaves:** Telessaúde, processamento de sinais, análise acústica, gráficos de recorrência, Computação em nuvem

## ABSTRACT

Information and Communication Technologies (ICTs) play a driving role in the most varied segments, such as finance, electricity, agribusiness and education. In health it is no different, where the concept of telehealth or, in English, *e-Health* is developed. Through the Internet, it is possible to carry out teleservices, in the form of a conference, between the health professional and the patient in a practical and comfortable way, even facilitating the monitoring of people with mobility restrictions. Concomitantly with the telecare concept, the use of supplementary tools that help with the diagnosis has shown to be of great interest to voice-oriented health professionals, such as speech therapists and otorhinolaryngologists. One of these tools is the acoustic analysis, which allows the characterization of the vocal production system through different measures, whether these are derived from linear or non-linear analysis. Within the non-linear context, graphs and recurrence quantification measures enable a new way of characterizing the vocal production system, enabling a better differentiation between healthy and pathological voices. In order to enable the use of a diagnostic aid solution via the Internet, using graphs and recurrence quantification measures, allowing its horizontal and vertical scaling for public access, the *Athena Acoustic Analyzer* is proposed. Therefore, an environment based on Cloud Computing is used to enable its scalability, allowing the development of a modular application focused on the Web. As final results, an interface for the acquisition of voice signals was developed, which allows the sending of files in WAV format or their direct recording through the Internet browser, and the analysis of the signals using visual and quantitative analysis. In this work, the acoustic analysis of voice signals is carried out, using digital signal processing techniques. Signal measurements are extracted from linear analysis (fundamental frequency, formants, cepstral and linear predictive analysis, Fourier spectrum and spectrogram) and non-linear analysis (recurrence graphs and their quantification measures). The achieved result can be checked via a web browser.

**Key-words:** Telehealth, signal processing, acoustic analysis, recurrence plots, cloud computing

## LISTA DE FIGURAS

Figura 1 – Modelo simplificado de produção da fala . . . . .	29
Figura 2 – Modelo fonte e filtro representativo do trato vocal . . . . .	30
Figura 3 – Processo de amostragem de um sinal analógico . . . . .	32
Figura 4 – Janela de Hamming (a) e sua Transformada de tempo discreto de Fourier (b)	33
Figura 5 – Seção da forma de onda da fala com janelas de análise a curto intervalo de tempo . . . . .	33
Figura 6 – AMDF de um sinal de voz saudável com os mínimos identificados . . . . .	34
Figura 7 – Representação gráfica de um sinal sonoro e seu espectro de Fourier . . . . .	36
Figura 8 – Representação gráfica de um sinal fricativo e seu espectro de Fourier . . . . .	36
Figura 9 – Composição de um sinal senoidal . . . . .	37
Figura 10 – Espectro de Fourier de um sinal . . . . .	38
Figura 11 – Trecho de um sinal real (a) e o espectro de Fourier (b) . . . . .	39
Figura 12 – (a) Sinal $x[n]$ de duração finita; (b) sinal periódico $\tilde{x}[n]$ construído para ser igual a $x[n]$ por um período . . . . .	40
Figura 13 – Interpretação gráfica da Equação 22 . . . . .	41
Figura 14 – Espectrograma apresentando altas e baixas frequências . . . . .	44
Figura 15 – Espectrograma faixa estreita (superior) e faixa larga (inferior) . . . . .	45
Figura 16 – Computando o cepstro complexo utilizando a DTFT . . . . .	46
Figura 17 – A inversa do sistema característico para convolução (cepstro complexo inverso)	47
Figura 18 – CPP de um sinal classificado como normal . . . . .	48
Figura 19 – Modelo preditivo linear . . . . .	49
Figura 20 – Espectro LPC (azul) e espectro da FFT (vermelho). As formantes são identificadas pelos picos $F_1, F_2, F_3$ . . . . .	53
Figura 21 – (a) Interpretação clássica da propagação do som através do sistema vocal. (b) Interpretação da dinâmica não linear de fluidos . . . . .	54
Figura 22 – Apresentação da visualização pela análise não linear de um sinal saudável ou normal: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: $\text{dim}=3, \tau = 7, \varepsilon = 0,139176$ . . . . .	58
Figura 23 – Apresentação da visualização pela análise não linear de um sinal com edema: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: $\text{dim}=3, \tau = 5, \varepsilon = 0,151222$ . . . . .	59
Figura 24 – Apresentação da visualização pela análise não linear de um sinal com paralisia: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: $\text{dim}=4, \tau = 8, \varepsilon = 0,137660$ . . . . .	60

Figura 25 – Apresentação da visualização pela análise não linear de um sinal com nóduo: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: $\dim=3$ , $\tau = 5$ , $\varepsilon = 0,143660$ . . . . .	61
Figura 26 – Fluxo da análise e levantamento de requisitos . . . . .	70
Figura 27 – Pilha tecnológica utilizada . . . . .	71
Figura 28 – Fluxograma da solução proposta . . . . .	72
Figura 29 – Tela inicial . . . . .	73
Figura 30 – Módulo de aquisição . . . . .	74
Figura 31 – Módulo de análise da informação - Seleção de sinal . . . . .	75
Figura 32 – Módulo de análise da informação - Interface . . . . .	75
Figura 33 – Módulo de análise da informação - Seção visual não linear (a) Sinal de voz, (b) retrato de fase e (c) gráfico de recorrência . . . . .	76
Figura 34 – Módulo de análise da informação - Seção quantitativa linear Frequências Fundamental (F0) e Formantes (F1, F2, F3) . . . . .	76
Figura 35 – Módulo de análise da informação - Medidas de quantificação de recorrência. Seletor de medida (a), <i>BoxPlots</i> (b) e Sumarização (c) . . . . .	77
Figura 36 – Saída dos elementos visuais de análise linear para o sinal de teste gravado .	78
Figura 37 – Saída dos elementos visuais de análise não linear para o sinal de teste gravado	79
Figura 38 – Frequências fundamental e formantes do sinal de teste . . . . .	79
Figura 39 – <i>BoxPlots</i> para a medida de recorrência $V_{max}$ . . . . .	80
Figura 40 – <i>BoxPlots</i> para a medida de recorrência $L_{max}$ . . . . .	80

## LISTA DE TABELAS

Tabela 1 – Medidas relacionadas à formação de pontos de recorrência, linhas diagonais e verticais . . . . .	63
Tabela 2 – Tabela com valores para a medida de recorrência $V_{max}$ para cinco classes .	78
Tabela 3 – Tabela com valores para a medida de recorrência $V_{max}$ para três classes . . .	80
Tabela 4 – Tabela com valores para a medida de recorrência $L_{max}$ para cinco classes . .	81
Tabela 5 – Tabela com valores para a medida de recorrência $L_{max}$ para duas classes .	81

## LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
AWS	Amazon Web Services
CSS	Cascading Style Sheets
DFT	Transformada Discreta de Fourier [DTFT] Discrete-Time Fourier Transform
FACIT	Função de autocorrelação a curto intervalo de tempo
GCP	Google Cloud Plataform
IaC	Infraestrutura como código
IU	Interface com o usuário
LPC	Linear Predictive Coding
NoSQL	Not Only Structured Query Language
SSL	Secure Sockets Layer
TIC	Tecnologia da Informação e Comunicação
WAV	WAVEform audio format

## LISTA DE SÍMBOLOS

$x[n]$	Sinal discretizado
$\hat{x}[n]$	Sinal discretizado aproximado
$X_{\hat{n}}(e^{j\hat{w}})$	Transformada de Fourier a curto intervalo de tempo
$D_*\{x[n]\}$	Sistema característico para convolução
$F_s$	Frequência de amostragem
$w_H[m]$	Janela de Hamming
$\phi$	Função de autocorrelação determinística
$F_1, F_2, F_3$	Frequências formantes
$\Omega$	Frequência radiana
$I[n]$	Informação mútua em [n]
$\tau$	Passo de reconstrução
$\Theta$	Função de Heaviside
$\frac{\sigma}{r}$	Proporção da distância máxima e o raio de vizinhança
$dim$	Dimensão de imersão
$P_0$	Probabilidade de encontro de um ponto de recorrência no gráfico de recorrência
$DET$	Determinismo
$\varepsilon$	Raio de vizinhança
$\rho$	Probabilidade de um determinado estado
$L_{max}$	Comprimento máximo das estruturas diagonais
$REC$	Densidade dos pontos/Taxa de recorrência
$TREND$	Tendência
$LAM$	Laminaridade
$v_{min}$	Comprimento mínimo das estruturas verticais

$TT$  Tempo de permanência

$V_{max}$  Comprimento máximo das estruturas verticais



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>18</b>
<b>1.1</b>	<b>Motivação e Definição do Problema</b>	<b>19</b>
<b>1.2</b>	<b>Objetivos</b>	<b>22</b>
1.2.1	Objetivo geral	22
1.2.2	Objetivos específicos	22
<b>1.3</b>	<b>Estrutura do Documento</b>	<b>22</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>24</b>
<b>2.1</b>	<b>Princípios da anatomia, fisiologia e patologia vocal</b>	<b>24</b>
2.1.1	Anatomia e fisiologia da fala	24
2.1.2	Patologias laríngeas	25
2.1.2.1	<i>Edema de Reinke</i>	25
2.1.2.2	<i>Nódulo</i>	25
2.1.2.3	<i>Paralisia</i>	26
2.1.3	Etiologias dos distúrbios vocais	26
<b>2.2</b>	<b>Das medidas tradicionais a longo intervalo de tempo e da análise preditiva linear</b>	<b>27</b>
<b>2.3</b>	<b>Modelo fonte/filtro de produção da fala e análises associadas</b>	<b>28</b>
2.3.1	Seleção e manipulação de amostras em um sinal	30
2.3.1.1	<i>Amostragem</i>	30
2.3.1.2	<i>Janelamento</i>	31
2.3.2	Frequência fundamental	33
2.3.2.1	<i>AMDF - Average Magnitude Difference Function</i>	34
2.3.2.2	<i>Função de Autocorrelação a Curto Intervalo de Tempo (FACIT)</i>	35
2.3.3	Faixa de frequência da voz humana	35
2.3.4	Análise de Fourier	35
2.3.4.1	<i>Transformada Discreta de Fourier (DFT)</i>	38
2.3.4.2	<i>Transformada de tempo discreto de Fourier (DTFT)</i>	39
2.3.4.3	<i>Transforma de Fourier a Curto Intervalo de Tempo (STFT)</i>	42
2.3.4.4	<i>Análise espectrográfica</i>	43
2.3.5	Análise cepstral	44
2.3.5.1	<i>O cepstro a curto intervalo de tempo</i>	47
2.3.5.2	<i>Proeminência de pico cepstral - (CPP)</i>	47
2.3.6	Análise preditiva linear	48
2.3.6.1	<i>Método da covariância</i>	51

2.3.6.2	<i>Método da autocorrelação</i> . . . . .	52
2.3.7	Frequências Formantes . . . . .	52
<b>2.4</b>	<b>A presença da não linearidade na voz</b> . . . . .	<b>53</b>
<b>2.5</b>	<b>Gráficos de recorrência e suas medidas de quantificação</b> . . . . .	<b>55</b>
2.5.1	Espaços de fase . . . . .	55
2.5.1.1	<i>Método da informação mútua para determinação do passo de reconstrução</i> .	56
2.5.1.2	<i>Falsos vizinhos próximos</i> . . . . .	56
2.5.1.3	<i>Gráficos de recorrência</i> . . . . .	57
2.5.1.4	<i>Medidas de quantificação de recorrência</i> . . . . .	58
<b>2.6</b>	<b>Da arquitetura de <i>software</i> e infraestrutura computacional</b> . . . . .	<b>63</b>
2.6.1	Arquitetura de microsserviços . . . . .	63
2.6.2	<i>Computação em nuvem e Infrastructure as Code (IaC)</i> . . . . .	64
2.6.2.1	<i>Docker</i> . . . . .	65
<b>2.7</b>	<b>Trabalhos Relacionados</b> . . . . .	<b>65</b>
<b>3</b>	<b>ATHENA ACOUSTIC ANALYSER</b> . . . . .	<b>69</b>
<b>3.1</b>	<b>Aplicabilidade</b> . . . . .	<b>69</b>
<b>3.2</b>	<b>Metodologia</b> . . . . .	<b>69</b>
3.2.1	Base de dados de teste . . . . .	72
<b>3.3</b>	<b>Resultados obtidos</b> . . . . .	<b>73</b>
3.3.1	Aquisição do sinal de voz . . . . .	73
3.3.2	Análise da Informação . . . . .	74
<b>3.4</b>	<b>Análise comparativa com base de dados de referência</b> . . . . .	<b>77</b>
<b>4</b>	<b>CONCLUSÕES</b> . . . . .	<b>82</b>
<b>4.1</b>	<b>Contribuições da pesquisa</b> . . . . .	<b>82</b>
<b>4.2</b>	<b>Sugestões para trabalhos futuros</b> . . . . .	<b>83</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>85</b>

# 1 INTRODUÇÃO

A palavra comunicação permite uma interpretação ubíqua, permitindo seu uso em diversas áreas, como psicolinguística, crítica literária, na área de marketing e por especialistas em voz (GELDARD, 1965). Sendo um entre tantos outros recursos de comunicação, a fala permite reconhecer um locutor (FECHINE, 2000), o seu humor (VIEIRA, 2018), bem como sua situação de saúde (KAVYA; SHIVRAM, 2018; CHANDRASHEKAR; KARJIGI; SREEDEVI, 2019; ISMAIL; DESHMUKH; SINGH, 2021), podendo ser considerado um reflexo do estado do sistema de produção da fala (COSTA, 2008; COSTA, 2012).

Devido à importância da fala bem como aos distúrbios a ela relacionados, foram desenvolvidas metodologias para avaliação dos transtornos da voz, metodologias as quais podem, ou não, ser invasivas. As técnicas invasivas, como o exame estroboscópico, podem gerar bastante desconforto, pela necessidade de inserção de instrumentos pelas vias áreas ou pela boca (COSTA, 2008). As técnicas não invasivas podem ser de caráter subjetivo ou objetivo. Nos exames subjetivos, como na avaliação percepto-auditiva, é realizada a escuta de um sinal de voz por um profissional treinado e este emite um parecer sobre a qualidade do sistema vocal (COSTA, 2012). A avaliação pode ser comprometida, entretanto, pelo tratamento do sinal de áudio ou do parecer do profissional, traído por suas impressões subjetivas. Já para as métricas objetivas, a análise acústica se utiliza de medidas quantificáveis, sendo bem conhecidas a frequência fundamental, *pitch*, *jitter*, *shimmer* e análise espectral (COSTA, 2008; VIEIRA, 2014). Tais medidas são classificadas como lineares, pois são baseadas no modelo linear de produção da fala, diferindo das medidas não lineares, que estão relacionados à teoria dos sistemas dinâmicos caóticos (FERREIRA, 2010). Em virtude do seu comportamento fisiológico, o sistema de produção da fala é considerado não linear, para cuja análise podem ser utilizados os gráficos de recorrência (GR) e as medidas de quantificação de recorrência (MQR) (VIEIRA *et al.*, 2018).

Além das medidas quantificáveis, que descrevem de maneira objetiva as mais variadas características valendo-se de um sinal de voz, gerando índices para a criação de classes, como Patológicas e Saudáveis, por exemplo, é possível realizar uma análise visual e inferir uma determinada classe de diagnóstico patológico por intermédio de gráficos, a partir de uma característica do sinal - como a periodicidade, por exemplo -, da *Ceprtal Peak Proeminence* (HILLENBRAND; CLEVELAND; ERICKSON, 1994) ou das frequências formantes em um espectrograma (RABINER; SCHAFER, 2007). Além disso, esforços para a classificação de elementos gráficos de forma automatizada - como gráficos de recorrência -, com o objeto do diagnóstico de patologias são desenvolvidos, obtendo-se resultados satisfatórios (BARROS *et al.*, 2020).

Com o avanço da Tecnologia da Informação e Comunicação (TIC), por intermédio dos computadores pessoais e celulares, assim como o aumento da cobertura e capilaridade da Internet, viabilizou-se o uso da TIC para a triagem, diagnóstico e tratamento de doenças, desenvolvendo-se assim o conceito de Telemedicina e Telessaúde (FERNANDES *et al.*, 2015; EREN; WEBSTER, 2015; TEIXEIRA *et al.*, 2021; BITAR; ALISMAIL, 2021).

O uso de soluções baseadas em *e-Health* tem se tornado crescente, alinhado com as necessidades de atendimento em qualquer lugar, sem que o paciente precise deslocar-se do local onde está para ser atendido ou o profissional para o atender. Outro ponto relevante a ser evidenciado no tocante ao uso do sistema Telessaúde, durante a recente pandemia da COVID-19 diz respeito à proteção de pessoas, em especial as que apresentavam comorbidades de um possível contato com o referido vírus, (BITAR; ALISMAIL, 2021). Nesse sentido, o desenvolvimento de soluções *e-Health* voltadas para a detecção de patologias laríngeas, utilizando-se a análise acústica, por intermédio da Internet, se apresentam como promissor.

## 1.1 Motivação e Definição do Problema

Nesta Seção, são apresentados o contexto público epidemiológico recente, do ponto de vista financeiro, legal, clínico, matemático, de comunicações e computacional lastreando-se, assim a necessidade premente do desenvolvimento da solução proposta nesta dissertação.

Utilizando-se do recorte da situação da recente pandemia da COVID-19, várias especialidades e serviços de urgência e emergência tiveram um aumento no atendimento de pacientes por intermédio da Telemedicina (MANN *et al.*, 2020). Para além das vantagens da redução da circulação de indivíduos e sua possível contaminação, assim como diminuição da carga nos serviços de atendimentos presencial, foi possível a redução dos encargos hospitalares, resultantes de uma visita presencial, em torno de US\$ 700,00 (SEVERINI *et al.*, 2022). Outro aspecto importante é a avaliação dos usuários do sistema de Telemedicina. Segundo Kruze *et al.* (2017), que conduziram uma análise de 2.193 artigos publicados em algumas fontes de dados, foi apontado que a satisfação do paciente pode ser associada à telessaúde, além de ser uma opção a ser considerada na expansão da prática médica a áreas remotas sem a necessidade de realocação.

Em alinhamento com o desenvolvimento tecnológico e a expansão das oportunidades na área de telemedicina, impulsionaram-se, também, os serviços de atendimento de profissionais da área de fonoaudiologia (QUEIROZ *et al.*, 2021; DIMER *et al.*, 2020). Tais atividades no Brasil se desenvolvem desde 2003, quando foram iniciados os primeiros estudos voltados para a área, tendo em sua consecução a criação da Resolução CFFa nº 580 de 20 de agosto de 2020. Esta resolução é considerada o marco regulatório para que a telefonaudiologia seja integrada à prática em questão (CFFA, 2020).

Os serviços de telefonaudiologia podem ser classificados conforme a sincronicidade das interações, sendo síncronos, assíncronos, híbridos e automáticos (CFFA, 2020). Nos atendimentos

síncronos, a interação entre paciente e profissional é realizada em tempo real, sendo utilizado áudio e vídeo, desenvolvendo-se o conceito de teleconferência (CONSTANTINESCU *et al.*, 2011; FERRARI; BERNARDEZ-BRAGA; CAMPOS, 2011). Quando a interação é realizada de forma *offline* ou assíncrona, os atores do processo não realizam contato em tempo real, sendo os dados (áudios, vídeos ou texto) coletados, gerados e enviados posteriormente ao contato estabelecido entre os interlocutores (LAPLANTE-LÉVESQUE; PICHORA-FULLER; GAGNÉ, 2006; DOARN *et al.*, 2019). O modelo híbrido combina as formas síncrona e assíncrona (CONSTANTINESCU *et al.*, 2009). Já na forma automática, os dados sobre a saúde do paciente são registrados e transmitidos, como o próprio nome já antecipa, forma automática, ou seja, imediatamente, podendo ser gerados relatórios que possibilitam o monitoramento dos dados à distancia (KEIDSER; CONVERY, 2016).

No contexto das atividades desempenhadas por fonoaudiologia, estas podem ser segmentadas em telemonitoramento, teleinterconsulta, teleconsulta, segunda opinião formativa, teleconsultoria e serviços interpretativos. Nos serviços interpretativos, o fonoaudiólogo, na especialidade relacionada à patologia, emite um laudo baseado nos gráficos, arquivos de imagens e de áudio bem como dados adicionais. Estas informações subsidiam a tomada de decisão, viabilizando a diagnose e o possível tratamento da patologia (CFFA, 2020).

Para auxiliar no diagnóstico da patologia, o profissional pode se utilizar de alguns exames, estes classificados entre invasivos e não invasivos. Os invasivos, como o exame estroboscópico, estão relacionados à inserção de equipamentos pelas vias aéreas, sejam estes sondas, câmeras ou espelhos, podendo gerar bastante desconforto durante o procedimento (ZHANG *et al.*, 2019). Já os não invasivos se baseiam na avaliação do estado do sistema fonatório, seja por métricas objetivas (LOPES *et al.*, 2014) ou por avaliações subjetivas (SERVILHA; PENA, 2010). Uma das formas de se realizar uma avaliação subjetiva é a perceptivo-auditiva. Neste tipo de avaliação, é realizada uma entrevista com o paciente, momento em que o especialista realiza uma descrição subjetiva da qualidade vocal do paciente (STEMPLE; ROY; KLABEN, 2018).

De forma complementar à avaliação subjetiva, métricas objetivas ou instrumentais são utilizadas. Estas são consideradas medidas indiretas do sistema produtor da voz, haja vista os profissionais inferirem o estado fisiológico da laringe e pregas vocais. Os instrumentos mais usuais são as mudanças aerodinâmicas de pressão ou fluxo, imagens visuais da vibração das pregas vocais e a análise acústica (STEMPLE; ROY; KLABEN, 2018; SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022).

Baseada no escopo do processamento digital de sinais, a análise acústica se utiliza de medidas quantificáveis, sendo bem conhecidas a frequência fundamental, *pitch*, *jitter*, *shimmer*, e a análise espectral (SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022). Tais medidas são classificadas como lineares, pois são baseadas no modelo linear de produção da fala, diferindo das medidas não lineares que estão relacionados à teoria dos sistemas dinâmicos caóticos (FERREIRA, 2010).

Com a difusão da TIC nas últimas décadas, se tornou cada vez mais usual a utilização de serviços diversos pelo computador ou dispositivos móveis, como *smartphones*. O papel da tecnologia web na telessaúde permitiu o desenvolvimento de portais e sistema de computação voltados ao diagnóstico, ambientes onde podem ser realizadas aquisições de sinais e posterior extração de medidas acústicas, como as lineares (PRAAT, 2022) e de recorrência. Estes serviços rapidamente demonstraram que possuem um custo efetivo de grande valia, permitindo o acesso facilitado das soluções tanto para os provedores de serviço quanto para os pacientes (DOARN *et al.*, 2019; EREN; WEBSTER, 2015).

Na área da fonoaudiologia, diversos esforços foram desenvolvidos para a criação de ferramentas de apoio ao diagnóstico (PRAAT, 2022; KTH, 2022; CTS, 2022). Para a utilização dessas ferramentas, é necessário, entretanto, que se faça o *download* e posterior instalação do programa no computador, restringindo-se o seu uso e guarda de informações obtidas por esse programa à máquina que o hospeda. Além disso, o programa em questão não possibilita a coleta de sinais de fala de maneira remota, dificultando a tarefa de acompanhamento diagnóstico por parte do profissional. Nestes sistemas, que pode ser gratuitos ou pagos são apresentados medidas como *shimmer*, *jitter*, MFCC. Devivo, entretanto, à natureza não linear do sistema de produção da voz, torna-se mais adequada a aplicação de análise que lhe seja compatível, ou seja, também não linear, como o enfoque na recorrência, portanto.

Cabe ressaltar que, realizada uma revisão na literatura, não foi possível localizar nenhuma ferramenta de suporte ao diagnóstico de desordens vocais a qual, por meio da Web, realizasse o acompanhamento de pacientes, a partir da utilização de métricas não lineares - como os gráficos e medidas de quantificação de recorrência - e que fosse classificada com sistema de suporte à decisão clínica. A abordagem de utilização da ferramenta que aqui defendida dispensa o desenvolvimento e instalação de *softwares* desenvolvidos para sistemas operacionais específicos, bastando apenas o utilizador da solução possuir um navegador, tal qual Mozilla Firefox, Google Chrome ou Safari. Essa ferramenta tem por objetivo propiciar o treinamento especializado de profissionais na área da voz, estando no mesmo contexto da plataforma Ouvindo Vozes, cuja função é treinar alunos e profissionais da Fonoaudiologia para que desenvolvam habilidades de percepção necessárias, a fim de que realizem o julgamento perceptivo-auditivo da voz, também por meio da Web, de forma confiável e consistente (LIEV, 2022).

Para além disso, com a ferramenta apresentada, é possível conduzir análises diversas sobre os mais variados aspectos, tanto computacionais, quanto ligados à área de processamento de sinais. A título de exemplificação, pode-se citar a análise de impactos na avaliação da qualidade vocal, devido ao uso de determinados codecs de áudio (CAVALCANTE, 2018), latência em aplicações Web (NAIK; JENKINS, 2016) ou com base nos custos associados à implantação de aplicações em ambientes de nuvem (VILLAMIZAR *et al.*, 2016).

Com o intuito de desenvolver uma solução *e-Health* moderna, na forma de um *software* como Serviço (*SaaS*), que atenda aos requisitos de demanda, escalabilidade e seja de fácil manu-

tenção, implementando uma arquitetura de microsserviços em uma infraestrutura de computação na nuvem propõe-se o Athena Acoustic Analyser (ALSHUQAYRAN; ALI; EVANS, 2016; WASEEM; LIANG; SHAHIN, 2020). Essa plataforma permite a modelagem e desenvolvimento da solução na forma de serviços, os quais se comunicam por *APIs* ou serviços de *mensageria* (FU, 2019; DONCA *et al.*, 2020), permitindo a criação de novas instâncias e seu balanceamento de carga na infraestrutura computacional, desenvolvendo o conceito de disponibilidade (SAYFAN, 2020; NEWMAN, 2022; FRASER; ZIADE, 2021).

## 1.2 Objetivos

### 1.2.1 Objetivo geral

Desenvolver uma solução de *e-Health* que possibilite a triagem e auxílio ao diagnóstico de pessoas com distúrbios ou patologias vocais bem como o treinamento de profissionais, utilizando análise acústica baseada em medidas lineares e de quantificação de recorrência em uma arquitetura monolítica num ambiente de Computação em Nuvem.

### 1.2.2 Objetivos específicos

- Realizar a análise e levantamento de requisitos relacionados às necessidades dos atores da solução;
- Modelar e implementar um modelo de dados, para as necessidade dos atores, representação dos sinais e dados na solução;
- Desenvolver serviços que implementem pré-processamento e análise de séries temporais, incluindo amostragem, janelamento e cálculo de medidas lineares e de quantificação de recorrência;
- Configurar uma infraestrutura computacional baseada em contêineres, possibilitando a sua escalabilidade e disponibilidade;
- Implementar a regra de negócio, na forma de serviços, que possibilite seu uso por intermédio da Internet e de uma interface voltada para Web;
- Realizar uma análise comparativa entre um sinal adquirido pelo sistema e existente na base de dados, utilizando a ferramenta proposta, enquadrando-a na classe de diagnóstico;
- Possibilitar o treinamento de profissionais na área da fala, utilizando a análise acústica, com o enfoque em quantificação de recorrência e suas medidas associadas.

## 1.3 Estrutura do Documento

Os capítulos subsequentes estão organizados da seguinte maneira:

- A fundamentação teórica é apresentada na Seção 2, descrevendo não só a parte matemática envolvida mas também a teoria envolvida no suporte e desenvolvimento da solução computacional.
- Na Seção 3 são apresentados a solução Athena Acoustic Analyser desenvolvida, as ferramentas computacionais utilizadas e os resultados obtidos.
- As considerações finais e as propostas de trabalhos futuros são descritos na Seção 4.



## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Princípios da anatomia, fisiologia e patologia vocal

#### 2.1.1 Anatomia e fisiologia da fala

A anatomia é o estudo da estrutura e a fisiologia é o estudo de como as estruturas funcionam para gerarem uma determinada função (SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022). No campo da voz, o entendimento da anatomia e fisiologia da laringe é de absoluta importância. A laringe é um tubo cartilaginoso que conecta o sistema respiratório inferior (traqueia e pulmões) ao trato superior, que são o trato vocal e a cavidade oral. Por conseguinte, é possível determinar três subsistemas: a fonte de energia pulmonar, a válvula laríngea e o ressoador supraglótico do trato vocal (STEMPLE; ROY; KLABEN, 2018).

No contexto da produção da voz, as estruturas respiratórias, como os pulmões, asseguram a força motriz necessária para iniciar e sustentar a vibração das pregas vocais. A vibração provê uma fonte de som para fonação, no processo de abertura e fechamento cíclico, modulando assim a pressão subglotal e o fluxo transglotal, na forma de pulsos curtos de energia (SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022).

O trato vocal serve como uma cavidade ressoante, que modela e filtra a energia acústica para produzir o som que reconhecemos como a voz humana. A laringe consiste em um complexo arranjo de cartilagens, músculos, tecidos conectivos e mucosa que permite uma alta taxa de variação na posição, movimento e tensão, suportando três funções básicas: a preservação da passagem do ar (ao abrir) para ventilação, proteção da passagem de ar (ao fechar), momento ou situação em que é bloqueada ou repele infiltrações, e a fonação (vibração das pregas vocais), a fim de se realizar a vocalização comunicativa e o canto (STEMPLE; ROY; KLABEN, 2018). Na função de fonação, a laringe promove a vibração das pregas vocais, a partir das mudanças de tensão e longitude dessas pregas, ampliação da abertura glótica e a intensidade do esforço respiratórios, processos que provocam variações no tom da voz (COSTA, 2012).

O movimento das pregas vocais é descrito em teorias como a da aerodinâmica-mioelástica de Van den Berg, em que o movimento se baseia no processo físico do fluxo induzido de oscilação, onde uma corrente contínua de ar flui pelos tecidos, criando um padrão repetitivo de abertura e fechamento. Para Van den Berg (BERG, 1958 apud STEMPLE; ROY; KLABEN, 2018) as propriedades do fluxo de ar e as musculares estão associadas, estão associadas aos movimentos passivos convergentes e divergentes das pregas vocais enquanto é realizada a fonação. Já no modelo de Hirano e Sato (1993), as pregas vocais são duas dobras de músculo e mucosa que se estendem horizontalmente na laringe, fixando-se anteriormente na face interna da cartilagem tireoidea e posteriormente na cartilagem aritenóidea, cobrindo-a de mucosa. No sistema fonatório,

um dos conceitos mais utilizados é que as pregas vocais são uma estrutura multilaminada, cujas camadas, cada uma a seu tempo, apresentam propriedades mecânicas diferentes (BEHLAU, 2001 apud COSTA, 2012).

### 2.1.2 Patologias laríngeas

Os distúrbios vocais podem ser provocados por patologias laríngeas que afetam a qualidade vocal. As patologias podem ser de origem orgânica ou neurológica e ainda podem ser classificadas como malignas ou benignas (SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022). Para que seja possível realizar uma validação do sistema proposto nesta dissertação, serão abordadas duas patologias de origem orgânica (edema de Reinke e nódulos vocais) e uma neurológica (paralisia das pregas vocais), embora não se descarte a possibilidade de, se necessário, serem adicionados outros tipos de patologia no avançar deste estudo.

Ademais, novas formas de classificar as patologias disfônicas estão em decurso, podendo ser discriminadas em comportamental e orgânica. A disfonia comportamental é caracterizada por alterações vocais relacionadas ao comportamento vocal do sujeito, a partir do uso inadequado da voz ou da exposição a fatores de risco para distúrbios de voz; a orgânica apresenta alterações teciduais ou de estrutura nos órgãos envolvidos na fonação ou em outros sistemas, as quais impedem a produção natural da voz, independentemente do comportamento vocal do indivíduo. Diante disso, os autores constataram que tanto os fatores genéticos quanto os fatores ambientais auxiliam no surgimento dos problemas de voz e essa interferência pode ser ainda maior quando o indivíduo desenvolve algum tipo de ocupação que exige muito esforço vocal (JANI *et al.*, 2008; SIMBERG *et al.*, 2009; ALENCAR *et al.*, 2020).

#### 2.1.2.1 Edema de Reinke

O Edema de Reinke ocorre quando a camada superficial da lâmina fica preenchida com um fluido viscoso, devido a um trauma contínuo por um longo período de tempo. Na forma mais severa, toda a porção da membrana das pregas vocais fica infiltrada com um fluido gelatinoso e grosso, apresentando uma aparência de bolsas ou balões preenchidos por fluidos. Na forma extrema, é chamado de degeneração polipoide. Como causas mais comuns do Edema de Reinke podem ser citados o abuso vocal e o tabagismo. A assinatura primária do Edema de Reinke é uma moderada disfonia, com a redução dramática do *pitch*, apresentando uma voz similar à de uma pessoa ébria. A cirurgia para remoção do Edema de Reinke é usual, sendo, entretanto, necessárias sessões de terapia, antes e depois da intervenção cirúrgica (STEMPLE; ROY; KLABEN, 2018).

#### 2.1.2.2 Nódulo

Os nódulos se apresentam como patologias benignas mais comuns, sendo mais registradas em pessoas que falam demais, ou muito alto, gritam, tosse ou pigarreiam com frequência. Os grupos mais propensos à afetação dessa patologia são os das mulheres, das crianças, de

peessoas que atuam como líderes de torcida, de atores, cantores, entre tantos outros grupos que utilizam a voz como instrumento ou recurso profissional. Os nódulos são inflamações na camada superficial da lâmina, com associação de fibrose e edema. Usualmente se formam bilateralmente e se desenvolvem na margem ou na junção nas partes anterior e nos dois terços médios das cordas vocais, áreas que recebem alto impacto durante seu funcionamento. São descritos três tipos de nódulos: agudo, crônico e reativo à mudança nodular (SAPIENZA; RUDDY; HOFFMAN-RUDDY, 2022). As pessoas acometidas por nódulos relatam dor na região do pescoço, necessidade de pigarrear e dificuldade de atingir determinadas frequências (VIEIRA, 2014).

### 2.1.2.3 *Paralisia*

Na categoria das desordens vocais neurológicas, a paralisia é relacionada à interrupção da inervação nervosa, o que acaba comprometendo a provisão de estímulos centrais e periféricos à laringe. Dependendo da localização da lesão no nervo, se determina o tipo de paralisia e a qualidade da voz daí resultante. Quanto mais próxima do envolvimento do nervo vago (nervo craniano X), ou seja, da sua origem no tronco cerebral, mais a paralisia afetará os nervos laríngeos superior e recorrente. Se o dano for, entretanto, mais distal, é possível que afete apenas um nervo (STEMPLE; ROY; KLABEN, 2018). Os sintomas mais perceptíveis da paralisia unilateral são a sopro e rouquidão, sendo ocasionalmente presente a diplofonia. Já a paralisia bilateral do tipo adutor causará sopro severa ou afonia (COSTA, 2008).

### 2.1.3 *Etiologias dos distúrbios vocais*

As etiologias podem ser causadas pelos mais variados fatores. No serviço de atendimento na área da fala, são constantes os ingressos de pacientes, tendo em vista as motivações recorrentes para desordens vocais, o que exige a classificação dessas desordens, para melhor compreensão de sua origem ou natureza e, conseqüentemente, se emitir um diagnóstico correto. As etiologias de trauma, são originadas de maus comportamentos que contribuem para as desordens vocais. Entre esses comportamentos, citam-se: gritar, falar alto, ruídos vocais e tosses.

Há, entretanto, outras etiologias que não provêm de ações contrárias ao bom desenvolvimento da vocalização. Trata-se das etiologias causadas por intervenções médicas ou cirúrgicas, relacionadas, portanto, a tratamentos que podem, indiretamente, contribuir para a manifestação de disfunções vocais. Algumas dessas etiologias são laringectomia, mandibulotomia, cirurgias na tireoide, cardíaca, intubação, trauma mecânico ou queimaduras (STEMPLE; ROY; KLABEN, 2018).

## 2.2 Das medidas tradicionais a longo intervalo de tempo e da análise preditiva linear

A análise objetiva, que inclui a análise acústica, é independente do viés humano e pode auxiliar os especialistas na tomada de decisões. A detecção de patologias pode ser realizada por vários tipos de análises, a longo intervalo de tempo ou a curto intervalo de tempo. Os parâmetros de longo intervalo de tempo são derivados da análise acústica (BOYANOV; HADJITODOROV, 1997; MARTINEZ; RUFINER, 2000) e os de curto intervalo de tempo podem ser calculados, utilizando-se os coeficientes de predição linear e os coeficientes de predição cepstral, por exemplo (AL-NASHERI *et al.*, 2017).

Várias características de curto intervalo de tempo, nominadamente o *pitch*, *shimmer* e *jitter*, consideradas métodos tradicionais (LOPES; CAVALCANTE; COSTA, 2014; LOPES *et al.*, 2020), são frequentemente utilizadas para a detecção de patologias na voz - a frequência fundamental (F0) corresponde ao número de vibrações cíclicas completas, realizadas pelas pregas vocais por segundo. Os indicadores *jitter* e *shimmer* - amplamente utilizados para propósitos de pesquisas clínicas, possibilitando identificar as pessoas que sofrem ou não de patologias laringeas (AL-NASHERI *et al.*, 2017; LOPES *et al.*, 2020).

A ideia básica da predição linear reside no fato de que a voz amostrada pode ser aproximada como uma combinação linear das amostras de voz em um momento anterior e de valores presentes de uma entrada hipotética de um sistema cuja saída é o sinal dado. No domínio da frequência, o processo da predição linear é equivalente ao de modelar o sinal por um espectro de polos e zeros (COSTA *et al.*, 1994).

Por intermédio da minimização da soma das diferenças quadradas (sobre um intervalo finito) entre as amostras reais da fala e as amostras obtidas através da combinação linear das primeiras, um conjunto único de coeficientes do preditor pode ser determinado. Os coeficientes do preditor são os coeficientes de ponderação usados na combinação linear (FECHINE, 2000).

A filosofia da predição linear está intimamente relacionada com o modelo de produção da voz, apresentada na Seção 2.3, que mostra como o sinal de voz pode ser modelado como a saída de um sistema linear variante no tempo, composto por pulsos quase periódicos (para sons sonoros), ou ruído aleatório (para sons não sonoros). As técnicas de predição fornecem um método robusto, realizável e correto para estimação dos parâmetros que caracterizam o sistema linear variante com o tempo (CHENGALVARAYAN, 1999).

Os métodos de predição linear estão disponíveis na literatura de engenharia há um longo tempo e têm sido vastamente empregados, principalmente em sistemas de controle, automação, telecomunicações e teoria da informação e codificação (RABINER; SCHAFER, 1978)

As técnicas de predição linear também podem ser aplicadas à quantização para reduzir a taxa de bits na representação digital do sinal de voz (RABINER; SCHAFER, 1978).

O princípio básico da predição linear leva a um conjunto de técnica de análise que pode ser usado para estimar parâmetros da fala. Esse conjunto geral de técnicas é frequentemente denominado de Análise por Codificação Preditiva Linear ou Análise LPC (*Linear Prediction Coding*) (COSTA *et al.*, 1994).

Muitos sistemas de reconhecimento de fala e de locutor têm, tradicionalmente, utilizado os parâmetros obtidos da análise LPC, em virtude das vantagens que esses parâmetros propiciam em termos de generalização da envoltória espectral, independência do *pitch* das harmônicas e sua habilidade para modelar, razoavelmente bem, os picos espectrais (COSTA *et al.*, 1994).

Ainda que haja vantagens em sua aplicação, existe uma dificuldade associada à análise por predição linear: a de determinar um conjunto de coeficientes do preditor diretamente a partir do sinal de voz, a fim de se obter uma boa estimativa das propriedades espectrais desse sinal de voz. Devido à natureza variante do tempo do sinal de voz, os coeficientes do preditor devem ser estimados em segmentos de curtos intervalos de tempo (FECHINE, 2000).

Os coeficientes podem ser obtidos a partir da análise LPC, denominados coeficientes LPC, ou a partir de técnicas derivadas dessa análise, como os coeficientes cepstrais (FECHINE, 2000; COSTA, 2008).

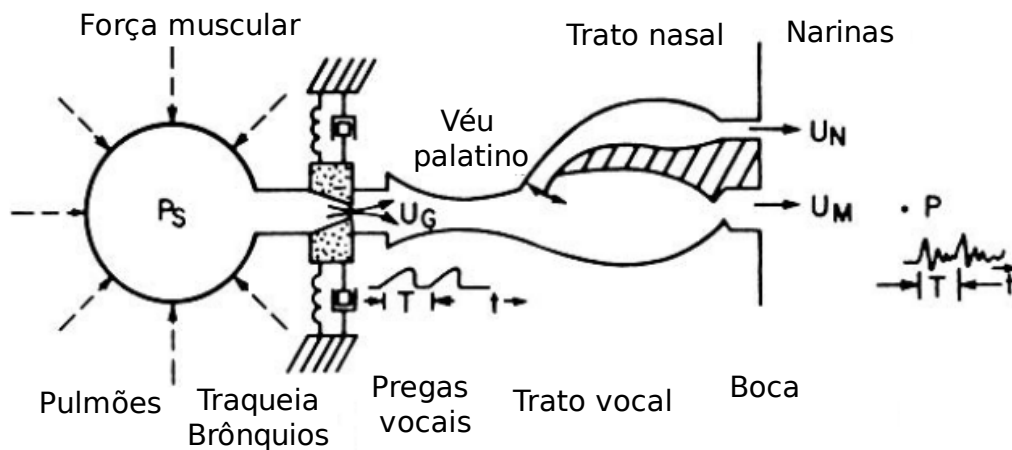
### 2.3 Modelo fonte/filtro de produção da fala e análises associadas

O estudo da linguagem e dos sons remontam aos gramáticos gregos e sânscritos dos séculos III e IV antes de Cristo. Os primeiros estudos explícitos relacionados à área da fala datam do século XVII, com Ferrein (1741), tentando explicar como as pregas vocais produzem a fonação. Outros desenvolveram estudos, como Kratzenstein (1780) e Kempelen (1791) tentaram explicar como as vogais e as consoantes da fala humana podiam ser produzidas pela sintetização de sons, utilizando máquinas artificiais chamadas de "máquinas falantes". No meio do século XIX, Muller (1848) formulou a teoria de fonte-filtro da produção da fala, que é consistente com os dados mais atuais e ainda guia as pesquisas na área da fala humana bem como na comunicação de outros animais. A continuação dos trabalhos de Muller passam para a Idade Moderna, quando, em 1930, dispositivos avançados para a época, como como espectrográfo sonoro, cineradiografia, e a eletromiografia permitiram a geração de novos dados. Estudos quantitativo, como os de Chiba e Kajiyama (1941), Joos (1948), Peterson e Barney (1952), Stevens e House (1955), e Fant (1960), refinaram e testaram as teorias tradicionais fonéticas do século XIX e proveram subsídios biológicos de fundamental importância para o desenvolvimento de estudos de produção, percepção e fonética da fala (LIEBERMAN; BLUMSTEIN, 1988).

Na Figura 1, é apresentado um corte longitudinal do mecanismo do trato vocal. O diagrama ressalta as partes físicas essenciais da anatomia humana que entram nos estágios finais do processo de produção da fala. O trato vocal como um tubo com área de seção reta, não uniforme, tendo, em uma das extremidades, as pregas vocais e, na outra, a abertura bucal. O trato

vocal, também chamado ramal principal, serve como um sistema de transmissão acústica para os sons ali gerados. Para criar sons nasais, como /M/, /N/, ou /NG/, o trato nasal é conectado a um ramal acústico principal, é conectado com ramal acústico principal pelo efeito de alçapão do véu palatino. Esse último ramal irradia o som pelas narinas. O formato (variação do corte transversal ao longo do eixo) do trato vocal varia com o tempo, devido às variações de lábios, mandíbula, língua e véu palatino (RABINER; SCHAFER, 2007).

Figura 1 – Modelo simplificado de produção da fala

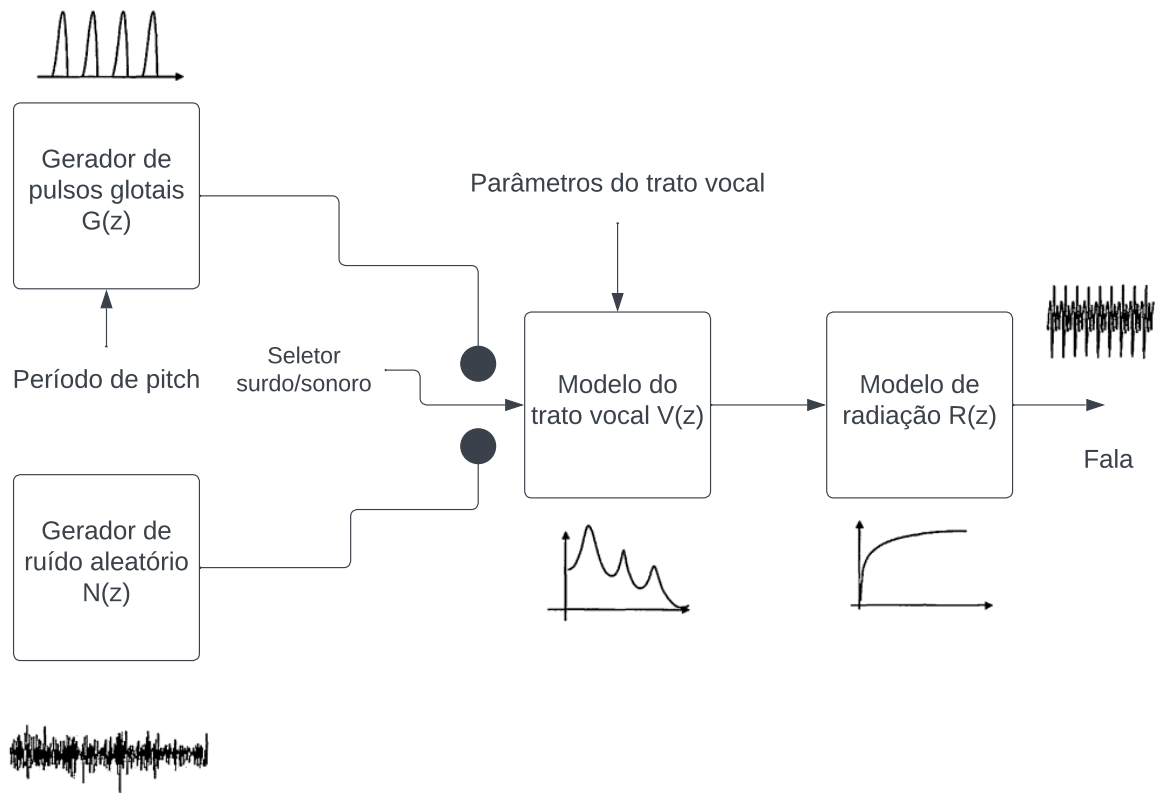


Fonte: Adaptado (RABINER; SCHAFER, 2007)

Como já abordado, o trato vocal pode ser visto como um filtro que molda espectralmente o fluxo das ondas que se originam nas pregas vocais para a produção de diversos tipos de sons. As pregas vocais produzem a excitação para o trato vocal, a qual pode ser periódica ou aperiódica, dependendo do estado das pregas vocais. Sons vozeados, ou sonoros, como vogais, são produzidos quando as pregas vocais estão no estado de vocalização (e vibram), enquanto os sons surdos, como algumas consoantes, são produzidas quando as pregas vocais estão no estado de não vocalização. Estes fatos sobre os papéis do trato vocal e pregas vocais possibilitaram que pesquisadores desenvolvessem um modelo de engenharia da produção da voz, como pode ser visualizado na Figura 2 mais adiante reproduzida (LOIZOU, 2005).

No modelo apresentado, o trato vocal é representado por um sistema quasi-linear que pode ser excitado por uma fonte periódica ou aperiódica, dependendo do estado das pregas vocais. A saída desse sistema é o sinal de voz, que é o único sinal que pode ser medido e/ou observado de forma mais acurada (LOIZOU, 2005).

Figura 2 – Modelo fonte e filtro representativo do trato vocal



Fonte: Adaptado (LOIZOU, 2005)

### 2.3.1 Seleção e manipulação de amostras em um sinal

#### 2.3.1.1 Amostragem

Em várias aplicações, como nas comunicações digitais, sinais analógicos são convertidos em sinais discretos, utilizando-se operações de amostragem e quantização. Com base na análise de Fourier, é possível descrever a operação de amostragem do ponto de vista do domínio da frequência, analisar seus efeitos e então realizar o processo de reconstrução do sinal (INGLE; PROAKIS, 2006).

Assumindo  $x_a(t)$  como um sinal analógico (e absolutamente integrável), a Transformada de Fourier de tempo contínuo é dada por

$$X_a(j\Omega) = \int_{-\infty}^{\infty} x_a(t) e^{-j\Omega t} dt \quad (1)$$

com  $\Omega$  sendo a frequência analógica em rad/s. A transformada inversa de tempo contínuo de Fourier é dada por

$$x_a(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_a(j\Omega) e^{j\Omega t} d\Omega. \quad (2)$$

O sinal  $x_a(t)$  é amostrado com intervalo de  $T_s$  segundos, com o objetivo de obter um sinal de tempo discreto  $x(n) = x_a(nT_s)$ .

Considerando  $X(e^{j\omega})$  como a Transformada de Fourier de tempo discreto de  $x[n]$ , é mostrado que  $X(e^{j\omega})$  tem uma soma contável em escala de amplitude e frequência, tendo versões transladadas da Transformada de Fourier  $X_a(j\Omega)$ .

$$X(e^{j\omega}) = \frac{1}{T_s} \sum_{l=-\infty}^{\infty} X_a\left[j\left(\frac{\omega}{T_s} - \frac{2\pi}{T_s}l\right)\right] \quad (3)$$

Essa relação é conhecida por fórmula de interpolação. As frequências analógicas e digitais estão relacionadas por intermédio de  $T_s$

$$\omega = \Omega T_s \quad (4)$$

enquanto que frequência de amostragem  $F_s$  é dada por

$$F_s = \frac{1}{T_s}, \quad \text{amostras/s} \quad (5)$$

A ilustração gráfica do processo realizado na Equação 3 é apresentada na Figura 3, de maneira geral, esclarecendo que o sinal discreto é uma versão interpolada do sinal analógico porque as frequências altas são interpoladas nas frequências baixas, se existir uma sobreposição. Entretanto, é possível recuperar a transformada de Fourier  $X_a(j\Omega)$  de  $X(e^{j\omega})$ , se suas réplicas infinitas de  $X_a(j\Omega)$  não se sobrepuserem para formar  $X(e^{j\omega})$ . Isto é verdade para sinais analógicos com banda limitada (INGLE; PROAKIS, 2006).

Cabe ressaltar que um sinal terá largura de banda limitada quando existir um frequência radiana  $\Omega_0$  quando  $X_a(j\Omega)$  é zero para  $|\Omega| > \Omega_0$ . A frequência  $F_0 = \Omega_0/2\pi$  é chamada de largura de banda do sinal em Hz. Além disso, um sinal com largura de banda  $F_0$  pode ser reconstruído a partir de suas amostras  $x(n) = x_a(nT_s)$ , se a frequência de amostragem  $F_s = \frac{1}{T_s}$  for maior que duas vezes a largura de banda  $F_0$  de  $x_a(t)$ .  $F_s \geq 2F_{MAX}$  define que a frequência de amostragem é duas vezes a máxima frequência encontrada no sinal. Esse processo gera uma taxa de amostragem, chamada de taxa de Nyquist, para um sinal de largura de banda limitado (INGLE; PROAKIS, 2006).

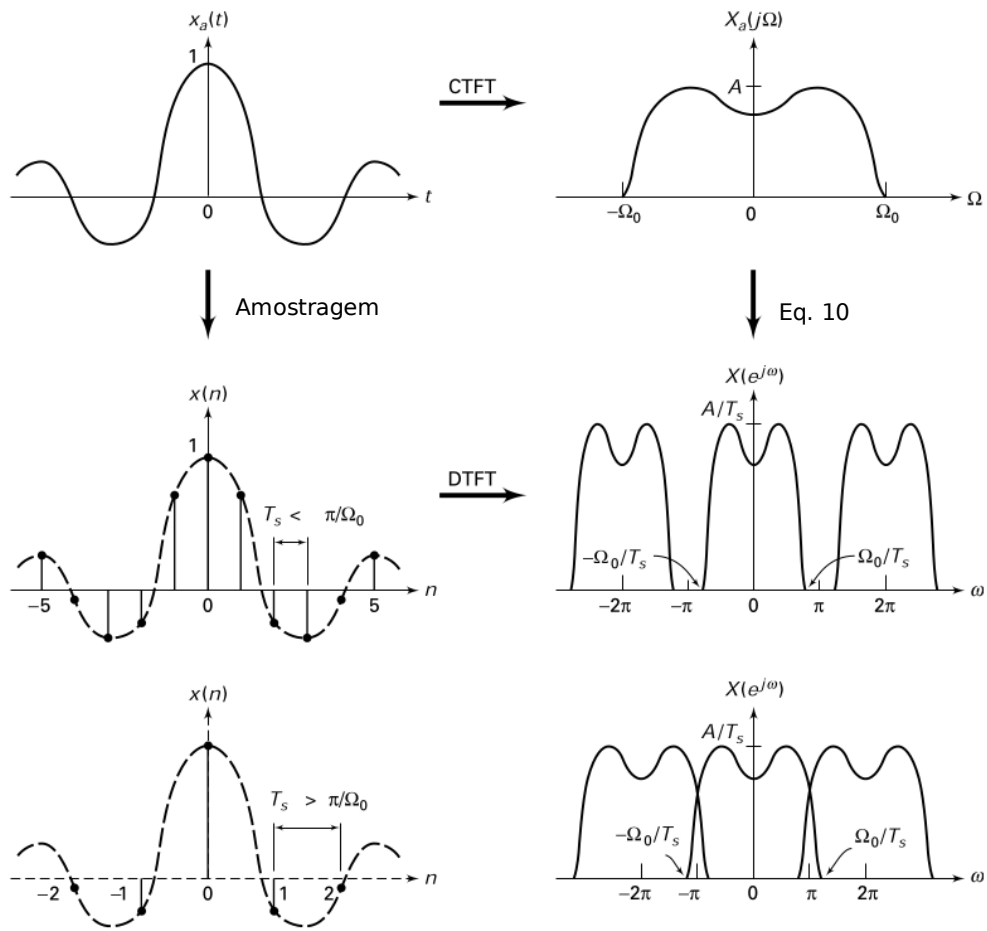
### 2.3.1.2 Janelamento

Em termos práticos de aplicações de processamento de sinais, é necessário trabalhar com pequenas seções ou quadros do sinal, a não ser que o sinal seja de curta duração. Isto é especialmente verdade, se são utilizadas técnicas convencionais de análise nos sinais (como a fala) como dinâmicas não estacionárias. Neste caso, é necessário selecionar a porção do sinal que possa ser considerada razoavelmente estacionária (DELLER; HANSEN; PROAKIS, 1999). Esta operação pode ser representada na forma geral por

$$X_{\hat{n}} = \sum_{m=-\infty}^{\infty} Tx[m]w[\hat{n} - m], \quad (6)$$



Figura 3 – Processo de amostragem de um sinal analógico



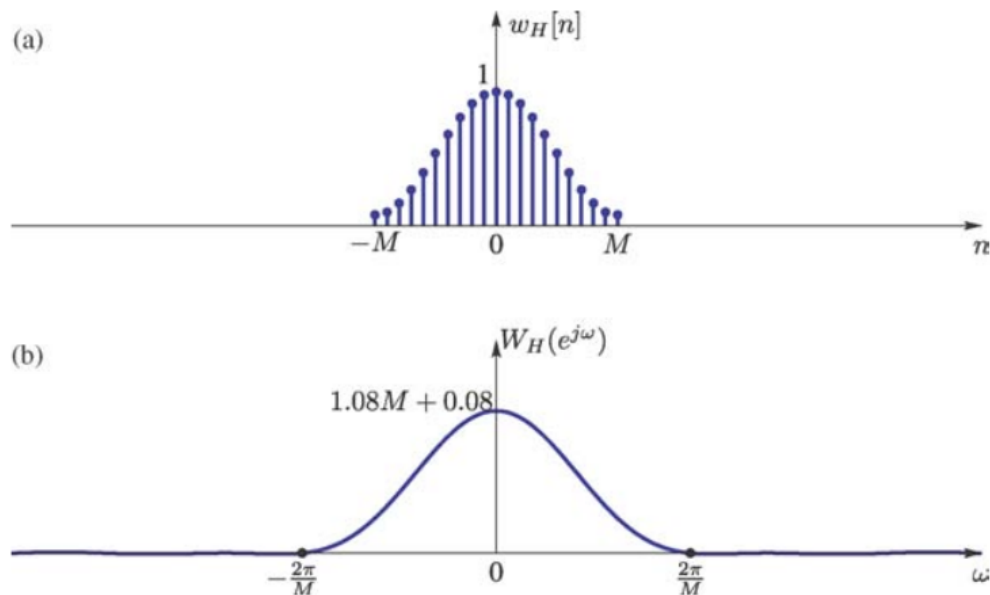
Fonte: Adaptado (INGLE; PROAKIS, 2006)

com  $X_{\hat{n}}$  representando o parâmetro da análise a curto intervalo de tempo (ou vetor de parâmetros) no tempo de análise  $\hat{n}$ . O operador  $T$  define a natureza da função de análise a curto intervalo de tempo, e  $w[\hat{n} - m]$  representa a sequência com a janela de tempo deslocada, com o propósito de selecionar os segmentos da sequência  $x[m]$  na vizinhança da amostra  $m = \hat{n}$ . Os limites infinitos na Equação 6 implica na soma de todos os valores não zeros do segmento janelado  $x_{\hat{n}}[m] = x[m]w[\hat{n} - m]$ ; i.e., para todos os  $m$  na região de suporte da janela. Uma das janelas de duração finita pode ser a janela de Hamming definida por

$$w_H[m] = \begin{cases} 0,54 + 0,46\cos(\frac{\pi m}{M}), & -M \leq m \leq M \\ 0, & \text{nos outros casos,} \end{cases} \quad (7)$$

A Figura 4 apresenta a janela de tempo discreto de Hamming e a sua Transformada de Fourier de tempo discreto. Esta Figura demonstra que a amostra  $(2M + 1)$  da janela de Hamming tem a banda do lóbulo da frequência principal de  $\frac{4\pi}{M}$ . Outras janelas possuem propriedades similares, i.e., isto é concentração em tempo e frequência, e a largura da janela em frequência será inversamente proporcional à largura da janela do tempo (OPPENHEIM, 1999).

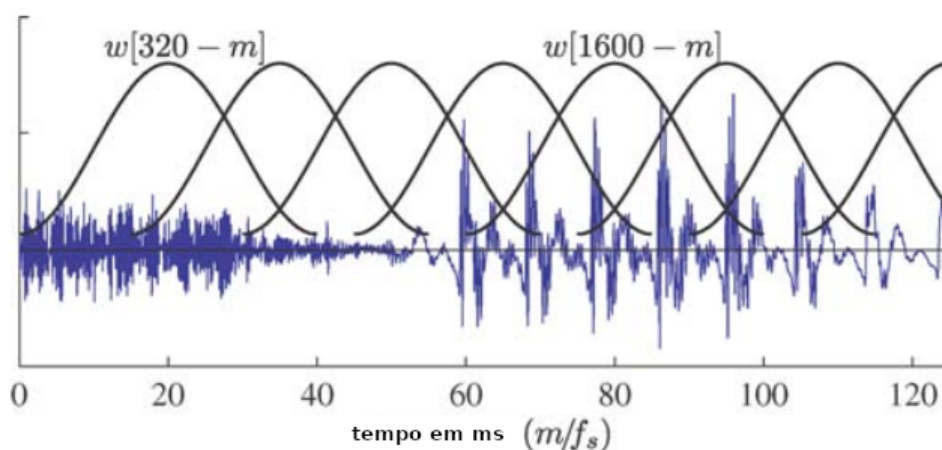
Figura 4 – Janela de Hamming (a) e sua Transformada de tempo discreto de Fourier (b)



Fonte: Adaptado (RABINER; SCHAFFER, 2007)

A Figura 5 apresenta um segmento da forma de onda da voz de 125 ms que inclui tanto fala do tipo surdo (0-50ms) quanto sonoro (50-125ms). Também apresenta a sequência de janelas de dados de duração de 40 ms com deslocamento de 15 ms (320 amostras a 16 kHz de taxa de amostragem) entre as janelas. Isto ilustra como a análise a curto intervalo de tempo é implementada.

Figura 5 – Seção da forma de onda da fala com janelas de análise a curto intervalo de tempo



Fonte: Adaptado (RABINER; SCHAFFER, 2007)

### 2.3.2 Frequência fundamental

A frequência fundamental é um dos mais importantes parâmetros da voz. Este inclui a altura e entonação de um sinal de voz, possuindo uma vasta variedade de aplicações nos

sistemas de processamento de sinais relacionados com a voz (ZHANG *et al.*, 2017). A frequência fundamental é a taxa da vibração das pregas vocais e é expressa em hertz (Hz) ou ciclos por segundo. Cabe ressaltar que a frequência fundamental é a medida acústica do seu percepto-auditivo correlato *pitch*, sendo que a frequência fundamental  $F_0$  é o inverso do período de *pitch*  $T_0$ . Para sujeitos com a fala normal, a digitalização das formas de onda da voz revelam um complexo, mas uma repetição quasi-periódica de períodos de *pitch* durante o tempo. A detecção do *pitch* permite, além de sua identificação em um sinal acústico também sua extração. Tais procedimentos podem ser feitos por meio de métodos específicos, a exemplo da seleção de picos, método da Função Média de Diferenças de Amplitudes (AMDF - *Average Magnitude Difference Function*) e método da função da autocorrelação (COSTA, 2008; STEMPLE; ROY; KLABEN, 2018).

### 2.3.2.1 AMDF - Average Magnitude Difference Function

O método AMDF convencional foi proposto por ROSS *et al.* (1974) e é definido por

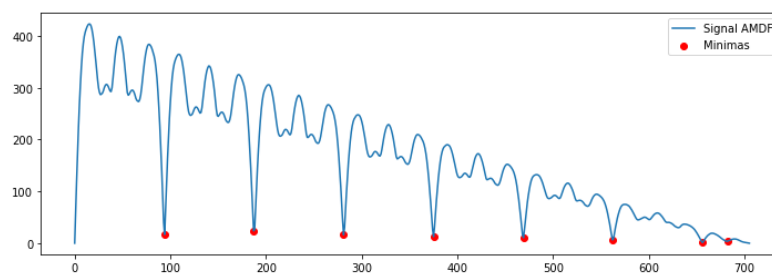
$$AMDF[m] = \sum_{n=0}^{N-m-1} |x[n] - x[n+m]|, \quad (8)$$

em que  $x[nm]$  denota um segmento sonoro multiplicado pela janela retangular de tamanho  $N$ , e  $m$  denota o índice de atraso. O intervalo de  $m$  está entre 0 e  $N - 1$ . Para um sinal periódico ou quasi-periódico com um período de  $T_0$ ,  $AMDF(m)$  deve exibir vales com o atraso  $T_p$  e em seus múltiplos. No contexto geral, o período de *pitch* bruto é estimado pela AMDF da seguinte forma

$$T_p = \arg \min_{m_{min}}^{m_{max}}(m) \quad (9)$$

considerando  $m_{max}$  e  $m_{min}$  são os possíveis valores máximo e mínimo do respectivo  $m$ . Com menos dados envolvidos para calcular  $D(m)$  nos atrasos maiores na Equação 8, é fácil presumir que uma tendência de queda é apresentada como uma característica global na curva da função. Além disso, um *pitch* incorreto pode ser detectado a partir do vale mínimo da AMDF. Em outras palavras, nem sempre o valor verdadeiro do *pitch* é menor vale (ZONG *et al.*, 2013). É possível observar tal método na Figura 6.

Figura 6 – AMDF de um sinal de voz saudável com os mínimos identificados



Fonte: Autor.

### 2.3.2.2 Função de Autocorrelação a Curto Intervalo de Tempo (FACIT)

A função de autocorrelação é ocasionalmente utilizada como meio para detectar a periodicidade em sinais, se configurando como um método básico para análise espectral, o que a torna uma ferramenta importante na análise da fala a curto intervalo de tempo, haja vista a facilidade em identificação de intervalos surdos e sonoros no sinal, tendo picos, na frequência, fundamental em um segmento sonoro, porém com ausência de picos em um segmento surdo. A FACIT é definida pela função de autocorrelação determinística da sequência  $x_{\hat{n}}[m] = x[m]w[\hat{n} - m]$  que é selecionada pela janela deslizante no tempo, isto é

$$\phi_{\hat{n}}[l] = \sum_{m=-\infty}^{\infty} x_{\hat{n}}[m]x_{\hat{n}}[m+l] = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n} - m]x[m+l]w[\hat{n} - m - l] \quad (10)$$

Utilizando-se uma propriedade de simetria par da autocorrelação,  $\phi_{\hat{n}}[-l] = \phi_{\hat{n}}[l]$ , a Equação 10 pode ser expressa em termos do filtro linear invariante no tempo (LTI) como

$$\phi_{\hat{n}}[l] = \sum_{m=-\infty}^{\infty} x[m]x[m-l]\tilde{w}_l[\hat{n} - m], \quad (11)$$

com  $\tilde{w}_l[m] = w[m]w[m+l]$ . É possível perceber que a FACIT é um função bi-dimensional com índice de tempo discreto  $\hat{n}$  (a posição da janela) e o índice  $l$  o atraso discreto. Se a janela tem duração finita, esta pode ser avaliada diretamente ou se utilizando técnicas de análise a curto intervalo de tempo. Para janelas exponenciais que decaem com duração infinita, a autocorrelação de curto intervalo de tempo da Equação 11 pode ser computada recursivamente no tempo  $\hat{n}$  utilizando-se um filtro diferente  $\tilde{w}_l[m]$  para valor de atraso  $l$  (RABINER; SCHAFER, 2007).

### 2.3.3 Faixa de frequência da voz humana

Para que a análise do sinal de voz seja possível, ou seja, que as características ou o estado do sistema produtor da fala sejam preservados no processo de captura, digitalização e disponibilização do sinal, é necessário que a faixa representativa do espectro de frequência seja capturada. Em virtude da limitação dos órgãos humanos de produção da voz e do sistema auditivo, a comunicação humana típica está limitada à faixa de 7-8 kHz (RABINER; SCHAFER, 1978).

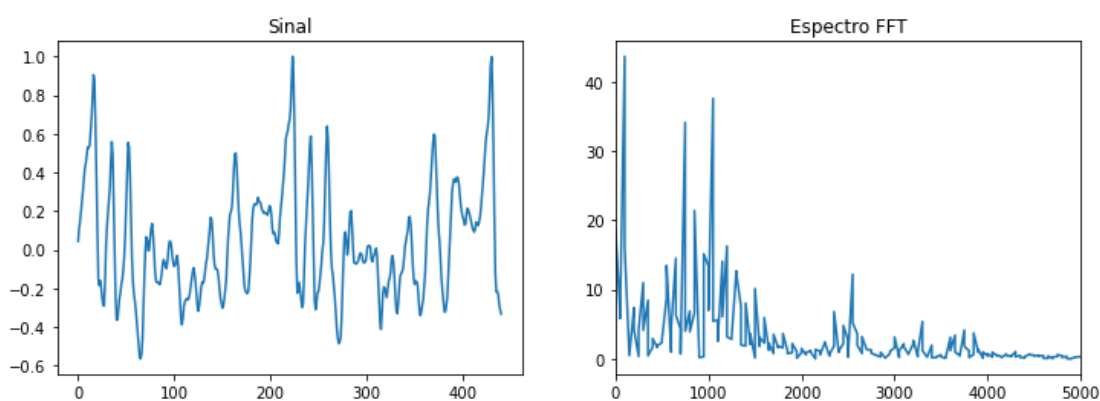
No caso da voz humana, utilizando-se uma amostra no estado normal, o menor valor para a componente é determinada pela frequência fundamental, que pode ir a patamares de 50 Hz em homens. Em situações de vozes roucas, contudo, pode ocorrer componentes sub harmônicos do som e estas frequências podem ir para patamares bem inferiores a 50 Hz (SVEC; GRANQVIST, 2010). A frequência fundamental de sons sonoros fica entre 80-120 Hz, para homens, 240 Hz para mulheres e 350 Hz para crianças (FELLBAUM, 1984; FECHINE, 2000)

### 2.3.4 Análise de Fourier

As vibrações senoidais são as formas mais simples de vibração. Quando adicionadas conjuntamente, as senoidais podem constituir assim todas as formas de vibração. A técnica

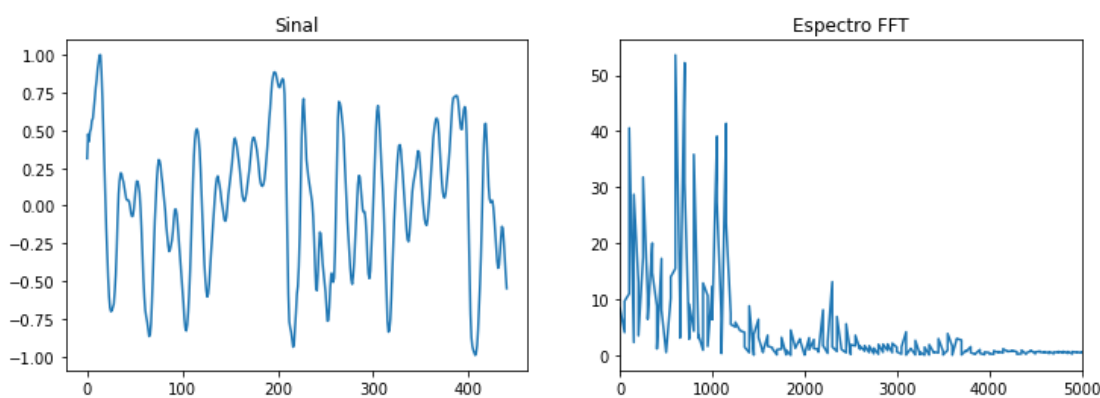
matemática de fatorar uma onda complexa em componentes senoidais é conhecida por análise de Fourier. Na Figura 9, é ilustrada a vibração periódica complexa consistindo de três componentes senoidais. A componente senoidal com a frequência mais baixa é a frequência fundamental, e as duas outras maiores são as segundas e terceiras harmônicas. Todas as três são componentes harmônicas da onda, com a frequência fundamental sendo a primeira. É possível perceber que, em ondas periódicas como as apresentadas na Figura 9, os valores de frequência das harmônicas são múltiplos inteiros da fundamental. Vibrações aperiódicas também podem ser analisadas como componentes sinusoidais, sendo chamadas de componentes de frequência. Entre os sinais vocais, os vocálicos são caracterizados como periódicos, enquanto os consonantais fricativos, como /f/ - /v/, por exemplo, são aperiódicos (CLARK; YALLOP; FLETCHER, 1995). Nas Figuras 7 e 8, é possível observar a representação de um sonoro, e sua transformada de Fourier, consistindo da vogal sustentada /a/ e de um sinal fricativo, representado pelo fonema /f/, num intervalo de 20 ms, com uma taxa de amostragem de 22050 Hz.

Figura 7 – Representação gráfica de um sinal sonoro e seu espectro de Fourier



Fonte: Autor.

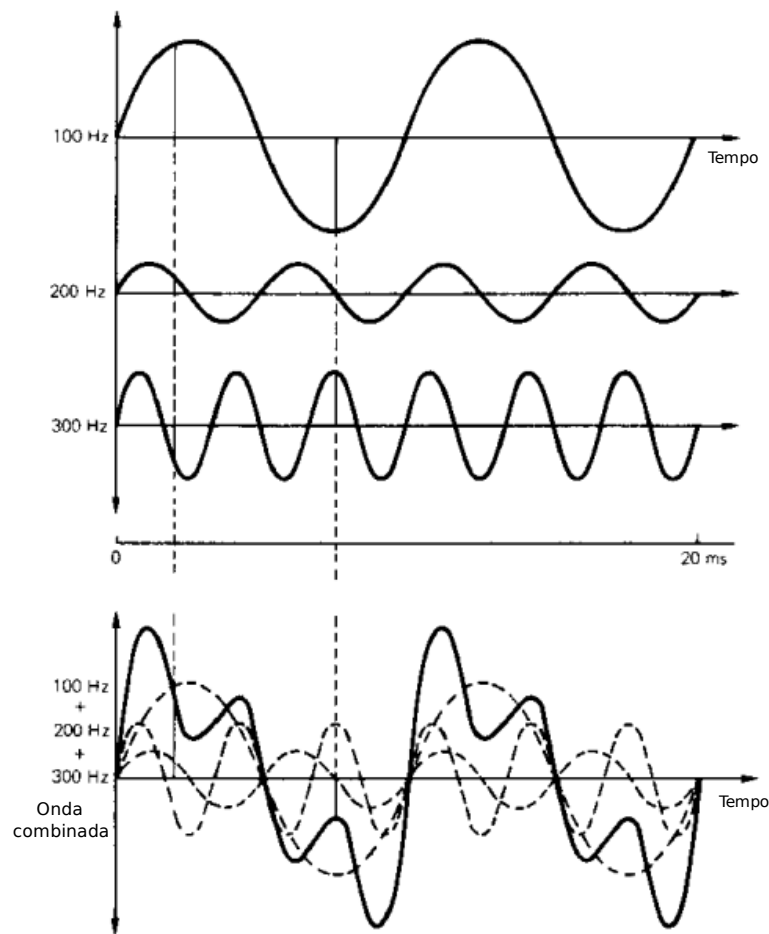
Figura 8 – Representação gráfica de um sinal fricativo e seu espectro de Fourier



Fonte: Autor.

A distribuição de frequência e amplitudes dos componentes harmônicos de uma onda pode ser representada por linhas no espectro. Neste modo, o eixo horizontal representa a

Figura 9 – Composição de um sinal senoidal



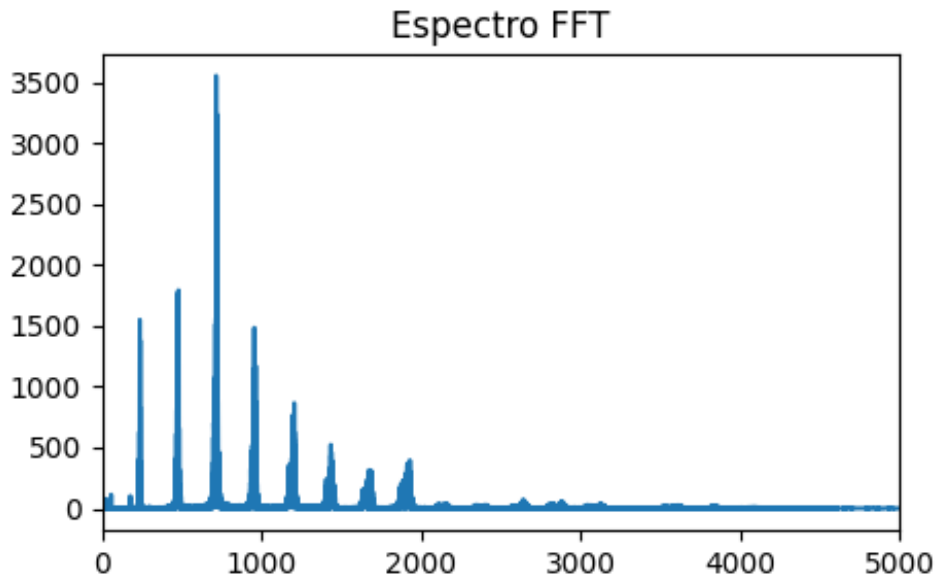
Fonte: Adaptado (CLARK; YALLOP; FLETCHER, 1995)

frequência, e o eixo vertical a amplitude. Cada harmônica é representada como uma única linha vertical, sendo sua altura a amplitude. O espectro computado a partir do sinal resultante da Figura 9 por ser visto na Figura 10 (CLARK; YALLOP; FLETCHER, 1995; KENT; KENT; READ, 2002).

Em sinais de voz extraídos de situações reais, o sinal sempre terá uma adição de ruído, seja este aperiódico ou não, do processo de digitalização ou até mesmo da irregularidade da frequência fundamental, conhecida por *jitter*. Titze (1995) sugere que sinais vozeados, como /a/, sejam classificados como aproximadamente periódico. Assim, diferentemente do espectro de um sinal considerado periódico, não existem linhas discretas, mas sim picos distintos no espectro contínuo, o que caracteriza uma forma de onda aperiódica. Cada pico pode ser considerado como sendo uma linha borrada (BAKEN; ORLIKOFF, 2000).

Além disso, os componentes de ruído aleatório adicionam uma pequena quantidade de energia em cada frequência e estas tendem a preencher os espaços entre as linhas das amplitudes do espectro. Esses picos possuem alturas variadas, refletindo na ideia de que linhas borradas do espectro têm diferentes amplitudes, assim como ocorre no espectro de um sinal sintético. É

Figura 10 – Espectro de Fourier de um sinal



Fonte: Autor.

possível perceber também que a amplitude dos picos não varia de forma aleatória em função da frequência: existem regiões com harmônicos mais evidentes e outras com harmônicos menos evidentes. Os harmônicos evidentes estão na vizinhança das frequências ressonantes do trato vocal, sendo chamadas de formantes (BAKEN; ORLIKOFF, 2000). Na Figura 11(b), é possível visualizar o espectro de Fourier.

#### 2.3.4.1 Transformada Discreta de Fourier (DFT)

A DTFT de uma sequência  $x[n]$  é uma função contínua de frequência  $\omega$ . Na prática, a sequência  $x[n]$  é finita na duração, consistindo de  $N$  amostras, e é possível amostrar a DTFT em  $N$  frequências espaçadas de maneira uniforme, tal que  $\omega_k = \frac{2\pi}{N}k, k = 0, 1, \dots, N-1$ . Essa amostragem produz uma nova transformada referenciada como transformada discreta de Fourier, do inglês *Discrete Fourier Transform* (DFT). A DFT de  $x[n]$  é dado por:

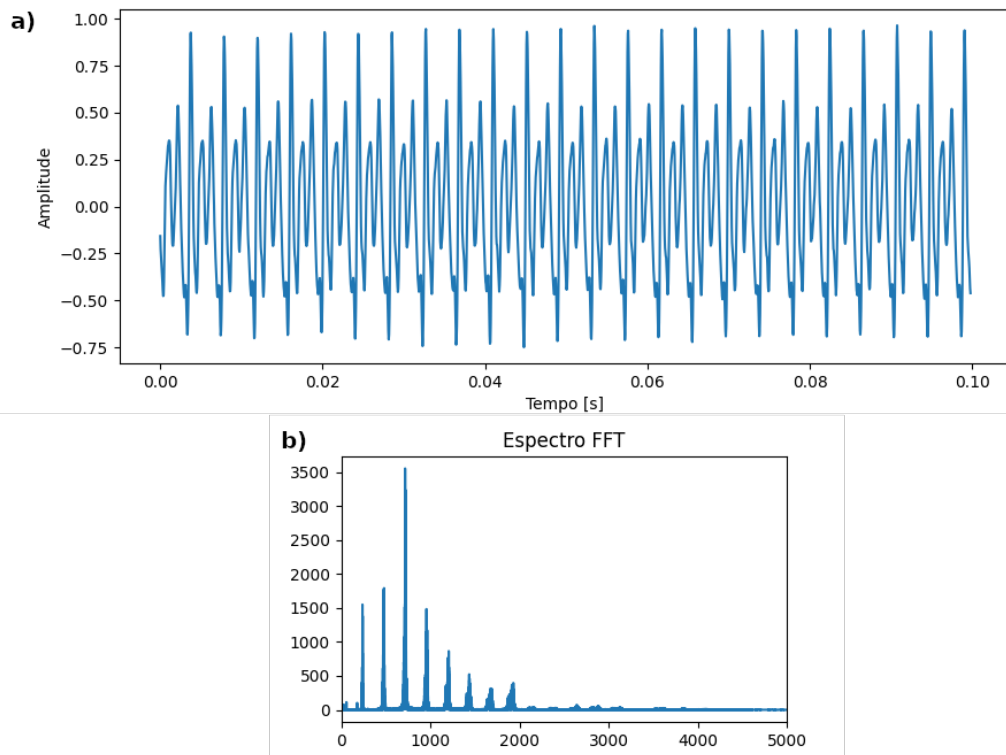
$$\tilde{X}[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi kn}{N}} \quad 0 \leq k \leq N-1 \quad (12)$$

Dado  $\tilde{X}[k]$ , é possível recuperar  $x[n]$  a partir de sua DFT, usando a DFT inversa (IDFT):

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}[k] e^{j\frac{2\pi kn}{N}} \quad 0 \leq n \leq N-1 \quad (13)$$

Muitas das propriedades da DFT são similares à DTFT. A principal diferença é que muitas das operações com a DFT são realizadas com módulo de  $N$ . Isto é possível devido à periodicidade implícita da sequência  $x[n]$  (LOIZOU, 2005).

Figura 11 – Trecho de um sinal real (a) e o espectro de Fourier (b)



Fonte: Autor.

#### 2.3.4.2 Transformada de tempo discreto de Fourier (DTFT)

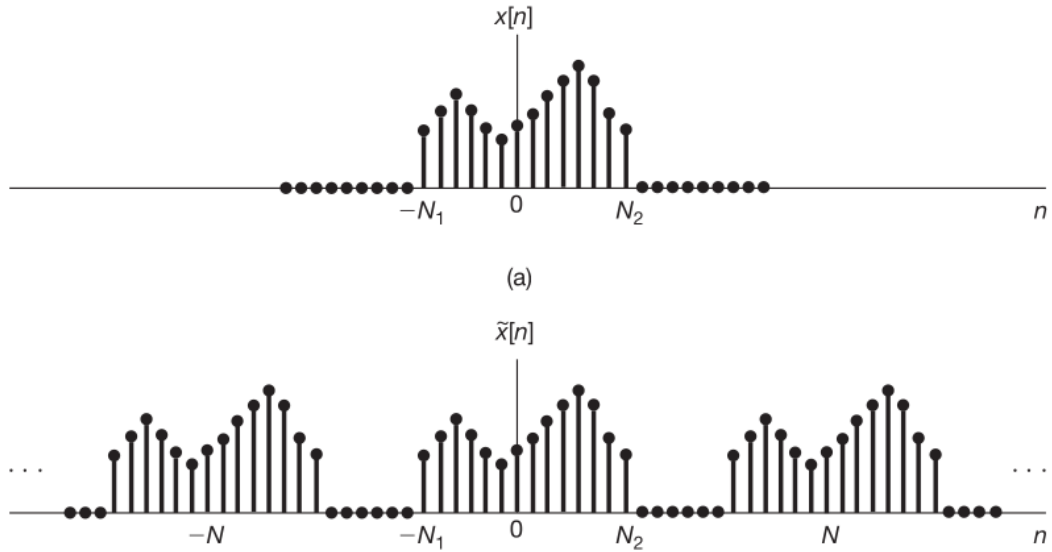
Considere-se uma sequência qualquer  $x[n]$  que tem duração finita, ou seja, para os inteiros  $N_1$  e  $N_2$  e  $x[n] = 0$  fora do intervalo  $-N_1 \leq n \leq N_2$ . Um sinal desse tipo é ilustrado na Figura 12(a). A partir desse sinal aperiódico, pode-se construir uma sequência periódica  $\tilde{x}[n]$  para a qual  $x[n]$  é um período, conforme ilustrado na 12(b). Quando escolhido o período  $N$  maior,  $\tilde{x}[n]$  é idêntico a  $x[n]$  por um intervalo maior, enquanto  $N \rightarrow \infty$ ,  $\tilde{x}[n] = x[n]$  para qualquer valor finito de  $n$ .

O par de equações de síntese e análise da série de Fourier de tempo discreto, descritas nas Equações 14 e 15, sendo análogos ao de tempo contínuo, os coeficientes da série de Fourier de tempo discreto  $a_k$  são frequentemente chamados coeficientes espectrais de  $x[n]$ . Esses coeficientes especificam a decomposição de  $x[n]$  em uma soma de  $N$  exponenciais complexas harmonicamente relacionadas.

$$x[\tilde{n}] = \sum_{k=\langle N \rangle} a_k e^{jk\omega_0 n} = \sum_{\langle k \rangle} a_k e^{jk(\frac{2\pi}{N})n} \quad (14)$$



Figura 12 – (a) Sinal  $x[n]$  de duração finita; (b) sinal periódico  $\tilde{x}[n]$  construído para ser igual a  $x[n]$  por um período



Fonte: (OPPENHEIN, 2010)

$$a_k = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] e^{-jk\omega_0 n} = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] e^{-jk(\frac{2\pi}{N})n} \quad (15)$$

A partir das Equações 14 e 15, tem-se:

$$\tilde{x}[n] = \sum_{k=\langle N \rangle} a_k e^{jk\omega_0 n} = \sum_{k=\langle N \rangle} a_k e^{jk(\frac{2\pi}{N})n} \quad (16)$$

$$a_k = \frac{1}{N} \sum_{n=\langle N \rangle} \tilde{x}[n] e^{-jk\omega_0 n} = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] e^{-jk(\frac{2\pi}{N})n} \quad (17)$$

Como  $x[n] = \tilde{x}[n]$  por um período que inclui o intervalo  $-N_1 \leq n \leq N_2$ , é conveniente escolher um intervalo do somatório na Equação 17 que inclua esse intervalo, de modo que  $\tilde{x}[n]$  possa ser substituído por  $x[n]$  no somatório. Portanto,

$$a_k = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] e^{-jk\omega_0 n} = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] e^{-jk(\frac{2\pi}{N})n} = \frac{1}{N} \sum_{n=-\infty}^{+\infty} x[n] e^{-jk(\frac{2\pi}{N})n} \quad (18)$$

sendo que, na segunda igualdade, usa-se o fato de que  $x[n]$  é nulo fora do intervalo  $-N_1 \leq n \leq N_2$ . Definindo a função

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n} \quad (19)$$

é visto que os coeficientes  $a_k$  são proporcionais às amostras de  $X(e^{j\omega})$ , ou seja,

$$a_k = \frac{1}{N} X(e^{jk\omega_0}) \quad (20)$$

em que  $\omega_0 = \frac{2\pi}{N}$  é o espaçamento entre as amostras no domínio de frequência. Combinando as Equações 14 e 20 resulta

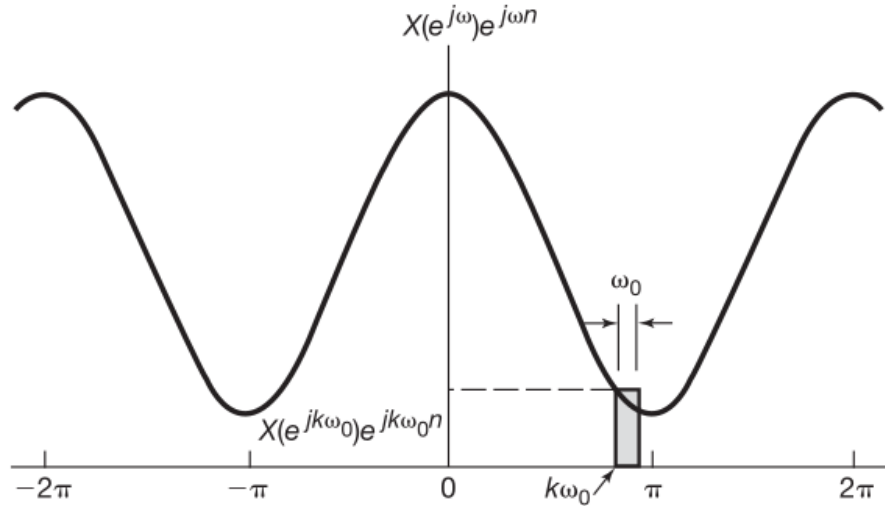
$$\tilde{x}[n] = \sum_{k=\langle N \rangle} \frac{1}{N} X(e^{jk\omega_0}) e^{jk\omega_0 n} \quad (21)$$

Como  $\omega_0 = \frac{2\pi}{N}$ , ou, de forma equivalente,  $\frac{1}{N} = \frac{\omega_0}{2\pi}$ , a Equação 21 pode ser reescrita como

$$\tilde{x}[n] = \frac{1}{2\pi} \sum_{k=\langle N \rangle} X(e^{jk\omega_0}) e^{jk\omega_0 n} \omega_0. \quad (22)$$

Quando  $N$  aumenta,  $\omega_0$  diminui, e quando  $N \rightarrow \infty$ , a Equação 22 torna-se uma integral. Para ver isso mais claramente, é necessário considerar  $X(e^{j\omega})e^{j\omega n}$ , conforme esboçado na Figura 13. Pela Equação 19, pode-se verificar que  $X(e^{j\omega})$  é periódico em  $\omega$  com o período  $2\pi$ , e o mesmo acontece com  $e^{j\omega n}$ . Portanto, o produto  $X(e^{j\omega})e^{j\omega n}$  também será periódico. Como representado na Figura 13, cada parcela do somatório da Equação 22 representa a área de um retângulo de altura  $x(e^{jk\omega_0})e^{jk\omega_0 n}$  e largura  $\omega_0$ . Quando  $\omega \rightarrow 0$ , o somatório torna-se uma integral.

Figura 13 – Interpretação gráfica da Equação 22



Fonte: (OPPENHEIN, 2010)

Além do mais, como o somatório é realizado sobre  $N$  intervalos consecutivos de largura  $\omega_0 = \frac{2\pi}{N}$ , o intervalo de integração total sempre terá uma largura de  $2\pi$ . Portanto, quando  $N \rightarrow \infty$ ,  $\tilde{x}[n] = x[n]$ , e a Equação 22 torna-se

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X(e^{j\omega}) e^{j\omega n} \omega \quad (23)$$

em que  $X(e^{j\omega})e^{j\omega n}$  é periódico com período  $2\pi$ , o intervalo de integração pode ser tomado como qualquer intervalo de comprimento  $2\pi$ . Assim, tem-se o seguinte par de equações:

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X(e^{j\omega}) e^{j\omega n} \omega \quad (24)$$

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-jn\omega} \quad (25)$$

A função  $X(e^{j\omega})$  é chamada de a transformada de Fourier de tempo discreto, e o par de equações denominado de par da transformada de Fourier de tempo discreto. A Equação 24 é a equação de síntese; a Equação 25, a equação de análise. A dedução dessas equações indica como uma sequência aperiódica pode ser considerada como uma combinação linear de exponenciais complexas. Em particular, a equação de síntese é, de fato, uma representação de  $x[n]$  como uma combinação linear de exponenciais complexas infinitesimalmente próximas em frequência e com amplitudes  $X(e^{j\omega})(\frac{\omega}{2\pi})$ . Por esse motivo, em tempo contínuo a transformada de Fourier  $X(e^{j\omega})$  comumente é chamada espectro de  $x[n]$ , pois nos oferece a informação sobre como  $x[n]$  é composto por exponenciais complexas em diferentes frequências (OPPENHEIM, 2010).

### 2.3.4.3 Transforma de Fourier a Curto Intervalo de Tempo (STFT)

A Transformada de Fourier a Curto Intervalo de tempo é definida como

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m] w[\hat{n} - m] e^{-j\hat{\omega} m}, \quad (26)$$

base para uma grande gama de análises de sistemas relacionados à voz, codificação e síntese. Por definição, para um tempo fixo de análise  $\hat{n}$ , a STFT é a Transformada de Tempo Discreto de Fourier (DTFT) do sinal  $x_{\hat{n}}[m] = x[m] w[\hat{n} - m]$ , i. e., a DTFT do sinal (usualmente de duração finita) selecionado e com amplitude ponderada por uma janela móvel  $w[\hat{n} - m]$ . Além disso, a STFT é uma função de duas variáveis;  $\hat{n}$  o índice de tempo discreto denotado pela posição da janela, e  $\hat{\omega}$  representando a frequência de análise. Como a Equação 26 é a sequência de DTFTs, a função de duas dimensões  $X_{\hat{n}}(e^{j\hat{\omega}})$  no tempo discreto  $\hat{n}$  é a função periódica da frequência radiana contínua  $\hat{\omega}$  com período  $2\pi$  (OPPENHEIM, 1999; RABINER; SCHAFER, 2007).

Como no caso de outras análises a curto intervalo de tempo, a STFT pode ser expressa em termo de uma operação de filtragem linear. Por exemplo, a Equação 27 pode ser expressa como uma convolução discreta

$$X_{\hat{n}}(e^{j\hat{\omega}}) = (x[n] e^{-j\hat{\omega} n}) * w[n] |_{n=\hat{n}}, \quad (27)$$

ou, alternativamente, 28 pode ser expressa como uma convolução discreta

$$X_{\hat{n}}(e^{j\hat{\omega}}) = (x[n] * (w[n] e^{-j\hat{\omega} n})) * e^{-j\hat{\omega} n} |_{n=\hat{n}}. \quad (28)$$

Lembrando que uma janela típica como a de Hamming, quando vista como um filtro linear de resposta ao impulso, tem uma resposta como um filtro passa-baixas em frequência com uma

frequência de corte variando inversamente ao comprimento da janela. Isto significa que um valor fixo de  $\hat{\omega}$ ,  $X_{\hat{n}}(e^{j\hat{\omega}})$  varia lentamente quando  $\hat{n}$  varia. A Equação 27 pode ser interpretada da seguinte forma: a modulação em amplitude  $x[n]e^{-j\hat{\omega}n}$  desloca o espectro de  $x[n]$  para baixo pelo fator de  $\hat{\omega}$ , e o filtro-janela (passa-baixas) seleciona as bandas de frequências resultantes no entorno da frequência zero. Esta é, com certeza, a banda de frequências de  $x[n]$  que foram originalmente centradas na frequência de análise  $\hat{\omega}$ . Uma conclusão idêntica segue da Equação 28:  $x[n]$  é a entrada de um filtro passa-banda com a resposta ao impulso  $w[n]e^{j\hat{\omega}n}$ , que seleciona as frequências centradas em  $\hat{\omega}$ . Então a banda de frequências é deslocada para baixo pela modulação em amplitude com  $e^{-j\hat{\omega}n}$ , resultando novamente no mesmo filtro passa-baixas (OPPENHEIM, 1999).

Sumariamente, a STFT tem três interpretações: (1) É a sequência da Transformada de Fourier de tempo discreto do sinal janelado como segmentos, i.e., a função periódica de  $\hat{\omega}$  em cada posição da janela  $\hat{n}$ . (2) Para cada frequência  $\hat{\omega}$  como um  $\hat{n}$  variante, é a sequência temporal de saída de um filtro passa-baixas que segue o corte em frequências dado por  $\hat{\omega}$ . (3) Para cada frequência  $\hat{\omega}$ , é a sequência temporal de saída do corte em frequências do filtro passa-banda (RABINER; SCHAFER, 2007).

#### 2.3.4.4 Análise espectral

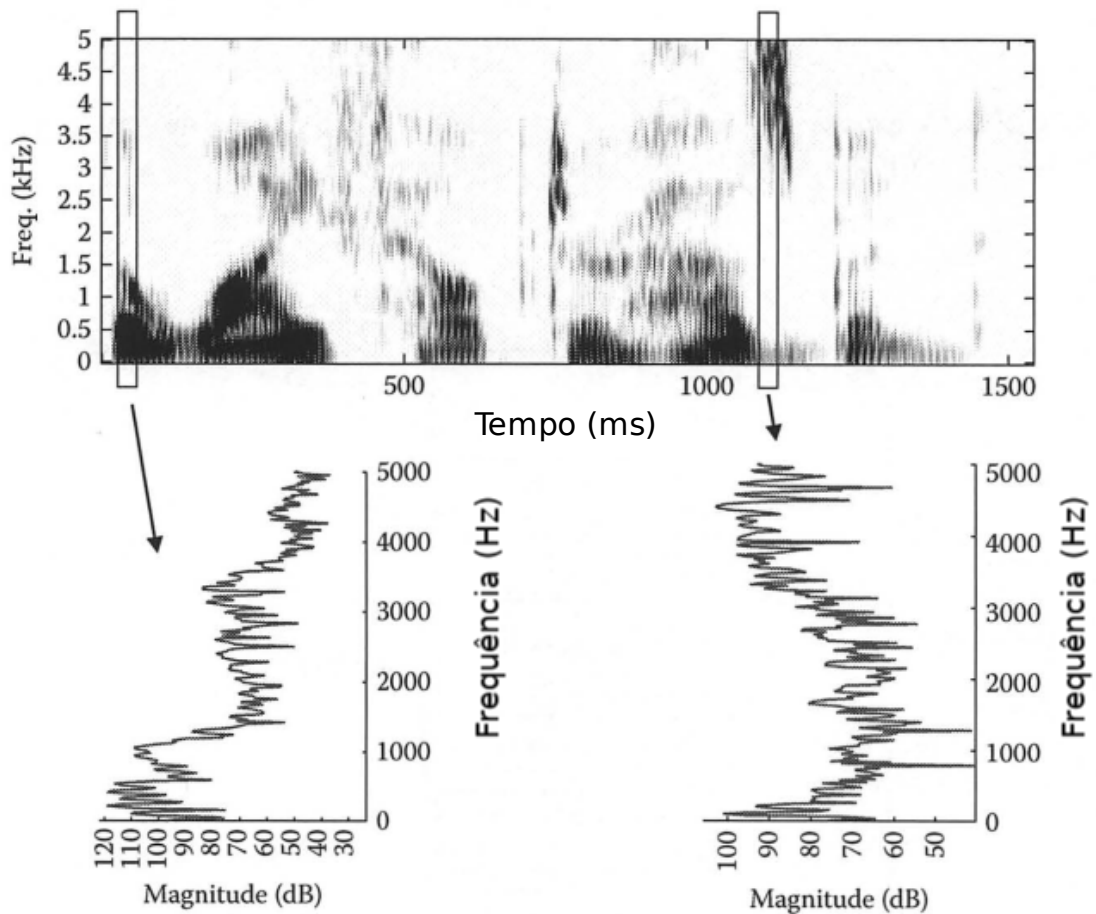
A transformada de Fourier a curto intervalo de tempo é utilizada ocasionalmente para análise de sinais de voz. O espectrograma é uma forma de apresentação gráfica do espectro de potência em função do tempo e dado por:

$$S(t_r, f_k) = 20 \log_{10} |\hat{X}_{rR}[k]| = 20 \log_{10} |X_{rR}[k]|, \quad (29)$$

com os eixos do gráfico rotulados em termos do tempo e frequência, por intermédio das relações  $t_r = rRT$  e  $f_k = \frac{k}{NT}$ , onde  $T$  é o período do sinal de tempo discreto  $x[n] = x_a(nT)$ ,  $f_k$  é o conjunto de frequência no tempo  $k$ . A quantidade  $S(t_r, f_k)$  pode ser visualizada como densidade espectral de potência bi-dimensional, sendo o tempo a segunda. O espectrograma descreve a concentração de energia relativa do sinal de voz na frequência como função do tempo e, sendo assim, reflete as propriedades variantes no tempo da forma de onda da voz. Espectrogramas podem ser apresentados em escala de cinza ou coloridos. Quanto maior a energia no espectro num frequência específica, mais escura a apresentação daquela determinada frequência. Grandes magnitudes no espectro, como picos, são apresentadas em preto ou em cores escuras, enquanto vales são apresentados em cores ou tons de cinza mais claros. É possível verificar a variação de cores na Figura 14. Valores variando entre picos e vales são apresentados utilizando-se uma variedade de tons de cinza (LOIZOU, 2005).

Dois tipos de espectrogramas, banda estreita e banda larga, podem ser produzidos, dependendo do tamanho da janela utilizada na computação de  $S(n, \omega)$ . Uma janela de longa duração, de pelos menos dois períodos de *pitch*, é tipicamente utilizada na computação do

Figura 14 – Espectrograma apresentando altas e baixas frequências



Fonte: Adaptado (LOIZOU, 2005)

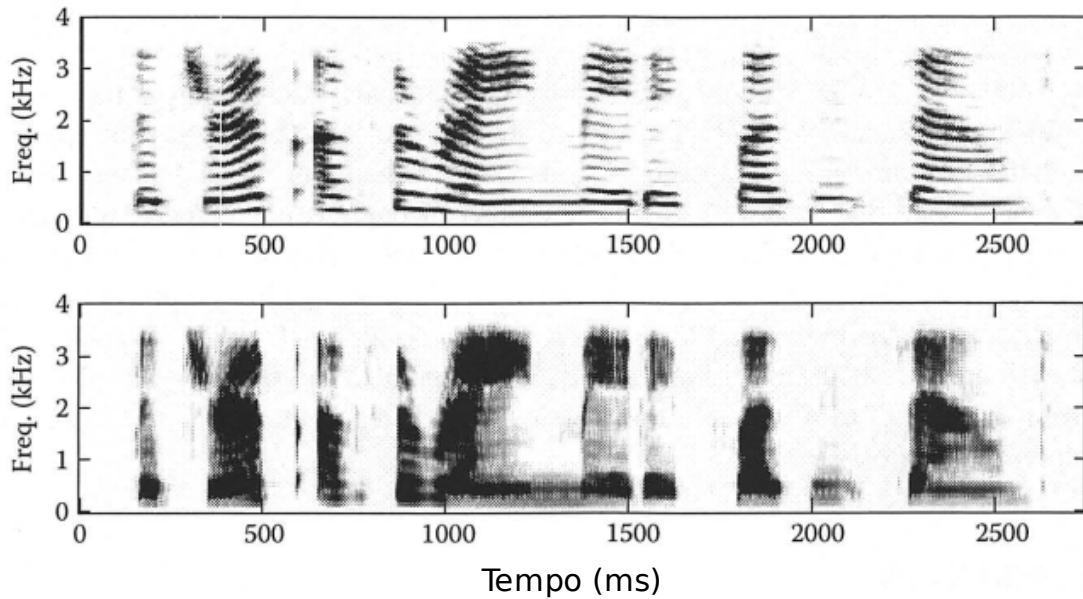
espectrograma de banda larga. O espectrograma de banda estreita dá uma boa resolução em frequência, mas uma péssima resolução no domínio do tempo. A resolução de frequência otimizada permite a visualização dos harmônicos individuais da fala. Estes harmônicos se apresentam como estrias horizontais no espectrograma, os quais podem ser visualizados na parte superior da Figura 15 (LOIZOU, 2005; RABINER; SCHAFER, 2007).

Já o espectrograma de banda larga usa janelas de curta duração (menor que um período de *pitch*) e dá uma boa resolução temporal, mas uma péssima resolução em frequência. A principal consequência de uma resolução pobre em frequência é o borramento de harmônicos individuais no espectro da fala, produzindo o envelope do espectro, como pode ser visto na parte inferior da Figura 15 (LOIZOU, 2005).

### 2.3.5 Análise cepstral

O cepstro foi definido por Bogert, Healy, e Tukey, sendo a transformada inversa de Fourier do logaritmo da magnitude do espectro do sinal (BOGERT, 1963). A definição original, vagamente enquadrada nos termos da análise espectrográfica de sinais, foi motivada pelo fato

Figura 15 – Espectrograma faixa estreita (superior) e faixa larga (inferior)



Fonte: Adaptado (LOIZOU, 2005)

de que o logaritmo do espectro de Fourier de um sinal contém um eco, o qual pode ser uma componente aditiva, dependendo apenas de seu tamanho e tempo de atraso. As consequentes análises do espectro logarítmico podem ajudar na detecção de presença desse eco. Oppenheim, Schaffer e Stockham desmostraram que o cepstro está mais relacionado ao conceito geral da filtragem homomórfica de sinais que são combinados pela convolução (OPPENHEIM, 1965; OPPENHEIM; SCHAFFER; STOCKHAM, 1968; SCHAFFER, 1969). A definição do cepstro de um sinal discretizado no tempo como

$$c[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|X(e^{j\omega})| e^{jn\omega} d\omega \quad (30)$$

com  $\log|X(e^{j\omega})|$  é o logaritmo da magnitude da DTFT do sinal, e esta estende o conceito pela definição de cepstro complexo como

$$\hat{x}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\{X(e^{j\omega})\} e^{jn\omega} d\omega, \quad (31)$$

com  $\log X(e^{j\omega})$  é o logaritmo complexo de  $X(e^{jn\omega})$  definido como

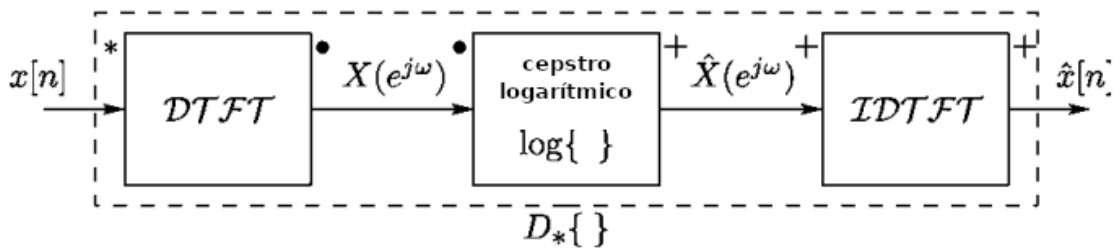
$$\hat{X}(e^{j\omega}) = \log|X(e^{j\omega})| + j\arg[X(e^{j\omega})]. \quad (32)$$

A transformação implicada pela Equação 31 é apresentada no diagrama de bloco na Figura 16. Esse mesmo diagrama representa o cepstro se o logaritmo complexo é substituído pelo logaritmo da magnitude da DTFT. Como foi restringida a atenção para sequências reais,  $x[n]$ , que é uma das propriedades da transformada de Fourier, é a parte par do cepstro complexo, *i.e.*,  $c[n] = \frac{(\hat{x}[n] + \hat{x}[-n])}{2}$  (OPPENHEIM, 1999). Como apresentado na Figura 16, a operação de

computação do cepstro complexo da entrada pode ser denotada como  $\hat{x}[n] = D_*\{x[n]\}$ . Na teoria dos sistemas homomórficos  $D_*\{\}$  é chamado de sistema característico para convolução. A conexão entre o conceito de cepstro e a filtragem homomórfica de sinais convoluídos é que o cepstro complexo tem a propriedade que se  $x[n] = x_1[n] * x_2[n]$  logo

$$\hat{x}[n] = D_*\{x_1[n] * x_2\} = \hat{x}_1[n] + \hat{x}_1[n] + \hat{x}_2[n] \quad (33)$$

Figura 16 – Computando o cepstro complexo utilizando a DTFT



Fonte: Adaptado (RABINER; SCHAFER, 2007)

Neste caso, o operador cepstro complexo transforma convolução em adição. Essa propriedade, verdadeira para ambos os cepstros, é útil para a análise da fala, envolvendo, no modelo de produção desta última, a convolução da excitação com a resposta ao impulso do trato vocal, sendo também objetivo ocasional a separação do sinal de excitação do sinal do trato vocal. No caso do cepstro complexo, a inversa do sistema característico existe como apresentado na Figura 17, demonstrando a cascata reversa das inversas dos operadores apresentados na Figura 16. A filtragem homomórfica de sinais convoluídos é alcançada pela formação de um cepstro complexo modificado, conforme se vê em:

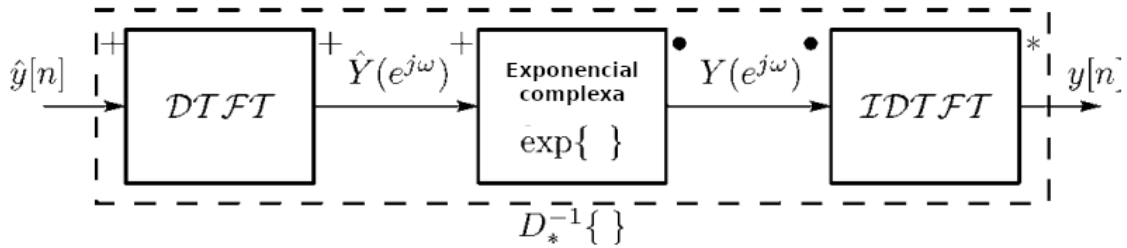
$$\hat{y}[n] = g[n]\hat{x}[n], \quad (34)$$

sendo  $g[n]$  a janela (um *lifter* na terminologia de Bogert et al.) que seleciona a porção do cepstro complexo no processamento inverso. O sinal de saída modificado  $y[n]$  pode ser obtido pela saída com  $\hat{y}[n]$  dado pela Equação 34 como entrada, podendo ser observado o processo na Figura 17. Essa equação define um operador linear de forma convencional, *i.e.*, se  $\hat{x}[n] = \hat{x}_1 + \hat{x}_2[n]$  então  $\hat{y}[n] = g[n]\hat{x}_1 + g[n]\hat{x}_2[n]$ . Entretanto, a saída do sistema característico inverso terá a forma  $y[n] = y_1[n] * y_2[n]$ , tendo  $\hat{y}_1[n] = g[n]\hat{x}_1[n]$  que é o cepstro complexo de  $y_1[n]$ , etc.

O ponto chave na definição do cepstro complexo é a computação do logaritmo complexo; mais especificamente, a computação do ângulo de fase  $arg[X(e^{j\omega})]$ , o qual deve ser concluído mantendo a combinação aditiva das fases dos dois sinais combinados pela convolução (OPPENHEIM; SCHAFER; STOCKHAM, 1968; SCHAFER, 1969).

A variável independente do cepstro e do cepstro complexo é nominadamente o *tempo*. A observação crucial que leva ao conceito de cepstro é que o espectro logarítmico pode ser tratado como uma forma de onda passível de realização da análise de Fourier. Para enfatizar

Figura 17 – A inversa do sistema característico para convolução (cepstro complexo inverso)



Fonte: Adaptado (RABINER; SCHAFER, 2007)

este intercâmbio de domínios de referência, Bogert et al. (BOGERT, 1963) cunhou a palavra *cepstro* pela transposição de algumas letras na palavra *espectro*. Os autores criaram vários outros termos especiais nesta forma incluindo *quefrência* como nome da variável independente do cepstro e *liftering* para a operação de filtragem linear do espectro logarítmico de magnitude pela operação descrita na Equação 34. Apenas os termos cepstro, quefrência e *liftering* são utilizadas amplamente atualmente (RABINER; SCHAFER, 2007).

### 2.3.5.1 O cepstro a curto intervalo de tempo

A aplicação dessas definições na fala requer que a DTFT seja substituída pela STFT. Por isso, o cepstro a curto intervalo de tempo é definido por

$$c_{\hat{n}}[m] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X_{\hat{n}}(e^{j\hat{\omega}})| e^{j\omega m} d\hat{\omega}, \quad (35)$$

com  $X_{\hat{n}}(e^{j\hat{\omega}})$  sendo a Transformada Rápida de Fourier a Curto Intervalo de Tempo (STFT), e o Cepstro Complexo a Curto Intervalo de Tempo é definido da mesma maneira pela substituição de  $X(e^{j\hat{\omega}})$  por  $X_{\hat{n}}(e^{j\hat{\omega}})$  na Equação 31. A similaridade com FACIT deve ser clara. O cepstro a curto intervalo de tempo é a sequência do cepstro de segmentos janelados de duração finita da forma de onda da voz. Por analogia, um "cepstrograma" será uma imagem obtida pela plotagem da magnitude do cepstro a curto intervalo de tempo como função da quefrência  $m$  e tempo de análise  $\hat{n}$  (RABINER; SCHAFER, 2007).

### 2.3.5.2 Proeminência de pico cepstral - (CPP)

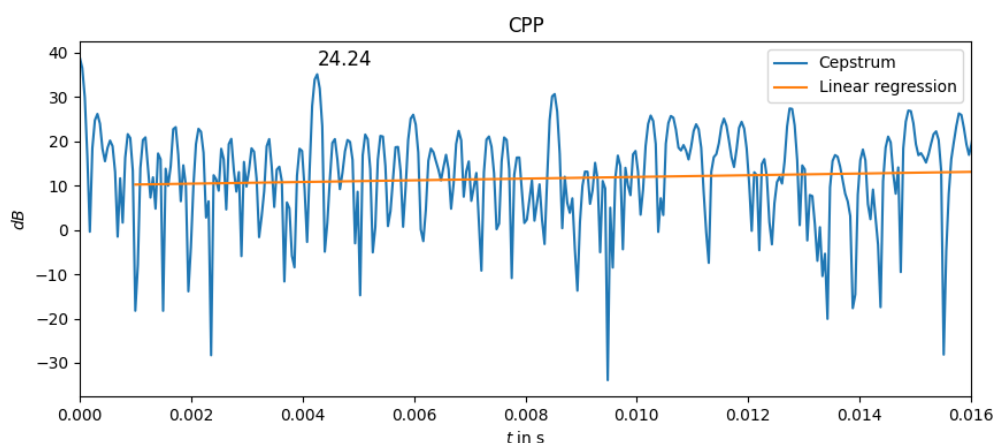
A Proeminência do Pico Cepstral, do inglês *Cepstral Peak Proeminence* (CPP), é a medida de amplitude do pico cepstral normalizado para uma amplitude média. O pico cepstral corresponde ao período fundamental. A ideia que subsidia a medida CPP é que um sinal altamente periódico deve mostrar uma harmônica bem definida e, conseqüentemente, um pico cepstral mais proeminente que um sinal menos periódico. O que é necessário é a medida de proeminência do pico ao invés da amplitude absoluta. Isso porque a amplitude do pico cepstral é afetada não somente pelo nível de periodicidade mas também pela energia média e o tamanho da



janela de análise (HILLENBRAND; CLEVELAND; ERICKSON, 1994; FRAILE; GODINO-LLORENTE, 2014).

Qualquer número de métodos pode ser utilizado para normalizar a amplitude do sinal. O método adotado para análises desta dissertação envolve o ajuste de uma curva de regressão linear relativa à quefrência e à magnitude cepstral. A linha é computada entre 1 ms e a máxima quefrência. Notadamente quefrências abaixo de 1 ms são ignoradas devido a estas componentes serem sensíveis ao formato do envelope espectral em detrimento da periodicidade do sinal. A medida CPP é a diferença na amplitude entra o pico cepstral e o valor correspondente na reta de regressão linear imediatamente abaixo do pico (i.e. a magnitude cepstral predita para a quefrência no pico cepstral). Em outras palavras, a medida CPP representa quão longe o pico cepstral emerge do "ruído de fundo" cepstral (HILLENBRAND; CLEVELAND; ERICKSON, 1994; FRAILE; GODINO-LLORENTE, 2014). Tal medida pode ser visualizada na Figura 18.

Figura 18 – CPP de um sinal classificado como normal



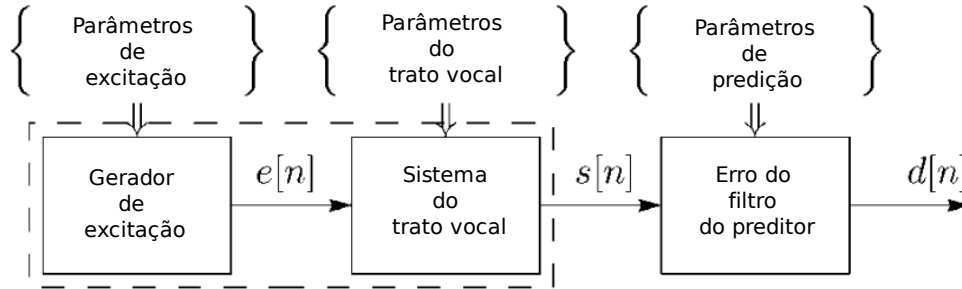
Fonte: Autor.

### 2.3.6 Análise preditiva linear

Embora a análise da voz por intermédio de espectrogramas baseados em Fourier seja a técnica mais comum de examinar as propriedades do espectro da fala, a análise por coeficientes de codificação por predição linear, do inglês *Linear Predictive Coding* (LPC), tem se demonstrado eficaz. A análise LPC representa o sinal de voz em termos de conjunto de coeficientes que tem por objetivo prever o sinal, a partir, dos valores anteriores no domínio do tempo, com o menor erro possível. Estes coeficientes podem ser utilizados para produzir uma representação espectral do sinal. Em essência, o processo captura o formato da resposta em frequência do filtro que representa o trato vocal, incluindo os efeitos de arredondamento e aproximação do espectro da fonte do sinal de voz. Além disso, deve ser incluído o processo de radiação do mecanismo de produção da fala, produzindo-se, assim, um formato da onda da fala no domínio do tempo que é similar ao formato de onda que será analisado (CLARK; YALLOP; FLETCHER, 1995).

Uma forma particular do modelo fonte e filtro implícito pela análise preditiva linear é apresentada na Figura 19. A área do modelo da fala corresponde à parte tracejada tracejada (RABINER; SCHAFER, 2007).

Figura 19 – Modelo preditivo linear



Fonte: Adaptado (RABINER; SCHAFER, 2007)

Em intervalos curtos de tempo, o sistema linear é descrito por uma função de um sistema só polos na forma:

$$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}}. \quad (36)$$

Na análise preditiva linear, a excitação é definida implicitamente pelo modelo do sistema de produção da fala, isto é, a excitação é qualquer estímulo necessário para produzir  $s[n]$  como saída do sistema. A maior vantagem desse modelo é o parâmetro de ganho,  $G$ , e os coeficientes do filtro  $a_k$ , que pode ser estimados de forma direta e de maneira computacionalmente eficiente pelo método da análise preditiva linear (RABINER; SCHAFER, 2007).

Para o sistema apresentado na Figura 19 com o modelo do trato vocal da Equação 36, as amostras da voz  $s[n]$  são relacionadas à excitação  $e[n]$  pela equação de diferenças

$$s[n] = \sum_{k=1}^p a_k s[n-k] + Ge[n]. \quad (37)$$

Um preditor linear com coeficientes de predição,  $\alpha_k$ , é definido como um sistema cuja saída é

$$\tilde{s}[n] = \sum_{k=1}^p a_k s[n-k], \quad (38)$$

e o erro de predição, definido como a quantidade a qual  $s[n]$  falha para prever de maneira exata a amostra  $s[n]$ , é

$$d[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^p a_k s[n-k]. \quad (39)$$

Da Equação 39, segue que o erro de predição da sequência é a saída de sistema linear FIR, tendo como função do sistema

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} = \frac{D(z)}{S(z)}. \quad (40)$$

É possível a comparação entre as Equações 37 e 39, a partir das quais se vê que o sinal de voz obedece ao modelo da Equação 37 exatamente, e se  $\alpha_k = a_k$ , então  $d[n] = Ge[n]$ . Além disso, o erro do filtro preditivo,  $A(z)$ , será um filtro inverso do sistema,  $H(z)$ , da Equação 36, isto é,

$$H(z) = \frac{G}{A(z)} \quad (41)$$

O problema básico da análise por predição linear é a determinação de um conjunto de coeficientes de predição  $\alpha_k$  diretamente do sinal de voz, com o objetivo de obter uma estimativa útil do sistema do trato vocal, que varia conforme o tempo, a partir do uso da Equação 41. A abordagem básica é achar um conjunto de coeficientes de predição que irá minimizar o erro médio de predição quadrático em um curto segmento da forma de onda da fala (RABINER; SCHAFER, 2007).

O erro de predição médio a curto intervalo de tempo é definido como

$$E_{\hat{n}} = \langle d_{\hat{n}}^2[m] \rangle = \langle (s_{\hat{n}}[m] - \sum_{k=1}^p \alpha_k s_{\hat{n}}[m-k])^2 \rangle, \quad (42)$$

com  $s_{\hat{n}}[m]$  sendo um segmento de fala que foi selecionado na vizinhança do tempo de análise  $\hat{n}$ , isto é,

$$s_{\hat{n}}[m] = s[m + \hat{n}] \quad -M_1 \leq m \leq M_2. \quad (43)$$

Isto posto, a origem temporal do segmento analisado é deslocada para a amostra  $\hat{n}$  de todo o sinal. A notação  $\langle \rangle$  denota a criação de uma média sob um número finito de amostras (RABINER; SCHAFER, 2007).

É possível encontrar valores para  $\alpha_k$  que minimizam  $E_{\hat{n}}$  na Equação 42 pela definição  $\frac{\partial E_{\hat{n}}}{\partial \alpha_i} = 0$ , para  $i = 1, 2, \dots, p$ , obtendo as equações

$$\sum_{k=1}^p \tilde{\alpha}_k \langle s_{\hat{n}}[m-i] s_{\hat{n}}[m-k] \rangle = s_{\hat{n}}[m-i] s_{\hat{n}}[m] \quad 1 \leq i \leq p, \quad (44)$$

sendo  $\tilde{\alpha}_k$  são os valores que minimizam  $E_{\hat{n}}$  em 42. Como  $\tilde{\alpha}_k$  são únicos, a remoção do til e o uso da notação  $\alpha_k$  denotam os valores que minimizam  $E_{\hat{n}}$ . Definindo

$$\varphi_{\hat{n}}[i, k] = s_{\hat{n}}[m-i] s_{\hat{n}}[m-k], \quad (45)$$

então a Equação 44 pode ser reescrita de maneira compacta como

$$\sum_{k=1}^p \alpha_k \varphi_{\hat{n}}[i, k] = \varphi_{\hat{n}}[i, 0] \quad i = 1, 2, \dots, p. \quad (46)$$

Se é sabido  $\varphi_{\hat{n}}[i, k]$  para  $i \leq i \leq p$  e  $0 \leq k \leq p$ , este conjunto de  $p$  equações em  $p$  incógnitas, que pode ser representadas pela matriz de equações:

$$\Phi \alpha = \psi \quad (47)$$

pode ser resolvida por um vetor  $\alpha = \{\alpha_k\}$  de coeficientes de predição desconhecidos que minimizam o erro médio quadrático de predição para o segmento  $s_{\hat{n}}[m]$ . Utilizando as Equações 42 e 44, o erro médio quadrático mínimo pode ser mostrado por:

$$E_{\hat{n}} = \varphi_{\hat{n}}[0,0] - \sum_{k=1}^p \alpha_k \varphi_{\hat{n}}[0,k]. \quad (48)$$

Desta forma, o erro quadrático mínimo consiste numa componente fixa igual ao valor da média quadrática do segmento do sinal menos o termo que depende dos coeficientes do preditor que satisfazem a Equação 46, isto é, coeficientes ótimos que reduzem  $E_{\hat{n}}$ , na Equação 48, ao máximo (RABINER; SCHAFER, 2007).

Para que seja possível determinar os coeficientes ótimos do preditor, deve-se primeiro computar as quantidades  $\varphi_{\hat{n}}[i,k]$  para  $1 \leq i \leq p$  e  $0 \leq k \leq p$ . Quando isto é concluído, resta apenas resolver a Equação 46 para obter os  $\alpha_k$ s. Desta forma, no princípio, a análise preditiva linear é bastante direta. Entretanto, os detalhes da computação de  $\varphi_{\hat{n}}[i,k]$  e a subsequente solução das equações de alguma forma estão intrincadas, e discussões adicionais são necessárias (RABINER; SCHAFER, 2007).

A análise preditiva linear e a Transformada de Fourier estão relacionadas a dois aspectos: a primeira à representação no tempo; a segunda, à frequência. A diferença principal é que o espectro de Fourier representa os harmônicos da frequência fundamental, enquanto o espectro LPC representa as frequência formantes e amplitudes (ou ressonâncias) (KENT; KENT; READ, 2002).

### 2.3.6.1 Método da covariância

Um forma de calcular os coeficientes de predição é baseada na seguinte definição:

$$E_{\hat{n}} = \sum_{m=-M_1}^{M_2} (d_{\hat{n}}[m])^2 = \sum_{m=-M_1}^{M_2} (s_{\hat{n}}[m] - \sum_{k=1}^p \alpha_k s_{\hat{n}}[m-k])^2, \quad (49)$$

com  $-M_1 \leq n \leq M_2$ . As quantidades  $\varphi_{\hat{n}}[i,k]$  necessárias na Equação 46 herdam a mesma definição do operador da média, isto é,

$$\varphi_{\hat{n}}[i,k] = \sum_{m=-M_1}^{M_2} s_{\hat{n}}[m-i]s_{\hat{n}}[m-k] \quad \begin{cases} 1 \leq i \leq p \\ 0 \leq k \leq p. \end{cases} \quad (50)$$

Ambas Equações - 49 e 50 - requerem valores de  $s_{\hat{n}}[m] = s[m + \hat{n}]$  no intervalo  $-M_1 - p \leq m \leq M_2$ . Com as mudanças dos índices do somatório, a Equação 50 pode ser expressa na forma equivalente:

$$\varphi_{\hat{n}}[i,k] = \sum_{m=-M_1-i}^{M_2-i} s_{\hat{n}}[m]s_{\hat{n}}[m+i-k] \quad (51a)$$

$$= \sum_{m=-M_1-k}^{M_2-k} s_{\hat{n}}[m]s_{\hat{n}}[m+k-i] \quad (51b)$$

das quais é obtido que  $\varphi_{\hat{n}}[i, k] = \varphi_{\hat{n}}[k, i]$  (RABINER; SCHAFER, 2007).

### 2.3.6.2 Método da autocorrelação

O método mais utilizado para análise preditiva linear é chamada de método da autocorrelação porque a função de covariância  $\varphi_{\hat{n}}[i, k]$ , necessária na Equação 46, pode ser reduzida a uma autocorrelação a curto intervalo de tempo  $\phi_{\hat{n}}[|i - k|]$ . No método da autocorrelação, a análise do segmento  $s_{\hat{n}}[m]$  é definida como

$$s_{\hat{n}}[m] = \begin{cases} s[n+m]w[m], & -M_1 \leq m \leq M_2 \\ 0, & \text{nos outros casos,} \end{cases} \quad (52)$$

nos outros casos, a janela de análise  $w[m]$  é utilizada para suavizar as bordas do segmento para zero. Como a análise do segmento é definido pelo janelamento da Equação 52 para ser zero fora do intervalo  $-M_1 \leq m \leq M_2$ , é verificado que sequência do erro de predição  $d_{\hat{n}}[m]$  não pode ser zero apenas no intervalo  $-M_1 \leq m \leq M_2 + p$ . Entretanto,  $E_{\hat{n}}$  é definido por:

$$E_{\hat{n}} = \sum_{m=-M_1}^{M_2+p} (d_{\hat{n}}[m])^2 = \sum_{m=-\infty}^{\infty} (d_{\hat{n}}[m])^2 \quad (53)$$

O janelamento da Equação 52 permite o uso de limites que tendem ao infinito para significar que o somatório é de todos os valores não nulos de  $d_{\hat{n}}[m]$ . Aplicando-se essa notação, as Equações 51a e 51b levam à conclusão de que

$$\varphi_{\hat{n}}[i, k] = \sum_{m=-\infty}^{\infty} s_{\hat{n}}[m]s_{\hat{n}}[m + |i - k|] = \phi_{\hat{n}}[|i - k|]. \quad (54)$$

Desta forma,  $\varphi[i, k]$  é uma função somente de  $|i - k|$ . Entretanto, pode-se substituir  $\varphi_{\hat{n}}[i, k]$  por  $\phi_{\hat{n}}[|i - k|]$ , que é a Função de autocorrelação a curto intervalo de tempo, definida pela Equação 11 como

$$\phi_{\hat{n}}[k] = \sum_{m=-\infty}^{\infty} s_{\hat{n}}[m]s_{\hat{n}}[m + k] = \phi_{\hat{n}}[-k]. \quad (55)$$

O conjunto resultante de equações para os coeficientes de predição otimizados é, entretanto, (RABINER; SCHAFER, 2007)

$$\sum_{k=1}^p \alpha_k \phi_{\hat{n}}[|i - k|] = \phi_{\hat{n}}[i] \quad i = 1, 2, \dots, p. \quad (56)$$

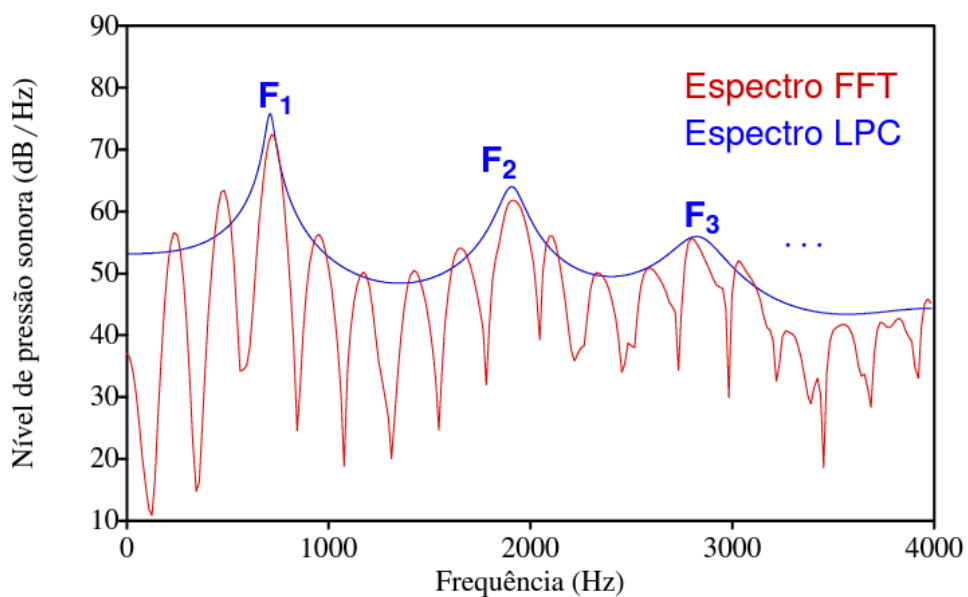
### 2.3.7 Frequências Formantes

Uma forma de reconhecer as regiões vocálicas de um sinal de voz é por intermédio da obtenção dos formantes, que são as frequências de ressonância do trato vocal, representadas por picos no espectro LPC, as quais refletem a configuração do trato no momento da fonação. Essas

são designadas por  $F_1, F_2, \dots, F_n$  (primeiro formante, segundo formante, ...,  $n$ -ésimo formante) (COSTA, 2008).

A obtenção da envoltória espectral pode ser realizada pela análise LPC. Na Figura 20, é possível visualizar o espectro LPC, mais suavizado que o espectro FFT, evidenciando os picos que correspondem aos formantes daquele instante de tempo do sinal de voz em análise. A técnica possibilita a visualização da resposta do trato vocal, extraindo justamente a informação de formantes que é de interesse para a análise (DRESCH, 2015).

Figura 20 – Espectro LPC (azul) e espectro da FFT (vermelho). As formantes são identificadas pelos picos  $F_1, F_2, F_3$



Fonte: Adaptado (DRESCH, 2015)

A quantidade máxima de coeficientes para a representação do sinal é realizada pelo valor da frequência de amostragem (em kHz) dividido por mil mais dois ( $n_{coef} = (f_s/1000) + 2$ ) (DRESCH, 2015).

## 2.4 A presença da não linearidade na voz

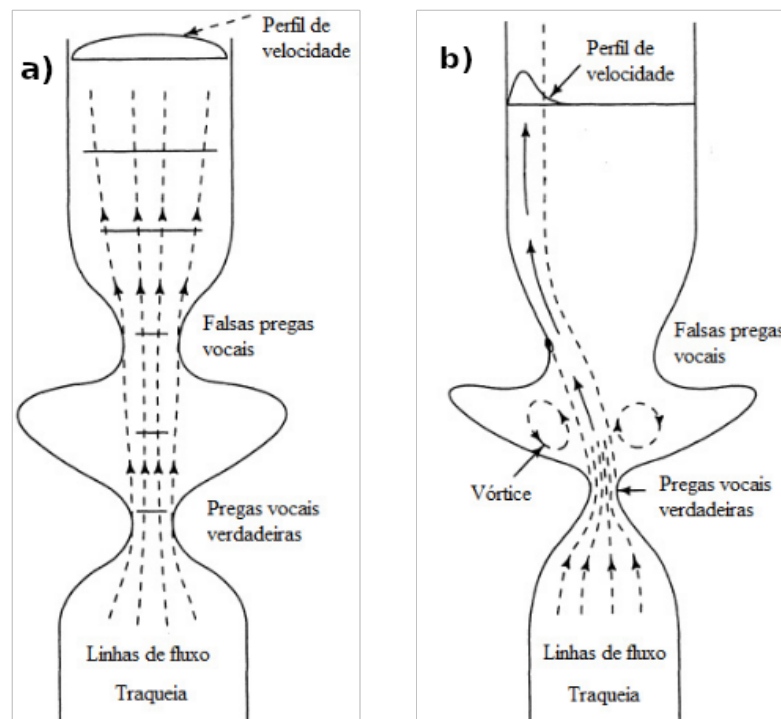
Como apresentado na Seção 2.3, a Teoria Fonte-Filtro tenta explicar o funcionamento do sistema de produção da fala, sendo a mais investigada nas diversas áreas do processamento digital de sinais de voz. Nesse caso, o modelo admitido é o de concatenação de subsistemas lineares (fonte, trato vocal e radiação dos lábios), sem perdas de acoplamento entre eles. Por exemplo, um único fonema sonoro, como uma vogal, pode ser representado como o produto de três funções de transferência, como mostrado a seguir:

$$S(z) = U(z)H(z)R(z) \quad (57)$$

em que  $U(z)$  representa a excitação da fonte,  $H(z)$  a dinâmica do filtro, e  $R(z)$  os efeitos da radiação.

Outra suposição feita na abordagem clássica é que o som da voz se propaga como uma onda plana, assumindo que, quando a prega vocal abre, um pressão sonora uniforme produzindo, se expandindo para preencher a área da seção transversal do trato vocal e se propaga de forma homogênea do trato vocal até a boca. Tal representação é apresentada na Figura 21 (a). Entretanto, estudos (TEAGER, 1980; TEAGER; TEAGER, 1983; TEAGER; TEAGER, 1990; NADA, a) sugerem que uma dinâmica não linear de fluidos (Figura 21 (b)) parece ser uma abordagem mais realística para caracterizar a propagação do som através do sistema vocal. Neste caso, os vórtices localizados nas falsas pregas vocais fornecem a excitação necessária da fonte durante a fase de fechamento das pregas vocais (DELLER; HANSEN; PROAKIS, 1999; COSTA, 2012).

Figura 21 – (a) Interpretação clássica da propagação do som através do sistema vocal. (b) Interpretação da dinâmica não linear de fluidos



Fonte: Adaptado (DELLER; HANSEN; PROAKIS, 1999) apud (COSTA, 2012)

Uma das formas de analisar sistemas não lineares é o uso da quantificação da recorrência. Esse se baseia no Teorema da Recorrência, de Poincaré (POINCARÉ, 1890), o qual certifica que, para um conjunto grande de sistemas dinâmicos, as trajetórias retornam infinitas vezes arbitrariamente próximas a quase todos os pontos iniciais, formando um conjunto infinito de instantes de retorno (ALTMANN, 2004). A análise de quantificação de recorrência pode ser empregada de duas formas: análise subjetiva, por intermédio da análise visual dos gráficos de recorrência e pela análise objetiva, utilizando-se medidas de quantificação de recorrência (COSTA, 2012; VIEIRA, 2014; BARROS *et al.*, 2020)

## 2.5 Gráficos de recorrência e suas medidas de quantificação

Métodos aprovados e já conhecidos para análise de dados se baseiam em conceitos lineares, como estimação de momentos, correlações, espectro de potência, ou análise de componente principal (TAKENS, 1981; KANTZ; SCHREIBER, 2003). Nas últimas duas décadas, os métodos da teoria não linear se mostraram bastantes promissores, ampliando os métodos analíticos. Nesse contexto, destacam-se os gráficos de recorrência (WEBBER; MARWAN, 2015).

Para que seja possível o uso da análise de recorrência, é necessário o uso de métodos preliminares para a determinação de parâmetros ótimos para a reconstrução do espaço de fase, que são a dimensão de imersão e o passo de reconstrução (MARWAN *et al.*, 2007). Além da necessidade de uma inspeção minuciosa dos parâmetros, necessitando de ferramentas de análise de dados, o seu cálculo demanda um elevado poder computacional, devido às grandes matrizes de parâmetros geradas, refletindo em um longo tempo para o processamento (RAWALD; SIPS; MARWAN, 2017a). Outrossim, o uso do método de quantificação de recorrência tem se demonstrado promissor na discriminação de patologias laríngeas (COSTA *et al.*, 2012; VIEIRA *et al.*, 2018; LOPES *et al.*, 2014; LOPES *et al.*, 2020).

Os gráficos de recorrência, assim como suas medidas de quantificação, dependem da geração do espaço de fase. Para que este seja reconstruído, parâmetros ótimos para o passo de reconstrução e dimensão de imersão devem ser definidos, utilizando-se métodos da informação mútua e falsos vizinhos próximos, respectivamente.

### 2.5.1 Espaços de fase

Alguns desses métodos são baseados na análise topológica da dinâmica de sistemas a partir do espaço de fase. O estado do sistema pode ser descrito por  $d$  variáveis de estados:

$$x_1[n], x_2[n], \dots, x_d[n], \quad (58)$$

sendo, por exemplo, duas variáveis de estado de um sistema termodinâmico, a temperatura e pressão. A variável de estado  $d$  no tempo  $n$  forma um vetor  $x[n]$  no espaço  $d$ -dimensional, o qual é chamado de espaço de fase. Esse vetor move-se no tempo e na direção que é especificada pelo vetor velocidade.

$$\vec{x}[n] = \partial_t x[n] = F[n], \quad (59)$$

sendo o campo da velocidade  $F[n]$  é tangente a trajetória. A sucessão de vetores do espaço de fase forma uma trajetória (trajetória do espaço de fase, ou órbita). Usualmente, apenas uma observação  $u[n]$  é disponível, que é uma série temporal. Como as medidas resultam numa série temporal discreta, as observações podem ser escritas como  $u_i$ , com  $n = i\Delta n$  e  $\Delta n$  representando a amostragem da medida (WEBBER; MARWAN, 2015).

Devido ao forte acoplamento entre as componentes do sistemas, estas individualmente carregam consigo informação suficiente sobre a dinâmica de todo o sistema. A observação



de um processo real normalmente não faz uso de todas as variáveis de estado, ora porque não estão disponíveis ora não foram todas medidas. Todavia, uma trajetória equivalente do espaço de fase, que preserve a estrutura topológica da trajetória do espaço de fase original, pode ser reconstruída apenas se utilizando uma observação da série temporal (TAKENS, 1981; PACKARD *et al.*, 1980). Este método é frequentemente chamado de método das versões defasadas:  $\vec{x}_i = (u_i, u_{i+\tau}, \dots, u_{i+[dim-1]\tau})^T$ , em que  $dim$  é a dimensão de imersão e  $\tau$  é o passo de reconstrução. Para que seja possível manter estruturas topológicas, é necessário que  $m \geq 2dim + 1$ , com  $dim$  sendo a dimensão do atrator (TAKENS, 1981). O espaço de fase pode ser visualizado nas Figuras 22, 23, 24 e 25, item (c).

### 2.5.1.1 Método da informação mútua para determinação do passo de reconstrução

Uma das formas de determinar o valor do passo de reconstrução  $\tau$ , o método da informação mútua, possibilita a reconstrução de vetores com o menor nível de informação redundante uns dos outros, mas, ainda sim, correlacionados. A informação mútua média entre  $x[n]$  e sua versão defasada  $x[n + \tau]$  é obtida a partir de um histograma de  $b$  intervalos de classe (bins), criado para estimar a distribuição de probabilidade dos dados do sinal  $x[n]$ . A informação mútua média entre  $x[n]$  e sua versão defasada  $x[n + \tau]$  é dado por (COSTA, 2012):

$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{i,j}(x[n], x[n + \tau]) \log_2 \left[ \frac{P_{i,j}(x[n], x[n + \tau])}{P_i(x[n])P_j(x[n + \tau])} \right], \quad (60)$$

em que  $P_i(x[n])$  é a probabilidade de que o sinal  $x[n]$  assuma um valor dentro do  $i$ -ésimo intervalo do histograma,  $P_j(x[n + \tau])$  é a probabilidade de que  $x[n + \tau]$  esteja no  $j$ -ésimo intervalo e  $P_{i,j}(x[n], x[n + \tau])$  é a probabilidade de que, simultaneamente,  $x[n]$  esteja no  $i$ -ésimo intervalo e  $x[n + \tau]$  no  $j$ -ésimo intervalo (COSTA, 2012; SAVI, 2017).

Tem-se ainda que  $I[n] \geq 0$ . Se  $x[n]$  e  $x[n + \tau]$  são iguais, então  $I[\tau]$  é máximo. Por outro lado, se  $x[n]$  e  $x[n + \tau]$  são completamente independentes, então o argumento do termo logarítmico é a unidade e  $I[\tau] = 0$ . É definido que o valor da defasagem ideal corresponde ao primeiro mínimo local, quando existir, da função de informação mútua média  $I[n\tau]$ , garantindo assim que o espaço de fase reconstruído seja topologicamente equivalente ao espaço de estados do sistema dinâmico original (COSTA, 2012; SAVI, 2017).

### 2.5.1.2 Falsos vizinhos próximos

Um dos métodos utilizados para encontrar a dimensão de imersão ótima é o método dos falsos vizinhos próximos. A ideia é que, em se diminuindo a dimensão, estar-se-á proporcionando-se um aumento do número de pontos no espaço de fase que serão projetados na vizinhança de qualquer ponto do espaço de fase, mesmo que estes não sejam realmente vizinhos. O método mais simples usa a quantidade destes falsos vizinhos próximos como a função da dimensão de imersão, com intuito para encontrar a dimensão ótima (KANTZ; SCHREIBER, 2003). Tal dimensão deve

ser tomada onde os falsos vizinhos próximos desaparecem. A Equação 61 determina o índice em que é apresentado a menor quantidade de vizinhos próximos (WEBBER; MARWAN, 2015).

$$f_{nn}(r) = \frac{\sum_{n=1}^{N-dim-1} \Theta\left(\frac{|s_n^{(dim+1)} - s_{k(n)}^{(dim+1)}|}{|s_n^{(dim)} - s_{k(n)}^{(dim)}|} - r\right) \Theta\left(\frac{\sigma}{r} - |s_n^{(dim)} - s_{k(n)}^{(dim)}|\right)}{\sum_{n=1}^{N-dim-1} \Theta\left(\frac{\sigma}{r} - |s_n^{(dim)} - s_{k(n)}^{(dim)}|\right)}, \quad (61)$$

em que  $s_{k(n)}^{(dim)}$  é o vizinho mais próximo de  $s_n$  em  $dim$  dimensões, isto é,  $k(n)$  é o índice do elemento da série temporal  $k$  diferente de  $n$  para que  $|s_n - s_k|$  seja mínimo. A primeira função degrau do numerador é unitária, se o vizinho próximo for falso, isto é, se a distância é incrementada por um fator maior que  $r$  quando a dimensão de imersão é aumentada de forma unitária; já a segunda função degrau suprime todos, desde que estes tenham distância inicial que seja maior que  $\frac{\sigma}{r}$ . Pares que possuam uma distância cujos espaços sejam maiores que  $\frac{\sigma}{r}$ , por definição não podem ser vizinhos falsos, a partir de que, na média, não existe espaço suficiente para se afastar mais que  $\sigma$  (KANTZ; SCHREIBER, 2003).

### 2.5.1.3 Gráficos de recorrência

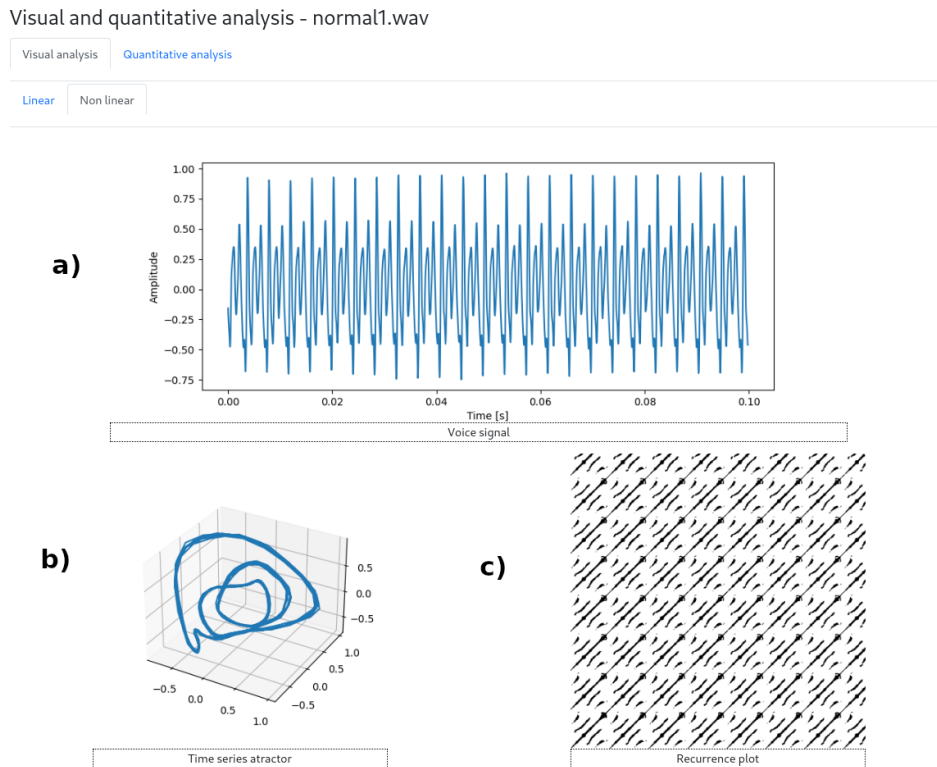
Uma tentativa de analisar as séries temporais e extrair informações relevantes dos dados baseia-se na busca por redundâncias. Caso os dados sejam aperiódicos, uma das formas é a busca por eventos repetitivos, viabilizando a construção de padrões. Esses eventos repetitivos podem ser chamados de recorrência, a qual pode ser definida como o retorno de uma trajetória dentro do espaço de estado, em sua região de vizinhança. No caso dos sinais produzidos pelo sistema vocal, é possível observar que o atrator caótico retorna de forma aleatória dentro de uma pequena vizinhança de qualquer um de seus pontos. Caso isso não ocorra, o sistema pode ser transiente (ECKMANN; KAMPHORST; RUELLE, 1987; KANTZ; SCHREIBER, 2003). Uma forma simples para visualizar recorrências é o gráfico de recorrência, que é gerado a partir da matriz definida em:

$$M_{ij} = \Theta(\varepsilon - |S_i - S_j|), \quad (62)$$

em que  $\Theta(\cdot)$  é a função de Heaviside,  $\varepsilon$  é o raio de vizinhança que deve ser selecionado e  $S_i$  são os vetores defasados com uma determinada dimensão de imersão. A matriz gerada é simétrica. Se a trajetória do espaço reconstruído retornar no tempo  $i$  na vizinhança  $\varepsilon$  onde já esteve em determinado instante  $j$ , então  $M_{ij} = 1$ , senão  $M_{ij} = 0$ , sendo o resultado positivo pintado de preto, e negativo de branco, formando assim o gráfico de recorrência (KANTZ; SCHREIBER, 2003; COSTA, 2012; VIEIRA, 2014). Dependendo de sua natureza, o sinal apresentará formas distintas: em vozes saudáveis, os sinais apresentarão estruturas diagonais bem definidas; em caso de patologias, como a paralisia, grandes estruturas recorrentes nas diagonais e nas linhas horizontais e verticais. Para o Edema de Reinke, são apresentadas pequenas estruturas horizontais e/ou verticais ao longo das diagonais e pequenas linhas diagonais espalhadas pelo

gráfico (VIEIRA *et al.*, 2018). Os gráficos de recorrência bem como o espaço de fase e o sinal de voz de sinais de voz saudável, com edema, paralisia e com nódulo são apresentados, respectivamente, nas Figuras 22, 23, 24 e 25

Figura 22 – Apresentação da visualização pela análise não linear de um sinal saudável ou normal: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros:  $\dim=3$ ,  $\tau = 7$ ,  $\varepsilon = 0,139176$



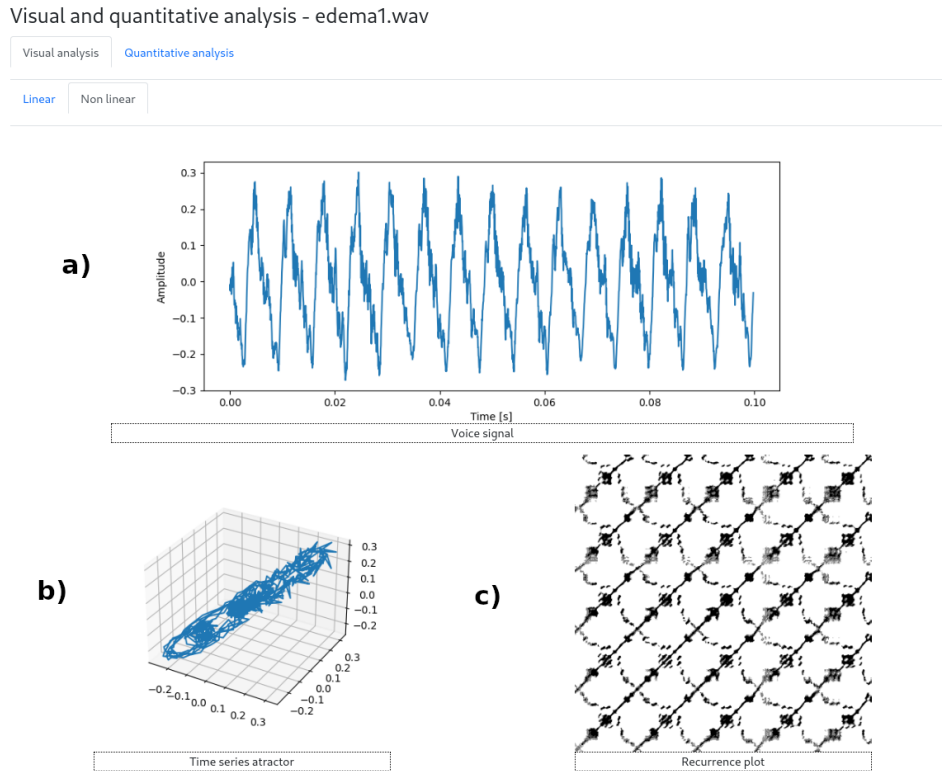
Fonte: Autor.

#### 2.5.1.4 Medidas de quantificação de recorrência

Os gráficos de recorrência indicam os comportamentos do sistema. Entretanto, caso o gráfico não seja gerado em uma resolução adequada, este poderá trazer um nível de subjetividade sobre os padrões associados ao sistema. Uma das formas de superar esta restrição é uso de medidas de quantificação de recorrência, definidos por Zbilut, Weber e Marwan (ZBILUT; WEBBER, 1992; WEBBER; ZBILUT, 1994; Marwan, 2003; ZBILUT; WEBBER, 2007), no início da década de 90. Existem diversas medidas para se quantificar a recorrência, e dentre elas Vieira (2014) apontou que o determinismo (DET), comprimento das linhas diagonais ( $L_{max}$ ), tendência (TREND), laminaridade (LAM), tempo de permanência (TT) e comprimento máximo da estrutura vertical ( $V_{max}$ ) como medidas robustas para a classificação de sinais de voz com e sem patologias laríngeas, como edema de Reinke, nódulo e paralisia.

A taxa de recorrência (REC) propicia a contagem dos pontos pretos excluindo a Linha de Identidade (LDI). Esta é uma medida da densidade relativa dos pontos de recorrência na matriz esparsa e está relacionada a definição da soma da correlação (GRASSBERGER; PROCACCIA,

Figura 23 – Apresentação da visualização pela análise não linear de um sinal com edema: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: dim=3,  $\tau = 5$ ,  $\varepsilon = 0,151222$



Fonte: Autor.

1983). Para que seja possível o uso da REC para a determinação da soma da correlação são necessários grandes segmentos de dados.

$$REC(\varepsilon, N) = \frac{1}{N^2 - N} \sum_{i \neq j=1}^N R_{i,j}^{m,\varepsilon}. \quad (63)$$

No limite de uma série temporal longa

$$P = \lim_{N \rightarrow \infty} REC(\varepsilon, N), \quad (64)$$

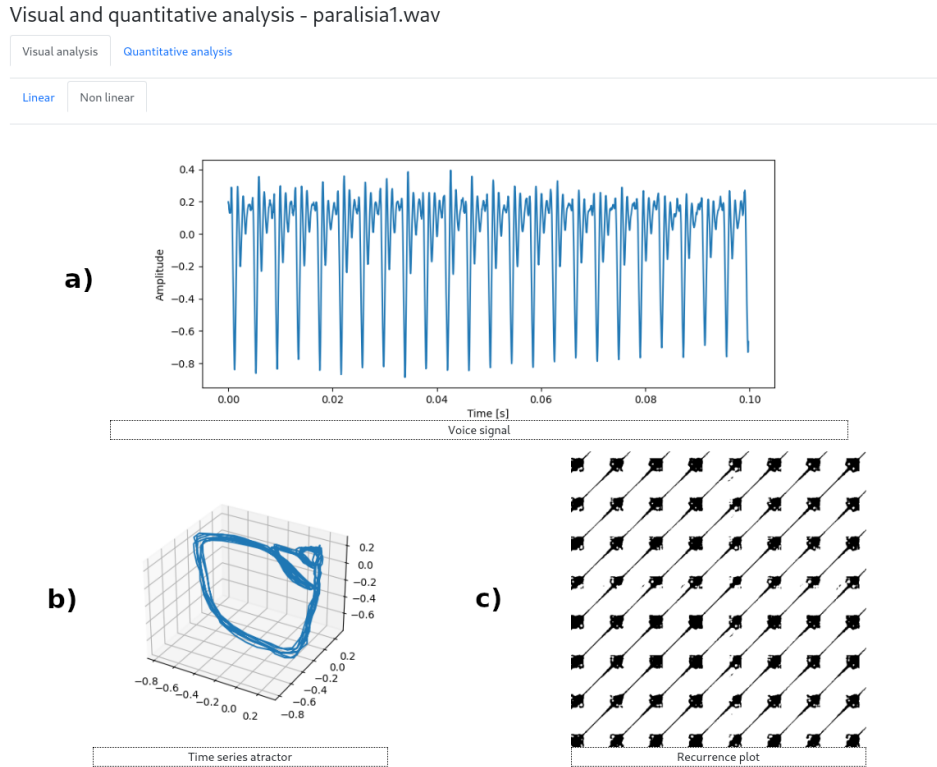
é a probabilidade de achar o ponto de recorrência dentro do gráfico de recorrência, bem como a probabilidade que o estado recorra. Sabendo que a probabilidade  $\rho(x)$  de um estado  $x$  de um processo estocástico e tendo a dimensão  $m = 1$  e utilizando a normal máxima, a REC pode ser computada analiticamente utilizando a convolução (THIEL; ROMANO; KURTHS, 2003):

$$P_0 = \rho(x) * \rho(x). \quad (65)$$

A probabilidade  $P_0$  pode ser utilizada para descrever analiticamente as medidas de quantificação de recorrência para alguns sistemas (THIEL *et al.*, 2002; THIEL; ROMANO; KURTHS, 2003).

As medidas a seguir são baseadas nas estruturas em forma de linha de um gráfico de recorrência. Primeiramente, é considerado o histograma  $H_D(l)$  dos comprimentos das estruturas

Figura 24 – Apresentação da visualização pela análise não linear de um sinal com paralisia: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros: dim=4,  $\tau = 8$ ,  $\varepsilon = 0,137660$



Fonte: Autor.

diagonais de um gráfico de recorrência.

$$H_D(l) = \sum_{i,j=1}^N (1 - R_{i-1,j-1})(1 - R_{i+1,j+1}) \prod_{k=0}^{l-1} (R_{i+k,j+k}) \dots \quad (66)$$

A porcentagem de determinismo (DET) é definida como a fração dos pontos de recorrência que formam linhas diagonais:

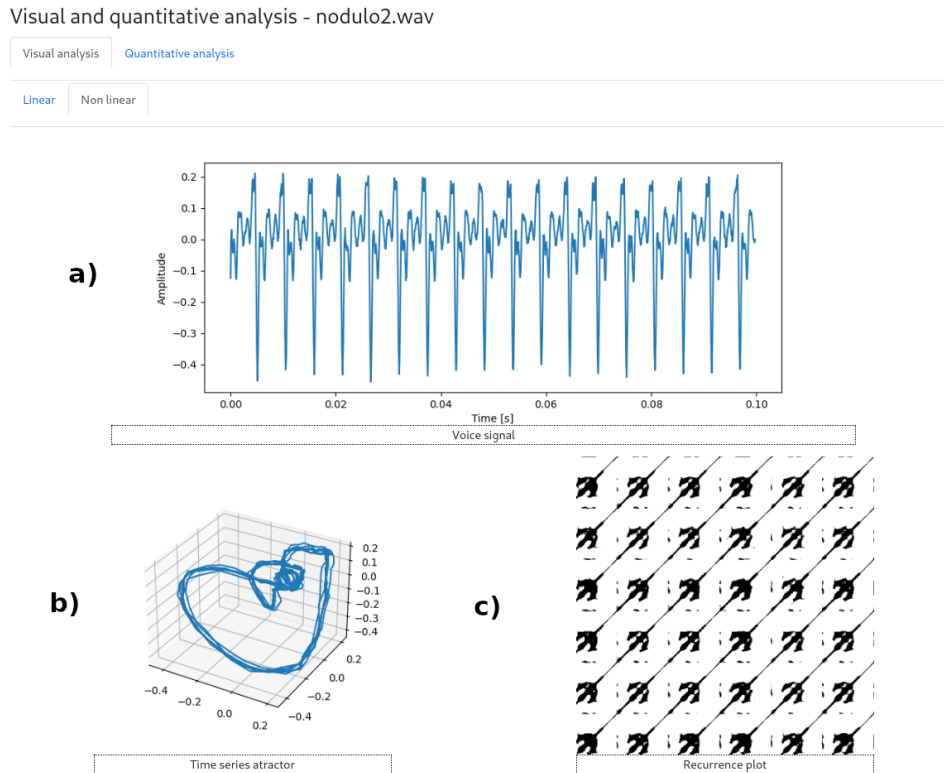
$$DET = \frac{\sum_{l=d_{min}}^N l H_D(l)}{\sum_{i,j=1}^N R_{i,j}} \dots \quad (67)$$

Sistemas que possuem uma dinâmica determinística são caracterizados por linhas diagonais indicando a repetição de recorrências dentro de um determinado estado. Para sinais periódicos as linhas diagonais são longas. Já para sinais caóticos as linhas diagonais são curtas. No caso de sinais estocásticos as linhas diagonais são ausentes, salvo em casos de recorrência que formam linhas muito curtas. O determinismo pode ser interpretado como a predicabilidade de um sistema ter um comportamento mais periódico do que caótico (WEBBER; MARWAN, 2015).

Outra medida é o comprimento máximo das linhas diagonais ( $L_{max}$ )

$$L_{max} = \arg \max_l H_D(l) \quad (68)$$

Figura 25 – Apresentação da visualização pela análise não linear de um sinal com nódulo: sinal de voz (a), atrator(b) e gráfico de recorrência (c). Parâmetros:  $\text{dim}=3$ ,  $\tau = 5$ ,  $\varepsilon = 0,143660$



Fonte: Autor.

que simplesmente indica o comprimento da diagonal única mais longa dentro de todo o gráfico de recorrência. Levando em consideração que as estruturas diagonais demonstram o intervalo em que um segmento da trajetória está próximo de outro segmento da trajetória em outro instante de tempo, essas linhas representam a divergência dos segmentos da trajetória. Quanto menor o  $L_{max}$  mais as trajetórias serão divergentes (WEBBER; MARWAN, 2015).

As medidas REC, DET e  $L_{max}$  podem ser computadas separadamente para diagonal paralela à LDI com a distância  $k$  até a LDI. Tomando por exemplo a densidade de pontos de recorrência ao longo de uma diagonal com a distância  $k$  da LDI dada por

$$REC_k = \frac{1}{N - K} \sum_{k,j=1}^{N-k} R_{k,j}. \quad (69)$$

veem-se tais medidas relacionadas a diagonais com um índice subscrito ou, no geral, com um asterisco, i.e.  $REC_*$  ou  $REC_i$ . As medidas de quantificação de recorrência baseadas em diagonais são importantes no estudo da periodicidade do sinal (ZBILUT; MARWAN, 2008), para indicar órbitas periódicas (MINDLIN; GILMORE, 1992; LATHROP; KOSTELICH, 1989; GILMORE, 1998) ou para investigar a inter-relação entre sistemas complexos (MARWAN; THIEL; NOWACZYK, 2002). Além disso,  $REC_k$  pode ser interpretado como a probabilidade que um sistema ocorra após  $k$  passos (WEBBER; MARWAN, 2015). A medida de tendência (TREND) demonstra o coeficiente de regressão linear sobre a densidade dos pontos de recorrência

$REC_*$  das diagonais paralelas à LDI, Equação 69, como função da distância temporal entre essas diagonais e a LDI:

$$TREND = \frac{\sum_{i=1}^{\tilde{N}} (i - \frac{\tilde{N}}{2})(REC_i - \langle REC_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \frac{\tilde{N}}{2})^2} \quad (70)$$

As medidas computadas relacionadas aos elementos diagonais com um índice subscrito ou, no geral, com um asterisco, são bastante úteis no estudo da periodicidade de um sinal, ou para indicar órbitas periódicas ou a inter-relação entre sistemas complexos. A tendência dá a informação sobre a relação entre estacionariedade e não estacionariedade no processo. Dinâmicas quasi-estacionárias tendem a ter valores de TREND próximos de 0. Dinâmicas não estacionárias possuem um valor longe de 0, revelando a deriva na dinâmica e possivelmente indicando que o sistema está em rota entre estados mais estacionários (WEBBER; MARWAN, 2015). Caso a tendência seja calculada a partir de medidas de quantificação de recorrência com base no tempo, com janelas deslizantes, o valor ficará altamente dependente do tamanho dessas janelas e poderá revelar resultados contraditórios para seus diversos tamanhos (MARWAN, 2011).

Os gráficos de recorrência não possuem apenas linhas diagonais mas também verticais e horizontais (MARWAN *et al.*, 2002; WEBBER; MARWAN, 2015). Neste contexto, as medidas de quantificação de recorrência de laminaridade (LAM), tempo de permanência (TT) e comprimento máximo das linhas verticais ( $V_{max}$ ) são apresentadas. A laminaridade pode ser descrita por:

$$LAM = \frac{\sum_{l=v_{min}}^N lH_V(l)}{\sum_{i,j=1}^N R_{i,j}} \quad (71)$$

com

$$H_V(l) = \sum_{i,j=1}^N (1 - R_{i,j-1})(1 - R_{i,j+1}) \prod_{k=0}^{l-1} R_{i,j+k} \quad (72)$$

definindo o histograma de comprimento das linhas verticais. A laminaridade carrega a definição análoga à definição do determinismo, ou seja, a laminaridade reporta a porcentagem dos pontos de recorrência nas estruturas verticais onde o DET define a porcentagem dos pontos de recorrência nas estruturas diagonais. A computação da laminaridade é realizada para aquelas estruturas cujo comprimento  $l$  exceda o comprimento mínimo  $v_{min}$  para que seja reduzida a influência dos pontos espúrios.

A definição de comprimento médio das estruturas verticais, chamada de tempo de permanência (TT), é representado por:

$$TT = \frac{\sum_{l=v_{min}}^N vH_V(l)}{\sum_{l=v_{min}}^N H_V^e(l)} \quad (73)$$

Essa equação utiliza o comprimento mínimo  $v_{min}$ , como no caso da laminaridade. O tempo de permanência contém informação sobre a quantidade e o tamanho das estruturas verticais no gráfico de recorrência, reportando o tempo médio que o sistema permanecerá num estado específico, ou seja, quanto o estado ficará aprisionado.

Por fim, o comprimento máximo das estruturas verticais  $V_{max}$  mede a linha vertical mais longa no gráfico de recorrência:

$$V_{max} = \arg \max_l H_V(l) \quad (74)$$

esta é análoga ao comprimento da diagonal  $L_{max}$  e está relacionada ao estado singular em que o sistema fica travado, mantendo um padrão inscrito em retângulos no gráfico de recorrência (WEBBER; MARWAN, 2015).

Segundo Lopes *et al.* (2020), podem-se organizar essas medidas em grupos. Assim estas podem se agrupar conforme estejam relacionados à formação de: i) pontos de recorrência; ii) diagonais principais, iii) estruturas verticais. Tais grupos estão detalhados na Tabela 1 a seguir.

Tabela 1 – Medidas relacionadas à formação de pontos de recorrência, linhas diagonais e verticais

Categoria	Medida	Descrição
Pontos de recorrência	TREND	Coefficiente de regressão linear. Provê informação da não estacionariedade da série temporal.
	DET	Razão entre o número de pontos de recorrência que formam as estruturas diagonais de todos os pontos de recorrência
	TRANS	Taxa local de recorrência
Estruturas horizontais	$L_{med}$	Tempo médio em que dois segmentos de uma trajetória evoluem em via similar ou paralela.
	$L_{max}$	Tempo máximo que dois segmentos permanecem com uma evolução similar no decorrer do tempo.
	ENTR	É uma medida escalar que representa a probabilidade da distribuição dos tamanhos das linhas diagonais
	DIV	Tempo em que a trajetória de dois segmentos divergem no espaço de fase
Estruturas verticais	TT	É o tempo gasto em um determinado estado.
	$V_{max}$	Tempo máximo que um estado permanece laminar (sem turbulência).
	LAM	Razão entre os pontos de recorrência que forma estruturas verticais e o conjunto de pontos de recorrência do gráfico
	RPDE	Grau no qual a série repete a mesma sequência no tempo

Adaptado (LOPES *et al.*, 2020)

## 2.6 Da arquitetura de *software* e infraestrutura computacional

### 2.6.1 Arquitetura de microsserviços

A implementação de microsserviços é um estilo de arquitetura que coloca a ênfase na divisão do sistema em serviços pequenos e leves, com o propósito de realizar um função do negócio de forma coerente, sendo uma evolução da tradicional arquitetura orientada a serviço (WASEEM; LIANG; SHAHIN, 2020). Essa arquitetura possibilita que um domínio da aplicação seja colocado em funcionamento, escalado e testado de forma independente, permitindo que cada microsserviço assuma uma única responsabilidade (THONES, 2015).

Com a quebra dos domínios da aplicação em projetos menores, a adição ou alteração das funcionalidades é realizada de forma mais rápida e eficiente, possibilitando a realização do refatoramento de forma mais usual, reduzindo-se, portanto, os tempos de geração de novas



versões. Além disso, caso a equipe de desenvolvimento deseje realizar a troca de uma linguagem de programação ou um *framework* de desenvolvimento, é possível a rápida prototipação da funcionalidade na mesma API do microsserviço (FRASER; ZIADE, 2021).

Sendo assim, cada microsserviço é desenvolvido por uma equipe, e essa parte do domínio é um projeto em si, podendo ter sua própria linguagem de programação e banco de dados, desde que possua uma interface de comunicação bem documentada (NEWMAN, 2022). Desta forma, a plataforma que dá suporte à arquitetura é bastante agnóstica, podendo ser implantada em servidores dedicados, em uma nuvem pública ou na forma de tecnologias de virtualização baseada em contêineres, como Docker e Kubernetes. Essas tecnologias facilitam a operação dos microsserviços e os proveem de funcionalidades de atualizações em esteira, escalabilidade automatizada e rebalanceamento se o nó associado falhar. (HEINRICH *et al.*, 2017; BUELTA, 2019).

### 2.6.2 Computação em nuvem e Infrastructure as Code (IaC)

Segundo Hurwitz e Kirsch (2020), o termo Computação em Nuvem é um método de prover recursos computacionais compartilhados, incluindo aplicações, tempo de computação, armazenamento, desenvolvimento e plataformas de implantação bem como os processos de negócios. Esses métodos permitem que o uso de recursos sejam mais fáceis propiciando a padronização e automação.

A padronização permite que os serviços possam ser implementados por intermédio de interfaces consistentes, facilitando sua automatização, processo este que depende das regras de negócio, recursos disponíveis e demandas de segurança. Neste sentido, a automação permite que, não sendo o serviço mais necessário, os recursos fiquem disponíveis para outras soluções. Este *pool* de recursos computacionais pode ser classificado segundo o seu compartilhamento, podendo ser público ou privado (COMER, 2021; HURWITZ; KIRSCH, 2020).

Na modalidade pública, o conjunto de *hardware*, rede, armazenamento, serviços, aplicações e interfaces é operada por um terceiro e de sua propriedade, para uso de outras companhias ou pessoas físicas. A partir desse conjunto, são criados *data centers* altamente escaláveis, os quais escondem os detalhes da infraestrutura subadjacente do consumidor. Os *data centers* disponibilizam interfaces de configuração, possibilitando o uso de aplicações de configuração de infraestrutura. Alguns exemplos de provedores de *clouds* públicas são Amazon Web Services (AWS), Google Cloud Platform (GCP) e Microsoft Azure. Já na modalidade privada, os recursos computacionais são alocados única e exclusivamente para o uso do ente, parceiros autorizados e clientes, tendo meios de acessos controlados e restrita ao uso do público (HABER; CHAPPELL; HILLS, 2022; HURWITZ; KIRSCH, 2020).

Com o intuito de automatizar as atividades de criação de recursos computacionais na infraestrutura pública ou privada, o conceito de Infraestrutura como Código, do inglês *Infrastructure as Code*, foi desenvolvido. Essa prática permite que os times reduzam o tempo e erros na realiza-

ção dos procedimentos de implantação, permitindo também a integração com fluxo de entrega e integração contínua. Existem diferentes tipos de linguagens e ferramentas para a implantação de IaC, podendo ser scripts, estes muito utilizados em provedores de *cloud*, declarativos, onde é descrito o estado desejado para o sistema ou infraestrutura (KRIEF, 2019).

Para a infraestrutura de *cloud*, a IaC é dividida em topologias de implantação e provisionamento de infraestrutura, configuração de servidores e padronização, containerização e configuração e implantação do Kubernetes. Para configuração de servidores, ferramentas como Ansible, Chef e Puppet são utilizadas, permitindo a criação de imagens que contêm todas as configurações necessárias dos servidores. A containerização, viabilizada pelo Docker, consiste na implantação de aplicações em contêineres, cuja configuração não é alterada durante a sua execução, ao invés de máquinas virtuais. No escopo de configuração e implantação, o Kubernetes é considerado um orquestrador, que implanta contêineres, arquitetura de rede e um gerenciador de volume de dados, tudo sendo representado num arquivo YAML (KRIEF, 2019).

### 2.6.2.1 Docker

Docker é uma ferramenta de código aberto popular em modelos de serviço baseado em *Cloud*, que provê orquestração em *Cloud* em um nível de infraestrutura bem como virtualização em nível de plataforma. Docker é uma tecnologia leve, que automatiza a implantação de aplicações em contêineres (NAIK, 2016).

Desenvolvedores podem construir, embarcar e distribuir aplicações variadas em qualquer lugar. Essas aplicações são executadas com a ajuda de pacotes de *softwares* chamados contêineres, os quais são criados a partir de uma imagem básica. A imagem consiste na combinação dessa imagem básica, proveniente de outros *softwares*, bibliotecas ou arquivos de configuração, possibilitando-se a criação de aplicações individualizadas. A maneira mais fácil de se criar uma imagem é escrever um arquivo chamado *Dockerfile*, o qual é composto por vários comandos, possibilitando assim a criação de uma nova imagem (YADAV; PAL; YADAV, 2021).

## 2.7 Trabalhos Relacionados

As aplicações em *e-Health* são inúmeras, especialmente no escopo da fonoaudiologia. Para realizar uma revisão no estado da arte, foram consultados alguns portais indexadores, como o *IEEEExplore*, *ScienceDirect*, *PubMed* e *SciELO*, utilizando-se a combinação de palavras, como *speech*, *voice*, *disorder*, *vocal*, *website*, *telehealth*, *telepractice*, *telessaúde*, *sistema*, *aplicação* e *fonoaudiologia*, *desordem vocal* e *patologia*. Além disso, trabalhos de dissertação e tese foram utilizados nesta revisão.

Doarn *et al.* (2019) descrevem o projeto e implementação de portal baseado na web desenvolvido para prover a prática a domicílio entre sessões semanais de terapia vocal realizada por intermédio de telessaúde para crianças com desordens vocais. A prática foi monitorada

por um patologista de fala/linguagem. O portal foi desenvolvido como uma plataforma na qual os participantes podiam realizar práticas completas em domicílio, acompanhados por um time interdisciplinar de especialistas em terapia vocal pediátrica, em telessaúde, engenheiros de *software* voltados para biomedicina e *designers* de interface. Esse sistema foi testado com um grupo de crianças com distúrbios vocais, participantes de um estudo mais abrangente em telessaúde. O estudo demonstrou que a prática em domicílio para terapia vocal em crianças é bem sucedida. O site final apresenta 11 questões voltadas para higiene vocal, exercícios tradicionais de terapia vocal agrupados em níveis, demonstrações audiovisuais de terapia vocal, um sistema de armazenamento e de recuperação de sinais de voz, uma funcionalidade de *chat*, guias escritos para os exercícios de terapia semanais, e questionários para os pais responderem após a realização de cada sessão de terapia.

No estudo piloto de Hill *et al.* (2006), exploram-se a viabilidade e a efetividade de uma aplicação de telerreabilitação, pela internet, para avaliação de distúrbios motores a nível de fala, em adultos com dano neurológico adquirido. A partir de um modelo de pesquisa que utiliza medidas contrabalanceadas e repetidas, dois patologistas de fala-linguagem avaliaram 19 falantes com disartria, aplicando uma bateria de avaliações perceptivas. As avaliações incluíam uma versão de 19 itens da Avaliação de Disartria Frenchay (ADF) (P. Enderby, 1983), a Avaliação de Inteligibilidade de Fala com Disartria (AIFD) (YORKSTON; BEUKELMAN, 1981), análise perceptivas de sinais de fala e uma taxa média da severidade da disartria. Uma das avaliações foi conduzida na forma presencial tradicional, enquanto as demais foram conduzidas de forma online, por intermédio da aplicação customizada de telerreabilitação. A aplicação permite a videoconferência a uma taxa de 128Kb/s e com a transferência na modalidade armazenar e transferir do áudio e vídeo entre os locais do paciente e o patologista da fala-linguagem. O estudo conseguiu como resultados que as medidas de severidade da disartria, percentual de inteligibilidade das sentenças e a grande parte das taxas perceptivas encontradas residiam dentro do critério clínico de aceitação. Entretanto, algumas das classificações no FDA não se comparavam entre os ambientes, onde foram exploradas as explicações sobre os resultados. Como conclusão, foi verificado que a avaliação de distúrbios motores vocais utilizando-se a telerreabilitação por intermédio da Internet é possível.

Lopes, Vieira e Behlau (2022), desenvolveram um estudo que compara e combina diferentes características acústicas na discriminação de sujeitos com e sem distúrbios vocais. Foi criado um banco de dados de 484 participantes da pesquisa. Todos os indivíduos gravaram a vogal sustentada /ε/ e realizaram um exame laringoscópico da laringe. Dos resultados do exame laríngeo realizado por um médico e o julgamento perceptivo-auditivo realizado por um fonoaudiólogo, os indivíduos foram alocados no grupo com (n=52) e sem (m=432) distúrbios vocais. Quatro tipos de características vocais foram utilizadas: medidas tradicionais, medidas cepstrais, medidas não lineares e medidas de quantificação de recorrência. Gravações compreenderam a emissão da vogal /ε/. O classificador utilizado foi a análise quadrática discriminante. Características individuais no contexto de medidas tradicionais, cepstrais e de recorrência alcançaram uma performance

aceitável  $\geq 70\%$ . A combinação de medidas melhorou a performance do classificador, com um resultado  $\geq 86.43\%$  de acurácia obtida na combinação de medidas lineares tradicionais e medidas de quantificação de recorrência. Resultados demonstram que medidas tradicionais, cepstrais e de quantificação de recorrência são características promissoras que capturam informação relevante da produção de voz, provendo boa taxa de classificação. Os achados encontrados no estudo podem ser usados para o desenvolvimento de um ferramenta computacional para monitoramento e diagnóstico de desordens vocais.

Na Tese de Costa (2008), foi realizada a discriminação de vozes com patologias laríngeas, dentre elas o edema. Para tanto, foi utilizada a análise acústica de sinais de voz, utilizando-se medidas lineares como os coeficientes de predição linear, análise cepstral e mel-cepstral. Foi utilizada uma base de dados proveniente da *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab*. Dessa base de dados foram extraídas 44 vozes com edemas nas pregas vocais, 23 vozes de outras patologias (cistos, nódulos e paralisia) e 53 vozes normais. Para a realização da classificação, foram utilizados os Modelos Ocultos de Markov, do inglês *Hidden Markov Models*. Os resultados obtidos mostram que, para a discriminação entre vozes com edema e normais, os coeficientes de predição linear obtiveram uma correta aceitação de 100%. Para a classificação, entre outras patologias (paralisia, nódulos e cistos) e vozes normais, obteve-se uma taxa de correta aceitação de 96%.

Utilizando medidas não lineares, Costa (2012) desenvolveu no trabalho de tese a caracterização e classificação de sinais de vozes saudáveis e com patologias laríngeas (edema, paralisia e nódulos), utilizando análise dinâmica não linear (e teoria do caos) e também as medidas de quantificação de recorrência. Foi utilizada na classificação a técnica de análise discriminante com as funções linear e quadrática, com validação cruzada. Em outro estudo, o mesmo autor explorou uma base de dados proveniente da *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab*, extraíndo 53 sinais de vozes saudáveis e 118 vozes afetadas por patologias laríngeas (45 com edema, 55 com paralisia e 18 com nódulos). Para que fosse possível verificar os resultados, foram criados dois conjuntos: um deles contendo uma combinação de medidas de análise não linear (MNL) e outros com medidas de quantificação de recorrência (MQR), obtendo-se os seguintes percentuais: 95,44% e 100% entre vozes saudáveis e patológicas, 94,75% e 100% entre saudáveis e edema.

Em sua Dissertação de Mestrado, Vieira (2014) realizou duas abordagens, uma com análise de sinais de vozes de locutores afetados com patologias laríngeas e outra analisando vozes infantis diagnosticadas de acordo com o grau de intensidade de desvio vocal. Para a primeira abordagem, foi utilizada uma base de dados proveniente da *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab*, sendo que foram extraídas 53 sinais de voz de locutores com laringes saudáveis e 118 sinais de voz de locutores afetados por patologias laríngeas, sendo 55 com paralisia, 45 com edema de Reinke e 18 com nódulos. Na segunda abordagem, a base utilizada foi fornecida pelo Laboratório de Voz e Deglutição do Departamento de Fonoaudiologia

da Universidade Federal da Paraíba, com a avaliação de 93 sinais de voz de crianças entre 3 e 10 anos de idade, tendo 10 como normais, 70 com desvio leve e 13 com desvio moderado. Em ambas as abordagens, foram utilizadas as Medidas de Quantificação de Recorrência, podendo citar determinismo, comprimento máximo das linhas diagonais, laminaridade e tempo de permanência. Para a classificação dos sinais, foram empregados a análise discriminante (linear e quadrática) e redes neurais MLP. Como resultados, para a abordagem um, utilizando a rede neural MLP e com segmentação, a discriminação entre saudável e patológica obteve uma acurácia, para o comprimento máximo das linhas diagonais, de 94,19%. No cenário dois, utilizando a rede neural MLP e com segmentação, a discriminação entre vozes normais e com desvio moderado foi 84,04%.

Durante a revisão do estado da arte, não foram encontradas soluções *e-Health* voltadas para discriminação de desordens vocais que contemplassem o acompanhamento de pacientes, controle de usuários, geração de relatórios, utilizando arquitetura de microserviços, Computação em Nuvem e análise não linear, com enfoque em medidas de quantificação de recorrência, como proposto neste trabalho.

## 3 ATHENA ACOUSTIC ANALYSER

### 3.1 Aplicabilidade

Este trabalho propõe a criação de uma solução, denominada *e-Health Athena Acoustic Analyser* que auxilie o profissional na tomada da decisão no diagnóstico de distúrbios vocais utilizando medidas tradicionais baseadas na análise acústica, na análise preditiva linear e em medidas de quantificação de recorrência, por intermédio da Internet. Para que o objetivo da proposta fosse alcançado, foram desenvolvidas algumas funcionalidades, dialogando com sua função social:

- Criação de uma infraestrutura computacional baseada em Computação em Nuvem;
- Criação de uma interface voltada para WEB para a interação do usuário e a solução;
- Criação de módulos de aquisição e análise dos sinais de voz;
- Implementação de lógica de processamento de sinais e persistência da informação;
- Implementação de lógica para extração de medidas quantitativas lineares e de recorrência;
- Geração de gráficos baseados nas medidas mencionadas e
- Validação do uso por especialistas de quantificação de recorrência e suas medidas associadas.

A implementação dessas funcionalidades permite realizar o processo de avaliação diagnóstica e suporte à tomada de decisão, possibilitando o relacionamento das informações com os atores envolvidos no processo. Outro ponto é o seu uso por intermédio da Internet, sendo hospedado em um ambiente de computação em nuvem, permitindo a guarda das informações de maneira segura, com redundância e de forma reutilizável em quaisquer terminais conectados à rede mundial de computadores. Além disso, é possível escalar a aplicação dependendo da demanda de uso/atendimento, sem comprometer a sua disponibilidade. Uma informação relevante é que, caso sejam desenvolvidas novas funcionalidades ou incorporação de novos microserviços, estes ficam disponíveis para os usuários, sem a necessidade de *download* e instalação das atualizações em uma máquina local.

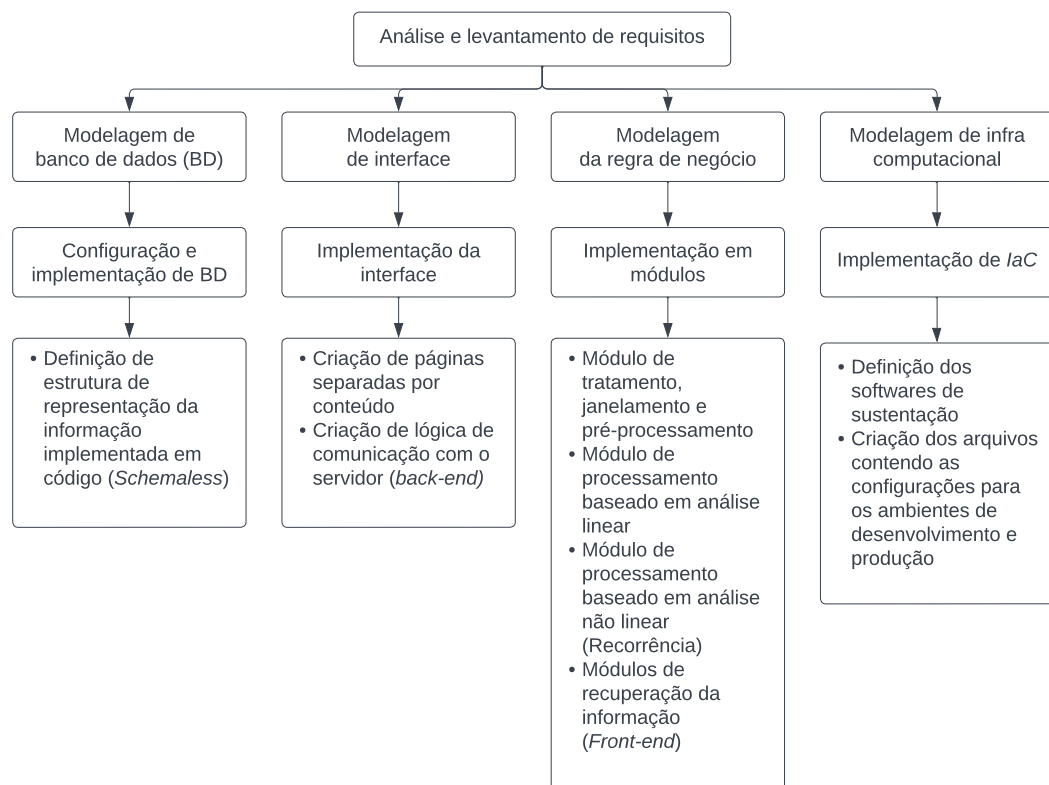
### 3.2 Metodologia

Nesta seção, são apresentadas as ferramentas necessárias para a consecução da solução Athena Acoustic Analyser. Foi realizada uma análise e levantamento de requisitos, processos a

partir dos quais se descreveram macro componentes integrantes da solução, vide Figura 26, bem como as tecnologias utilizadas por estas, como *Front-end*, *Back-end*, Sistema de Gerenciamento de Banco de Dados (SGBD), infraestrutura computacional, comunicação entre serviços e implementação de métodos para discriminação de desordens vocais, dispostos na Figura 27. Em Vieira (2014), foram apontadas as medidas com melhor desempenho na caracterização das patologias laríngeas estudadas neste trabalho (COSTA, 2008; COSTA, 2012; VIEIRA, 2014). Em relação à seção da framework de desenvolvimento, este foi selecionado pela melhor integração com as ferramentas de computação científicas, como NumPy e Scipy.

Além disso, também são descritos os métodos de análise linear (no domínio do tempo e da frequência) e não linear (gráficos de recorrência e medidas associadas) bem como a distribuição probabilística (*BoxPlots*) para as séries temporais utilizadas na solução, isto é, sinais de voz.

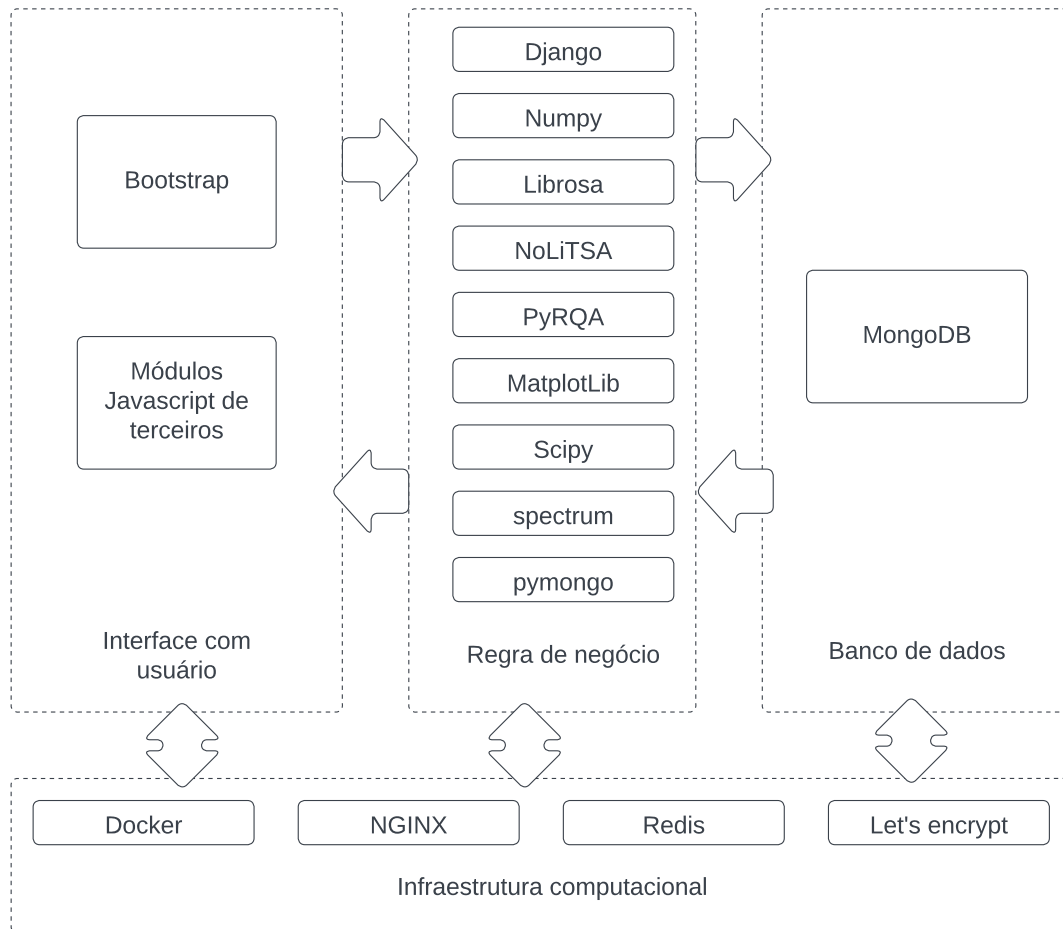
Figura 26 – Fluxo da análise e levantamento de requisitos



Fonte: Autor.

Para a criação da solução, foi utilizada para a *interface* com o usuário a biblioteca Twitter Bootstrap (SPURLOCK, 2013); para a implementação da regra de negócio o *framework* Django (NETGURU, 2022; DJANGOBOOK, 2022; TUTORIALSPPOINT, 2022); para o SGBD o MongoDB (MONGODB, 2022; MAKRIS *et al.*, 2020, 2020; BSONSPEC, 2022). Para a configuração e implementação de uma infraestrutura mínima para a execução da aplicação, fez-se uso da tecnologia de containerização Docker (NAIK, 2016; YADAV; PAL; YADAV, 2021). A

Figura 27 – Pilha tecnológica utilizada



Fonte: Autor.

implementação computacional dos gráficos de recorrência e suas medidas de quantificação foram realizadas utilizando-se a biblioteca PyRQA (RAWALD; SIPS; MARWAN, 2017b; RAWALD, )

A solução é composta por três grandes macro processos: processamentos de sinais, análise linear e não linear e persistência e disponibilidade da informação. O fluxo é iniciado na aquisição do sinal de voz a ser analisado pelo sistema, sinal este que precisa ter uma taxa de amostragem mínima de 22050 Hz e 16 bits por amostra. O sinal foi utilizado sem aplicação de filtros em frequência, e aplicada a pré-ênfase em alguns métodos de computação de características. Passando-se para etapa de detecção de atividade vocal com seus segmentos resultantes, foi aplicado o janelamento, aplicando-se a janela retangular com 32 ms de comprimento, sem sobreposição. Cada segmento janelado é enviado para o macro processo de análise, compreendendo as medidas baseadas nas análise linear e não linear. No processo de análise não linear, são obtidos os parâmetros ótimos do passo de reconstrução  $\tau$  e a dimensão de imersão  $dim$ . De posse dos parâmetros e do segmento janelado do sinal de voz, são gerados o espaço de fase e o gráfico de recorrência, possibilitando a extração das medidas de quantificação de recorrência.

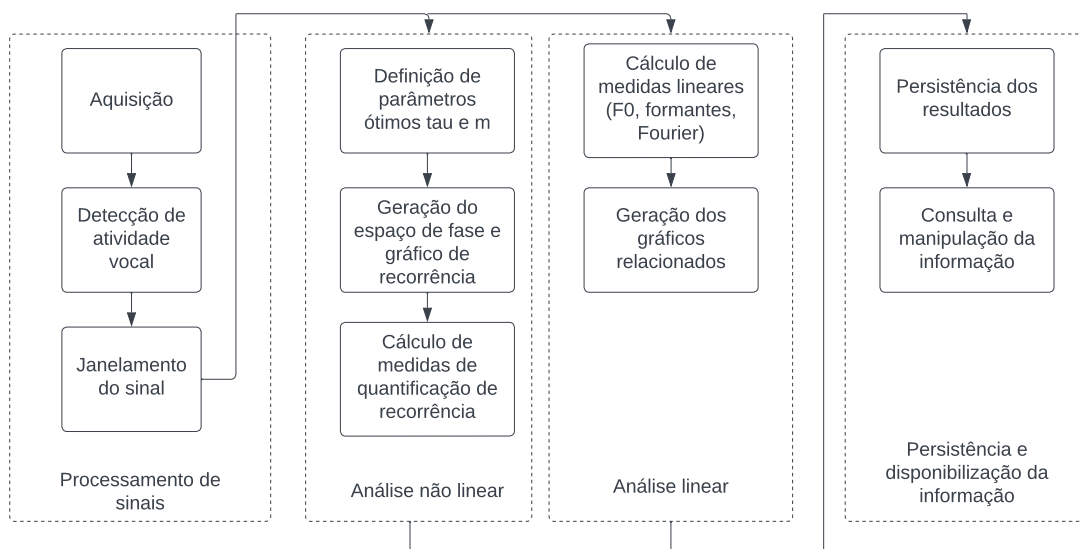


Estes resultados foram armazenados tanto em um sistema de gerenciamento de banco de dados, quanto num sistema de arquivos, o que favoreceu a composição de relatórios que auxiliem o diagnóstico de desordens vocais. O fluxo descrito pode ser visualizado na Figura 28.

Em relação ao critério de seleção do tamanho da janela de análise, as propriedades estatísticas dos sinais de voz podem ser consideradas invariantes no tempo, para curtos intervalos, até 32 ms, sendo um valor típico 16 ms. Utilizando-se essa característica, procura-se obter os parâmetros temporais do sinal a partir de segmentos que se situem nesse intervalo de interesse mantendo-se, assim, a estacionaridade do sinal de voz (RABINER; SCHAFER, 1978; FECHINE, 2000; COSTA, 2008). Já para a frequência de amostragem, esta segue o princípio apresentado na Seção 2.3.1.1, qual seja o de se manter pelo menos duas vezes a frequência máxima a ser considerada para análise. Nesta dissertação, o valor utilizado foi de 22050 Hz, sendo o padrão utilizado pela biblioteca de leitura dos arquivos do sinal de áudio, valor este passível de alteração.

Para que seja possível o uso da ferramenta de forma adequada, é necessário que a ferramenta atenda a um conjunto de requisitos mínimos, como uma placa de áudio que suporte a captura com a frequência de amostragem mínima (22050 Hz) bem como a seleção por intermédio de *software* da referida taxa de amostragem.

Figura 28 – Fluxograma da solução proposta



Fonte: Autor.

### 3.2.1 Base de dados de teste

Neste estudo de caso, é utilizada a base de dados *Disordered Voice Database, Model 4337*, da Kay Elemetrics, gravada pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (ELEMETRICS, 1994). Dessa base de dados, foram utilizados 53 sinais de voz de locutores com laringes saudáveis e 118 sinais de voz de locutores afetados por patologias

laríngeas (55 sinais de voz de laringes afetadas por paralisia nas pregas vocais, 45 sinais de voz de laringes afetadas por edema de Reinke e 18 sinais de voz de laringes afetadas por nódulos vocais). Os sinais são da vogal sustentada /a/. Os sinais de voz de laringes saudáveis, originalmente amostrados a 50.000 amostras/s, foram sub-amostrados a 25.000 amostras/s para serem equiparadas à taxa de amostragem dos sinais de voz de laringes patológicas (VIEIRA, 2014). Além disso, é possível a utilização dos sinais gravados pela ferramenta ou enviados pela interface de *uploads*.

### 3.3 Resultados obtidos

Para a verificação da viabilidade da proposta, foi construído um protótipo de um sistema Web, baseado em um recorte das tecnologias utilizadas na metodologia. Esse protótipo se divide, basicamente, em dois módulos: aquisição e análise da informação. Este também possui uma página inicial, na qual é informada a motivação, equipe de desenvolvimento e um breve tutorial apresentando suas funcionalidades do protótipo proposto.

É interessante a implementação de uma interface baseada em componentes, sendo implementada por um framework com React.JS, por exemplo, além da adequação do sistema para uma infraestrutura baseada em microsserviços.

#### 3.3.1 Aquisição do sinal de voz

A tela inicial do sistema possui uma logo com o nome do sistema, seguida da identificação do discente e dos orientadores do trabalho. Além disso, são apresentados o propósito e motivação do trabalho. Tal descrição é apresentada na Figura 29.

Figura 29 – Tela inicial



Fonte: Autor.

A Seção marcada em verde, na Figura 30, é o link do menu para realizar a aquisição. No item 1, é possível definir o nome do sinal ao ser realizada a gravação, conforme é visto no item 8b. Caso um nome não seja definido, o sistema gerará um nome, baseado na data no formato ISO, conforme apresentado no item 8a. Já o item 2 define a classe de diagnóstico do sinal, que pode ser "Sem diagnóstico", "Normal"(ou saudável), "Edema", "Paralisia"ou "Nódulo". Esse item define a classe tanto para a área de *upload* quanto para a gravação. No item 3, é possível selecionar pelo menos um áudio já gravado no computador, no formato *WAVEform audio format* (WAV), para que seja enviado para o servidor no item 4. O processo de gravação pode ser inicializado clicando no botão "Record"(item 5). Caso se deseje realizar uma pausa na gravação, deve-se clicar em "Pause"(item 6). Para finalizar o procedimento, deve-se clicar em "Stop"(item 7). Quando clicado, será gerada uma entrada da gravação, conforme apresentado no item 8. Depois de verificada a gravação, clica-se no link "Upload"associado a gravação (marcação azul). Tal descrição pode ser verificada na Figura 30.

Figura 30 – Módulo de aquisição

The screenshot shows the 'Voice signal acquisition' interface. On the left is a navigation menu with 'Acquisition' highlighted in green. The main area is divided into three sections: 'File Metadata', 'File Upload', and 'File recording'. In 'File Metadata', 'File name' is 'teste' (1) and 'Signal class' is 'Sem diagnostico' (2). 'File Upload' has a 'Choose Files' button (3) and an 'Upload' button (4). 'File recording' has 'Record' (5), 'Pause' (6), and 'Stop' (7) buttons. Below are 'Recordings' with two entries: '2022-06-29T13:42:08.wav' (8a) and 'teste.wav' (8b), each with a play button, progress bar, volume icon, and an 'Upload' link.

Fonte: Autor.

### 3.3.2 Análise da Informação

A tela de análise possibilita a visualização da computação realizada no processo de aquisição. Devido a sua natureza, o projeto se separa em duas grandes áreas de análise: linear e não linear. Para que seja possível realizar a análise, é necessário selecionar um sinal, conforme apresentando nos itens 1 e 2, na Figura 31.

Após selecionado o sinal de interesse, deve ser clicado o botão "Load", comando que permitirá a apresentação de uma nova janela contendo os gráficos e medidas associadas bem

como o título da amostra em análise. As análises podem ser realizadas de duas formas: visual e quantitativa. Tal cenário é apresentado na Figura 32. Ainda na Figura 32, na Seção de análise visual linear, é possível verificar a medida *Cepstral Peak Proeminence - CPP* (a), o espectrograma (b), o espectro LPC (c) e o espectro de Fourier (d). Para a geração de elementos visuais a partir da análise não linear, são apresentados um segmento de 100 ms do sinal (a), o retrato de fase do sinal (b) e o gráfico de recorrência (c). Tal visualização pode ser vista na Figura 33.

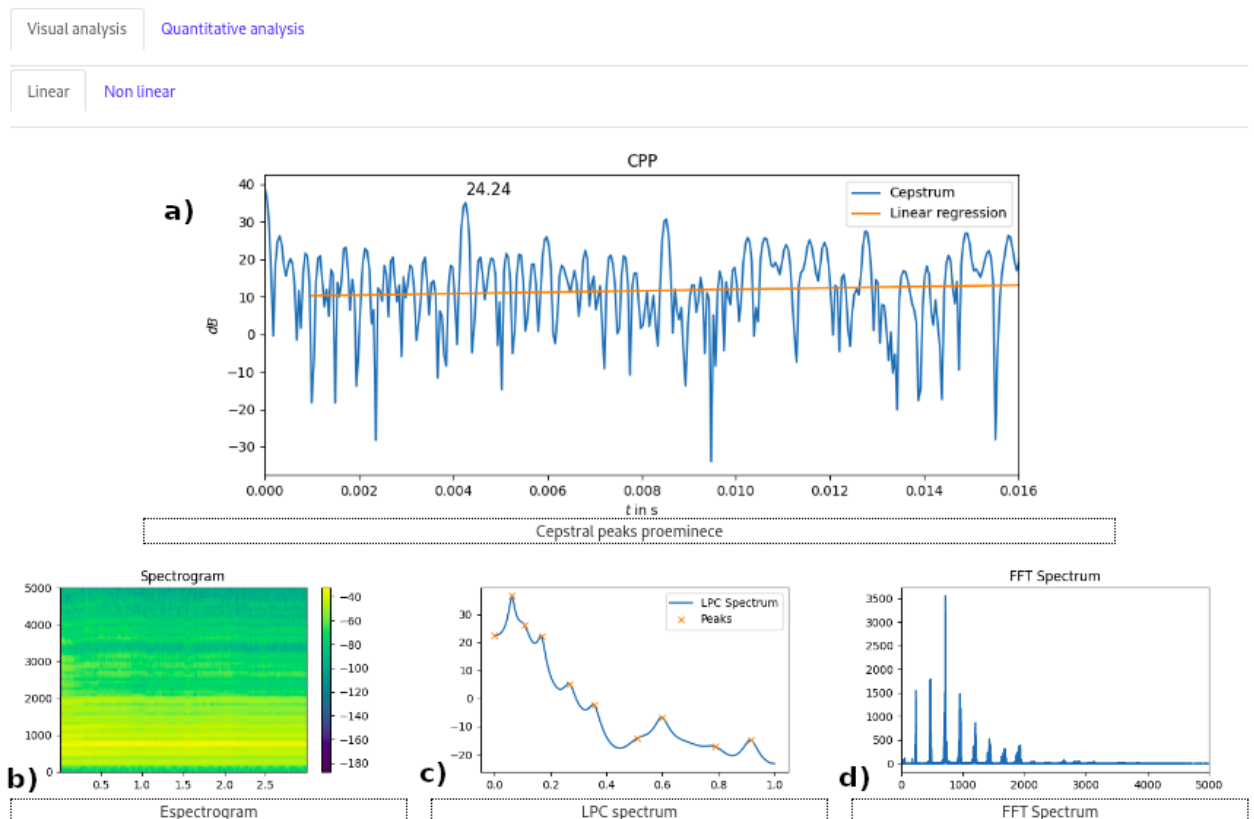
Figura 31 – Módulo de análise da informação - Seleção de sinal



Fonte: Autor.

Figura 32 – Módulo de análise da informação - Interface

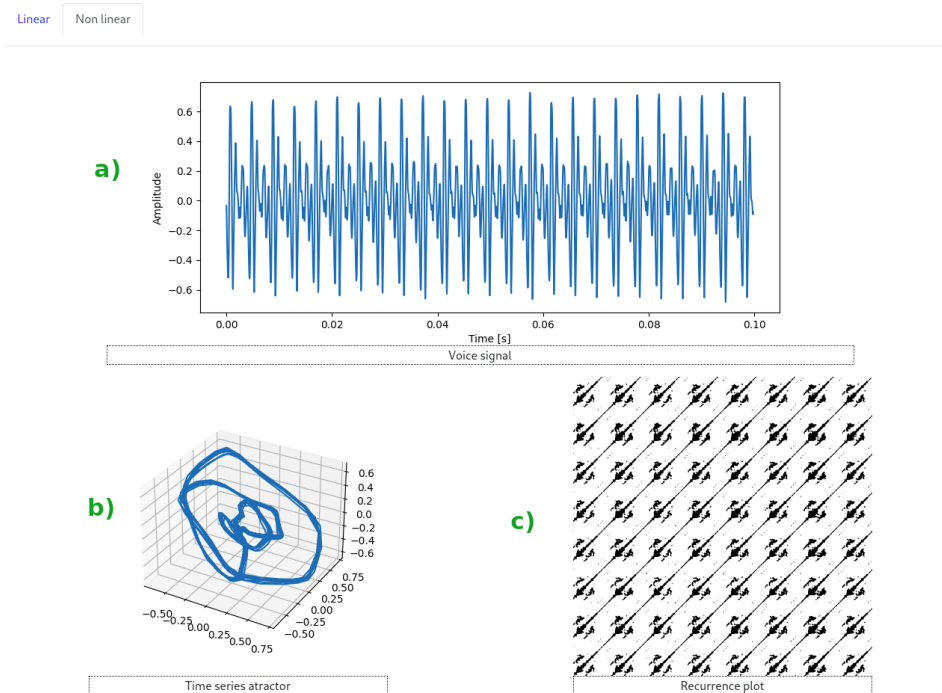
Visual and quantitative analysis - normal1.wav



Fonte: Autor.

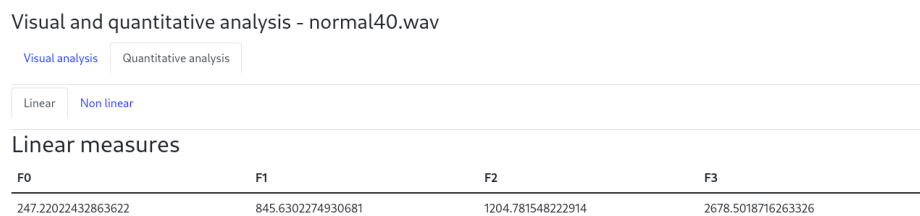
Na Seção de análise quantitativa, para as medidas lineares, são apresentadas a frequência fundamental (F0) e as frequências formantes (F1, F2 e F3), como apresentado na Figura 34.

Figura 33 – Módulo de análise da informação - Seção visual não linear (a) Sinal de voz, (b) retrato de fase e (c) gráfico de recorrência



Fonte: Autor.

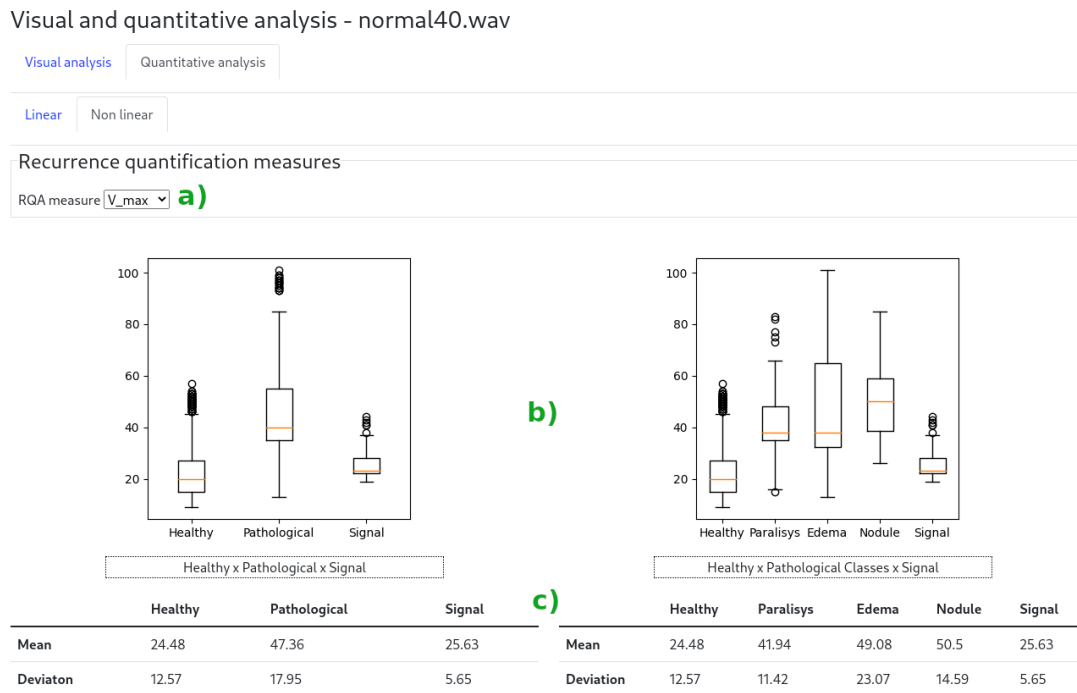
Figura 34 – Módulo de análise da informação - Seção quantitativa linear Freqüências Fundamental (F0) e Formantes (F1, F2, F3)



Fonte: Autor.

A aba não linear contém um seletor de medidas de quantificação de recorrência (a), sendo observados o comprimento máximo da diagonal ( $L_{max}$ ), o comprimento médio das linhas diagonais ( $L$ ), o determinismo (DET), o comprimento máximo das linhas verticais ( $V_{max}$ ) e o tempo de permanência (TT). Com o intuito de facilitar a visualização dos resultados, gráficos *BoxPlots* (b) exibem as comparações do sinal e as classes saudável e patológico e seu detalhamento por patologias (Paralisia, Edema e Nódulo). Também é apresentada uma sumarização em uma tabela, contendo as médias dos valores obtidos (c). Tal descrição pode ser observada na Figura 35.

Figura 35 – Módulo de análise da informação - Medidas de quantificação de recorrência. Seletor de medida (a), *BoxPlots* (b) e Sumarização (c)



Fonte: Autor.

### 3.4 Análise comparativa com base de dados de referência

Conforme mencionado na metodologia, foi utilizada uma base de dados contendo sinais saudáveis e patológicos, com classes de Paralisia, Edema e Nódulo (ELEMETRICS, 1994). Alguns desses sinais da base foram adquiridos pelo módulo de aquisição, os quais foram processados, extraídas suas medidas e gerados os gráficos que são armazenados no SGDB e no sistema de arquivos, respectivamente.

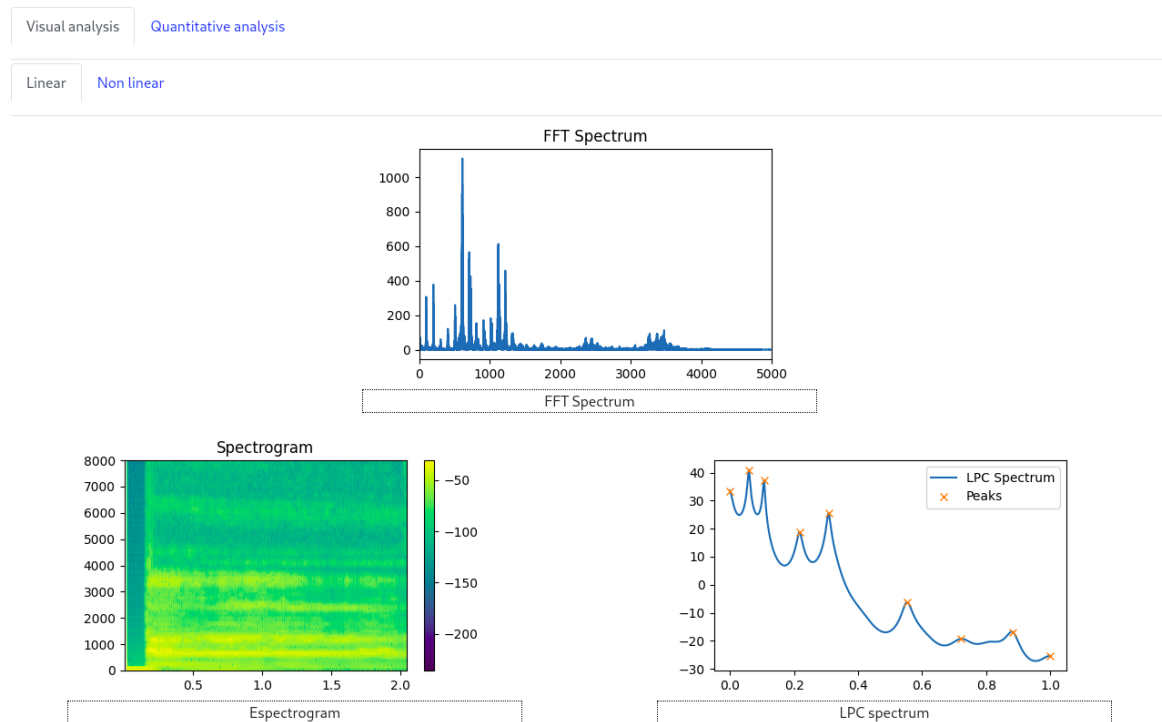
Para que fosse possível a geração dos resultados, o autor deste trabalho gravou sinais de voz, de aproximadamente 2 segundos, no sistema desenvolvido pela interface WEB, contendo a vogal sustentada /a/ para fins de teste, utilizando o microfone embutido de um notebook Dell, com a placa de som para aquisição Intel Corporation 7 Series/C216 Chipset. As saídas gráficas para as análises visual linear e não linear são apresentadas nas Figuras 36 e 37.

É possível observar no início do espectrograma, contido na Figura 36, um momento de silêncio, no início da gravação, o que pode ser observado pela baixa intensidade apresentada; isso ocorreu devido às difíceis condições de realização do procedimento. Além disso, a maior concentração de energia do sinal encontra-se nas frequências abaixo de 2000 Hz.

No gráfico de recorrência apresentado na Figura 37, é possível visualizar linhas diagonais bem definidas e estruturas definidas em diagonais paralelas a estas.

Figura 36 – Saída dos elementos visuais de análise linear para o sinal de teste gravado

Visual and quantitative analysis - elliott.wav



Fonte: Autor.

Para as medidas quantitativas, na seção linear puderam ser obtidos os valores da Frequência Fundamental em 102,53 Hz e as frequências formantes  $F1 = 660,30$  Hz,  $F2 = 1186,84$  e  $F3 = 2438,80$ . Os valores podem ser visualizados na interface apresentada na Figura 38.

Conforme apresentado na seção da metodologia, a análise quantitativa baseada em recorrência é apresentada de duas formas: por *BoxPlots* e resumida em uma tabela contendo a média e o desvio padrão. Para a medida de recorrência  $V_{max}$ , é apresentada, na Figura 39, a configuração dos *BoxPlots*, exibindo-se tanto as classes patológicas agrupadas quanto as segregadas.

Pelos *BoxPlots* apresentados, é possível observar que a mediana do sinal se aproxima da classe "Saudável" ou "Healthy". Já a sumarização é apresentada na Tabela 2 a seguir.

Tabela 2 – Tabela com valores para a medida de recorrência  $V_{max}$  para cinco classes

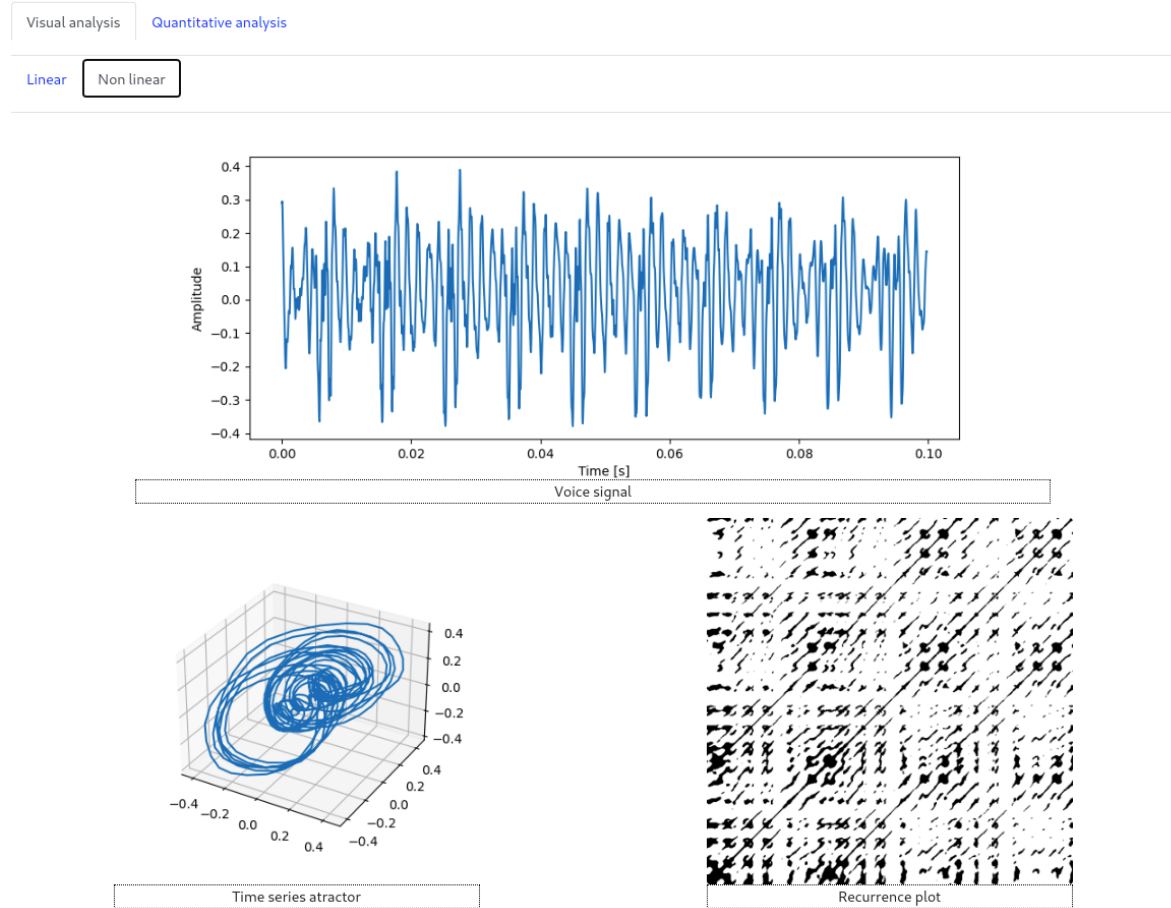
	<b>Saudável</b>	<b>Paralisia</b>	<b>Edema</b>	<b>Nódulo</b>	<b>Sinal</b>
Média	24.48	41.94	49.08	50.5	34.19
Desvio padrão	12.57	11.42	23.07	14.59	16.12

Fonte: Autor.

Já para a sumarização tabelas "Healthy" ou "Saudável" versus "Pathological" ou "Patológico" são apresentadas na Tabela 3.

Figura 37 – Saída dos elementos visuais de análise não linear para o sinal de teste gravado

Visual and quantitative analysis - elliott.wav



Fonte: Autor.

Figura 38 – Frequências fundamental e formantes do sinal de teste

Visual and quantitative analysis - elliott.wav

Visual analysis   Quantitative analysis

Linear   **Non linear**

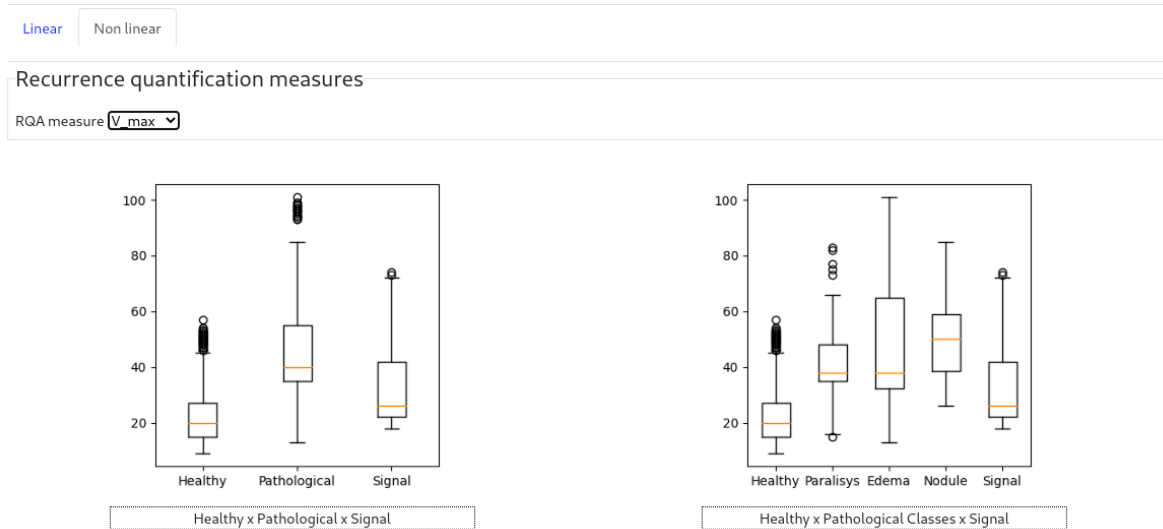
Linear measures

F0	F1	F2	F3
102.53670522057985	660.301513671875	1186.8453369140625	2438.800537109375

Fonte: Autor.



Figura 39 – *BoxPlots* para a medida de recorrência  $V_{max}$



Fonte: Autor.

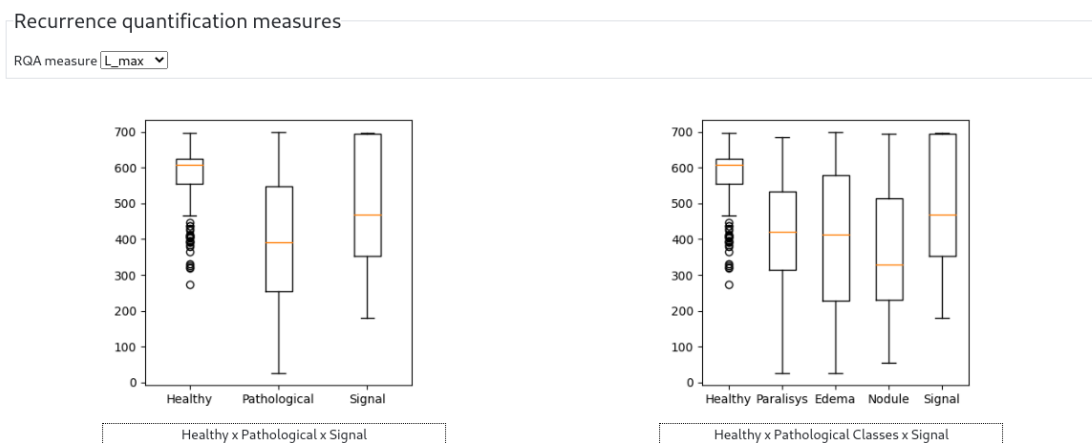
Tabela 3 – Tabela com valores para a medida de recorrência  $V_{max}$  para três classes

	<b>Saudável</b>	<b>Patológico</b>	<b>Sinal</b>
Média	595.78	398.63	473.02
Desvio padrão	73.26	186.95	175.14

Fonte: Autor.

Alterando a medida de recorrência para o comprimento máxima da diagonal ( $L_{max}$ ), os *BoxPlots* apresentam a conformação apresentada na Figura 40.

Figura 40 – *BoxPlots* para a medida de recorrência  $L_{max}$



Fonte: Autor.

Neste caso, o sinal de teste fica com a mediana dentro dos intervalos apresentados para as patologias, tendo o valor mais próximo, visualmente, a paralisia e o edema. Os valores referentes a sumarização dos valores para o agregado e para classes distintas são apresentados nas Tabelas 4 e 5.

Tabela 4 – Tabela com valores para a medida de recorrência  $L_{max}$  para cinco classes

	<b>Saudável</b>	<b>Paralisia</b>	<b>Edema</b>	<b>Nódulo</b>	<b>Sinal</b>
Média	595.78	413.13	415.22	364.74	473.02
Desvio padrão	73.26	148.81	201.05	197.56	175.14

Fonte: Autor.

Pelos valores apresentados na Tabela 4, pela mediana, é possível confirmar que o sinal apresentado está enquadrado entre as classes Paralisia e Edema.

Tabela 5 – Tabela com valores para a medida de recorrência  $L_{max}$  para duas classes

	<b>Saudável</b>	<b>Patológico</b>	<b>Sinal</b>
Média	595.78	398.63	473.02
Desvio padrão	73.26	186.95	175.14

As medidas  $L_{med}$  e TT analisam o tempo gasto, sendo o primeiro o tempo médio em que dois segmentos evoluem na mesma trajetória, e o segundo em um determinado estado específico. Já na formação de estruturas, as medidas DET e LAM definem a razão entre os pontos de recorrência para criação de estruturas diagonais e verticais, respectivamente.

A aplicação adequada de técnicas de pré-processamento (filtragem e pré-ênfase, por exemplo) permitem que sejam extraídas as características adequadas de forma fidedigna para a inferência da classe de diagnóstico, independentemente das condições de captação do referido sinal. Nesse contexto, é possível observar que, mesmo com a obtenção de medidas quantitativas objetivas, a análise de um profissional se torna indispensável para a verificação de qual medida irá melhor refletir a necessidade de auxílio ao diagnóstico, conforme seja sua interpretação quando da obtenção dos intervalos de classificação das medidas quantitativas.

Para além disso, a necessidade de adequação dos resultados às diferentes telas se faz necessária, haja vista a interface não estar adequada para ser uma interface responsiva. Para tanto, o desenvolvimento de aplicações direcionadas para plataformas específicas, como, por exemplo, dispositivos móveis, como o celular ou o uso de *frameworks* de desenvolvimento, como Flutter, Vue.JS e/ou React.

## 4 CONCLUSÕES

Este trabalho teve por objetivo a verificação da viabilidade do desenvolvimento de um solução *e-Health* voltada para aquisição, processamento, extração de características lineares e não lineares, armazenamento e posterior consulta por intermédio de um navegador Web. Foi possível realizar a implementação de tal sistema, favorecendo-se o desenvolvimento de módulos de aquisição, que permitindo-se o envio e gravação no próprio navegador de amostras de sinais de áudio, e análise visual e quantitativa.

No escopo da análise, é possível realizar a geração de gráficos a partir da análise linear da fala (espectro de Fourier, cepstro, espectrograma e LPC) e associados não linear, no escopo de recorrência, possibilitando a visualização do atrator e gráfico de recorrência associado. Ademais, na seção quantitativa, é possível obter as frequências fundamental e formantes, no escopo linear, e a exibição de seis medidas de recorrência, sendo possível a consulta de  $L_{max}$ ,  $L$ ,  $V_{max}$ , DET e TT. Para estas medidas, a consolidação pode ser realizada por intermédio de *BoxPlots* e de uma tabela, que as separa por classes gerais (Saúdavel *versus* Patológica) ou especializadas (Paralisia, Edema e Nódulo). Em ambas as situações, foi realizada a comparação com o sinal de interesse.

Para verificação funcional da solução, esta foi implantada em um servidor de produção localizado na AWS (Amazon Web Services), tendo um domínio registrado, possibilitando assim a simulação de uso real da aplicação.

### 4.1 Contribuições da pesquisa

Esta Tese contribuiu nos seguintes aspectos:

1. Implementação de solução voltada para web utilizando medidas lineares e de quantificação de recorrência por intermédio da Web;
2. Verificação funcional de pacotes de código aberto para extração de características de sinais de voz;
3. Implementação de processamento baseado em filas para cálculos computacionais pesados;
4. Implementação e verificação de gravador por intermédio da Web, permitindo a aquisição remota de sinais de voz;
5. Comparação de sinais de voz gravados com a base de dados de referência, para verificação funcional da solução;

6. Disponibilização de ferramenta que permita a análise de distúrbios vocais, visando à prevenção e ao auxílio ao diagnóstico, além do acompanhamento em casos de terapia vocal e,
7. Viabilização de uso da solução para treinar profissionais no uso de ferramentas de análise acústica, com enfoque em recorrência.

## 4.2 Sugestões para trabalhos futuros

O produto resultante deste trabalho pode, posteriormente, ganhar novas fases de desenvolvimento, com o intuito de agregar melhorias. Devido ao direcionamento da aplicação voltada para uso público, esta precisa ser adaptada para um processamento distribuído, levando em consideração a sua escalabilidade e disponibilidade, sendo viabilizada pelo processo de refatoração (FOWLER, 2020). Neste processo, o *software* atual, modelado em uma arquitetura monolítica, será segregado baseado em suas responsabilidades em microsserviços (NEWMAN, 2022), que poderão ser acessados por intermédio de uma API ou por comunicação de mensagens (JOHANSSON; DOSSOT, 2020). A arquitetura de microsserviços propicia a distribuição da aplicação em outros servidores, com o objetivo de distribuir a carga da aplicação, possibilitando a especialização do processo em uma determinada máquina. Pode-se citar o processo de computação do gráfico e de seus parâmetros ótimos. A rotina já se utiliza de pacotes de paralelização para a computação científica. No caso do uso de mais de um processo de forma paralelizada, este irá sobrecarregar a máquina hospedeira, reduzindo, então, sua performance.

Lastreado nessa prerrogativa, um modelo de arquitetura de *software* baseada em microsserviços é proposto. O modelo possui seis microsserviços essenciais, que possuem sua própria pilha tecnológica, ou seja, possuem linguagem de programação, *framework* e banco de dados individualizados, permitindo assim a completa segregação entre estes.

Assim, são definidos os microsserviços de interface com usuário (Afrodite), *gateway* e despacho de requisições (Hermes), análise de séries temporais (Chronos), geração de relatórios (Métis), armazenamento de mídias de áudio, imagens, figuras e arquivos em geral (Mnemósine) e armazenamento de informações de pacientes, profissionais e prontuário (Apolo).

O usuário da solução, ao acessar o sistema, irá realizar o acesso ao serviço Afrodite, o qual carregará as telas de login, módulos de gerenciamento, permitindo a operação da solução. A interface manterá a sua comunicação com o serviço que realiza a intermediação da comunicação, Hermes, que também realiza o controle de acesso de usuários ao sistema e define os seus respectivos papéis. Dependendo da natureza da requisição, este realizará o encaminhamento para o serviço responsável pelo processamento. Solicitações de processamento de sinais recém adquiridos, serão direcionados para o serviço Chronos, que se encarregará de extrair as medidas e gráficos programados. Os serviços podem se comunicar para completar determinadas tarefas, podendo citar a tarefa de armazenar os sinais de voz e gráficos gerados pelo serviço Chronos,

o qual enviar a requisição para o serviço Mnemósine. Tarefas de processamento prolongados, como a geração de relatórios gerais, será realizada pelo serviço Métis. Do armazenamento de informações relacionadas a pacientes, profissionais, prontuários de atendimento, contendo rodadas de avaliação, o serviço Apolo será encarregado. Com o intuito de aumentar a confiabilidade das sugestões ao profissional de saúde, é planejado o desenvolvimento de um serviço encarregado pela implementação de aprendizado de máquina, Temis.

Para possibilitar a implantação de uma arquitetura de microsserviços, é necessário o correto planejamento e implementação de uma infraestrutura computacional que seja provida de recursos computacionais escaláveis bem como o provisionamento e configuração. Neste sentido, será configurado um ambiente de Cloud Computing baseado em Kubernetes, Docker, por intermédio do Ansible e Terraform, tecnologias estas que permitem a implantação de soluções de forma rápida e escalável.

A interface com o usuário do sistema será desenvolvida utilizando React.JS, possibilitando a implementação de módulos de gerenciamento de usuários do sistema bem como a definição de papéis e permissões, gerenciamento de pacientes e profissionais, relatórios, rodadas de avaliação e gestão dos sinais de voz disponíveis. Para um melhor entendimento das necessidades que devem ser implementadas, são de relevante interessante o levantamento e análise de requisitos funcionais por um ou mais profissionais da área de fonoaudiologia bem como a sua validação com os referidos.

Além das limitações da solução computacional, o desenvolvimento de novas formas para a aquisição do sinal e seu tratamento pode ser implementado na solução, como a captura da vibração da pele do pescoço (SILVA, 2020; GUIMARÃES, 2020; LIMA, 2022), de forma suplementar ou combinada à aquisição de sinais de voz e sua correlação, por exemplo.

## REFERÊNCIAS

FERNANDA Soares Lima. Tese (Doutorado em Engenharia Elétrica) — Escola de Engenharia de São Carlos, Universidade de São Paulo. Disponível em: <<https://doi.org/10.11606/t.18.2002.tde-19112015-110520>>. Citado na página 54.

AL-NASHERI, A. *et al.* An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. *Journal of Voice*, Elsevier BV, v. 31, n. 1, p. 113.e9–113.e18, jan. 2017. Disponível em: <https://doi.org/10.1016/j.jvoice.2016.03.019>. Citado na página 27.

ALENCAR, S. A. L. de *et al.* Efetividade de duas modalidades terapêuticas na redução dos sintomas vocais em pacientes com disfonia comportamental. *Audiology - Communication Research*, FapUNIFESP (SciELO), v. 25, 2020. Disponível em: <https://doi.org/10.1590/2317-6431-2019-2126>. Citado na página 25.

ALSHUQAYRAN, N.; ALI, N.; EVANS, R. A systematic mapping study in microservice architecture. In: *2016 IEEE 9th International Conference on Service-Oriented Computing and Applications (SOCA)*. IEEE, 2016. Disponível em: <https://doi.org/10.1109%2Fsoca.2016.15>. Citado na página 22.

ALTMANN, E. G. "*Tempo de retorno em sistemas dinâmicos*". Dissertação (Mestrado em Ciências Físicas) — Instituto de Física, Universidade de São Paulo, 2004. Disponível em: <https://doi.org/10.11606/d.43.2004.tde-02032004-124530>. Citado na página 54.

BAKEN, R.; ORLIKOFF, R. *Clinical Measurement of Speech and Voice*. Singular Thomson Learning, 2000. (Speech Science). ISBN 9781565938694. Disponível em: <https://books.google.com.br/books?id=ElPyvaJbDiwC>. Citado 2 vezes nas páginas 37 e 38.

BARROS, L. *et al.* Classification of recurrence plots of voice signals using convolutional neural networks. In: *Anais de XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. Sociedade Brasileira de Telecomunicações, 2020. Disponível em: <https://doi.org/10.14209/sbrt.2020.1570661665>. Citado 2 vezes nas páginas 18 e 54.

BEHLAU, M. *Voz: o livro do especialista*. [S.l.: s.n.], 2001. Citado na página 25.

BERG, J. van den. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, American Speech Language Hearing Association, v. 1, n. 3, p. 227–244, set. 1958. Disponível em: <https://doi.org/10.1044/jshr.0103.227>. Citado na página 24.

BITAR, H.; ALISMAIL, S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: A rapid systematic review. *DIGITAL HEALTH*, SAGE Publications, v. 7, p. 205520762110093, jan 2021. Disponível em: <https://doi.org/10.1177%2F20552076211009396>. Citado na página 19.

BOGERT, B. P. The quefrency analysis of time series for echoes: Cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. In: *Proc. Symposium Time Series Analysis, 1963*. [S.l.: s.n.], 1963. p. 209–243. Citado 2 vezes nas páginas 44 e 47.

BOYANOV, B.; HADJITODOROV, S. Acoustic analysis of pathological voices. a voice analysis system for the screening of laryngeal diseases. *IEEE Engineering in Medicine and Biology Magazine*, v. 16, n. 4, p. 74–82, 1997. Citado na página 27.

BSONSPEC. *BSON (Binary JSON) Serialization* — *bsonspec.org*. 2022. <http://bsonspec.org/>. [Accessed 06-Aug-2022]. Citado na página 70.

BUELTA, J. *Hands-On Docker for Microservices with Python*. [S.l.: s.n.], 2019. ISBN 9781838823818. Citado na página 64.

CAVALCANTE, A. de V. *ANÁLISE DOS EFEITOS DE CODECS DE ÁUDIO NA AVALIAÇÃO DE DESVIOS VOCAIS*. Dissertação (Dissertação de Mestrado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, 2018. Citado na página 21.

CHANDRASHEKAR, H.; KARJIGI, V.; SREEDEVI, N. Breathiness indices for classification of dysarthria based on type and speech intelligibility. In: *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. [S.l.: s.n.], 2019. p. 266–270. Citado na página 18.

CHENGALVARAYAN, R. Hierarchical subband linear predictive cepstral (hslpc) features for hmm-based speech recognition. In: *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*. [S.l.: s.n.], 1999. v. 1, p. 409–412 vol.1. Citado na página 27.

CHIBA, T.; KAJIYAMA, M. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan, 1941. Disponível em: <https://books.google.com.br/books?id=tpkKAAAAMAAJ>. Citado na página 28.

CLARK, J.; YALLOP, C.; FLETCHER, J. *An introduction to phonetics and phonology*. 2. ed. London, England: Blackwell Publishing, 1995. (Blackwell Textbooks in Linguistics). Citado 3 vezes nas páginas 36, 37 e 48.

COMER, D. *The Cloud Computing Book*. [S.l.: s.n.], 2021. ISBN 9780367706807. Citado na página 64.

CONSELHO FEDERAL DE FONOAUDIOLOGIA. *Diretrizes de boas práticas em telefonaudiologia*: Volume 1. Brasília, 2020. 95 p. Citado 2 vezes nas páginas 19 e 20.

CONSTANTINESCU, G. *et al.* Treating disordered speech and voice in parkinson's disease online: a randomized controlled non-inferiority trial. *Int. J. Lang. Commun. Disord.*, v. 46, n. 1, p. 1–16, jan. 2011. Citado na página 20.

CONSTANTINESCU, G. A. *et al.* Home-based speech treatment for parkinson's disease delivered remotely: a case report. *Journal of Telemedicine and Telecare*, SAGE Publications, v. 16, n. 2, p. 100–104, dec 2009. Disponível em: <https://doi.org/10.1258%2Fjtt.2009.090306>. Citado na página 20.

COSTA, S. L. do N. C. *Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas*. Tese (Doutorado em Engenharia Elétrica) — Centro de Engenharia Elétrica e Informática - Program de Pós Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, Brasil, 2008. Acesso em: 26 jul. 2022. Citado 7 vezes nas páginas 18, 26, 28, 34, 53, 70 e 72.

COSTA, W. C. d. A. *et al.* Reconhecimento de fala utilizando modelos de markov escondidos (hmm's) de densidades contínuas. Universidade Federal de Campina Grande, 1994. Citado 2 vezes nas páginas 27 e 28.

COSTA, W. C. de A. *Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas*. Tese (Doutorado em Engenharia Elétrica) — Centro de Engenharia Elétrica e Informática - Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, Brasil, 2012. Acesso em: 26 jul. 2022. Citado 7 vezes nas páginas 18, 24, 25, 54, 56, 57 e 70.

COSTA, W. C. de A. *et al.* Pathological voice assessment by recurrence quantification analysis. In: *2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*. [S.l.: s.n.], 2012. p. 1–6. Citado na página 55.

CTS. *VOXMETRIA - ANLISE DE VOZ E QUALIDADE VOCAL* — *ctsinformatica.com.br*. 2022. <https://www.ctsinformatica.com.br/fonoaudiologia/voz/voxmetria-analise-de-voz-e-qualidade-vocal>. [Accessed 31-Jul-2022]. Citado na página 21.

DELLER, J. R.; HANSEN, J. H. L.; PROAKIS, J. G. *Discrete-time processing of speech signals*. Nashville, TN: John Wiley & Sons, 1999. Citado 2 vezes nas páginas 31 e 54.

DIMER, N. A. *et al.* Pandemia do COVID-19 e implementação de telefonaudiologia para pacientes em domicílio: relato de experiência. *CoDAS*, FapUNIFESP (SciELO), v. 32, n. 3, 2020. Disponível em: <https://doi.org/10.1590%2F2317-1782%2F20192020144>. Citado na página 19.

DJANGOBOOK. *Mastering Django: Structure - The Django Book* — *djangobook.com*. 2022. <https://djangobook.com/django-tutorials/mastering-django-structure/>. [Accessed 05-Aug-2022]. Citado na página 70.

DOARN, C. R. *et al.* Design and implementation of an interactive website for pediatric voice therapy—the concept of in-between care: A telehealth model. *Telemedicine and e-Health*, Mary Ann Liebert Inc, v. 25, n. 5, p. 415–422, may 2019. Disponível em: <https://doi.org/10.1089%2Ftmj.2018.0108>. Citado 2 vezes nas páginas 20 e 21.

DONCA, I.-C. *et al.* Autoscaled RabbitMQ kubernetes cluster on single-board computers. In: *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*. IEEE, 2020. Disponível em: <https://doi.org/10.1109%2Faqtr49680.2020.9129886>. Citado na página 22.

DRESCH, A. A. G. *Método para reconhecimento de vogais e extração de parâmetros acústicos para análises forenses*. Dissertação (Dissertação de Mestrado em Engenharia Biomédica) — CT - Programa de Pós-Graduação em Engenharia Biomédica, Universidade Tecnológica Federal do Paraná, 2015. Citado na página 53.

ECKMANN, J.-P.; KAMPHORST, S. O.; RUELLE, D. Recurrence plots of dynamical systems. *European Physics Letters (EPL)*, IOP Publishing, v. 4, n. 9, p. 973–977, nov 1987. Disponível em: <https://doi.org/10.1209/0295-5075/4/9/004>. Citado na página 57.

ELEMETRICS, K. *Kay elemetrics corp. disordered voice database*. 1994. [Accessed 31-Jul-2022]. Citado 2 vezes nas páginas 72 e 77.



- EREN, H.; WEBSTER, J. G. (Ed.). *Telehealth and Mobile Health*. London, England: CRC Press, 2015. Citado 2 vezes nas páginas 19 e 21.
- FANT, G. *Acoustic Theory of Speech Production*. Berlin, Boston: De Gruyter Mouton, 1960. ISBN 9783110873429. Disponível em: <https://doi.org/10.1515/9783110873429>. Citado na página 28.
- FECHINE, J. M. *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística*. Tese (Doutorado em Engenharia Elétrica) — Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica, Universidade Federal da Paraíba, Campina Grande, Brasil, 2000. Acesso em: 26 jul. 2022. Citado 5 vezes nas páginas 18, 27, 28, 35 e 72.
- FELLBAUM, K. *Sprachverarbeitung und Sprachübertragung*. Berlin, Germany: Springer, 1984. (Nachrichtentechnik). Citado na página 35.
- FERNANDES, N. M. da S. *et al.* Telemedicine: Development of a distance care system for pre-dialysis chronic kidney disease patients. *Jornal Brasileiro de Nefrologia*, FapUNIFESP (SciELO), v. 37, n. 3, 2015. Disponível em: <https://doi.org/10.5935%2F0101-2800.20150055>. Citado na página 19.
- FERRARI, D. V.; BERNARDEZ-BRAGA, G. R. A.; CAMPOS, P. D. Verificação da prótese auditiva realizada face a face e via teleconsulta: medidas repetidas. *Revista CEFAC*, FapUNIFESP (SciELO), v. 14, n. 6, p. 1061–1071, sep 2011. Disponível em: <https://doi.org/10.1590%2Fs1516-18462011005000104>. Citado na página 20.
- FERREIN, C. J. *Mémoires de l'Académie des Sciences des Paris*. [S.l.: s.n.], 1741. Citado na página 28.
- FERREIRA, M. T. *Métodos lineares e não lineares de análise de séries temporais e sua aplicação no estudo da variabilidade da frequência cardíaca de jovens saudáveis*. Dissertação (Mestrado em Biometria) — Instituto de Biociências de Botucatu, Universidade Estadual de São Paulo, 2010. Citado 2 vezes nas páginas 18 e 20.
- FOWLER, M. *Refatoração – 2ª edição*. [S.l.: s.n.], 2020. ISBN 9788575227244. Citado na página 83.
- FRAILE, R.; GODINO-LLORENTE, J. I. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, v. 14, p. 42–54, 2014. ISSN 1746-8094. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1746809414000986>. Citado na página 48.
- FRASER, S.; ZIADE, T. *Python Microservices Development*. [S.l.: s.n.], 2021. ISBN 9781801076302. Citado 2 vezes nas páginas 22 e 64.
- FU, L. Design and implementation of super mall system based on SOA distributed architecture. In: *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109%2Ficitbs.2019.00094>. Citado na página 22.
- GELDARD, F. A. Foreword. In: GELDARD, F. A. (Ed.). *Communication Processes*. Pergamon, 1965. p. ix–x. ISBN 978-0-08-010970-1. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780080109701500048>. Citado na página 18.

- GILMORE, R. Topological analysis of chaotic dynamical systems. *Rev. Mod. Phys.*, American Physical Society, v. 70, p. 1455–1529, Oct 1998. Disponível em: <https://link.aps.org/doi/10.1103/RevModPhys.70.1455>. Citado na página 61.
- GRASSBERGER, P.; PROCACCIA, I. Characterization of strange attractors. *Physical Review Letters*, American Physical Society (APS), v. 50, n. 5, p. 346–349, jan 1983. Disponível em: <https://doi.org/10.1103/PhysRevLett.50.346>. Citado na página 59.
- GUIMARÃES, A. M. *Avaliação de um acelerômetro capacitivo na mensuração de medidas acústicas relacionadas às vibrações das pregas vocais*. Dissertação (Dissertação de Mestrado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, 2020. Citado na página 84.
- HABER, M. J.; CHAPPELL, B.; HILLS, C. *Cloud Attack Vectors*. [S.l.: s.n.], 2022. ISBN 9781484282359. Citado na página 64.
- HEINRICH, R. *et al.* Performance engineering for microservices. In: *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion*. ACM, 2017. Disponível em: <https://doi.org/10.1145/3053600.3053653>. Citado na página 64.
- HILLENBRAND, J.; CLEVELAND, R. A.; ERICKSON, R. L. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, American Speech Language Hearing Association, v. 37, n. 4, p. 769–778, ago. 1994. Disponível em: <https://doi.org/10.1044/jshr.3704.769>. Citado 2 vezes nas páginas 18 e 48.
- HIRANO, M.; SATO, K. *Histological Color Atlas of the Human Larynx*. Singular Publishing Group, 1993. ISBN 9781565930773. Disponível em: <https://books.google.com.br/books?id=4OBqAAAAMAAJ>. Citado na página 24.
- HURWITZ, J. S.; KIRSCH, D. *Cloud Computing For Dummies*. [S.l.: s.n.], 2020. ISBN 9781119546658. Citado na página 64.
- INGLE, V. K.; PROAKIS, J. G. *Digital Signal Processing Using MATLAB*. 3. ed. Mason, OH: CENGAGE Learning Custom Publishing, 2006. Citado 3 vezes nas páginas 30, 31 e 32.
- ISMAIL, M. A.; DESHMUKH, S.; SINGH, R. Detection of covid-19 through the analysis of vocal fold oscillations. In: *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2021. p. 1035–1039. Citado na página 18.
- JANI, R. *et al.* Systematic review of the treatment of functional dysphonia and prevention of voice disorders. *Otolaryngology–Head and Neck Surgery*, Wiley, v. 138, n. 5, p. 557–565, maio 2008. Disponível em: <https://doi.org/10.1016/j.otohns.2008.01.014>. Citado na página 25.
- JOHANSSON, L.; DOSSOT, D. *RabbitMQ Essentials*. 2. ed. Birmingham, England: Packt Publishing, 2020. Citado na página 83.
- JOOS, M. *Acoustic Phonetics*, by Martin Joos ... Linguistic Society of America, 1948. Disponível em: [https://books.google.com.br/books?id=\\_zeTQwAACAAJ](https://books.google.com.br/books?id=_zeTQwAACAAJ). Citado na página 28.
- KANTZ, H.; SCHREIBER, T. *Nonlinear Time Series Analysis*. [S.l.: s.n.], 2003. ISBN 9780521821506. Citado 3 vezes nas páginas 55, 56 e 57.

KAVYA, S.; SHIVRAM, R. Detection of possibility of laryngeal cancer through mel frequency cepstrum coefficient analysis. In: *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*. [S.l.: s.n.], 2018. p. 575–579. Citado na página 18.

KEIDSER, G.; CONVERY, E. Self-fitting hearing aids. *Trends in Hearing*, SAGE Publications, v. 20, p. 233121651664328, jan 2016. Disponível em: <https://doi.org/10.1177/2331216516643284>. Citado na página 20.

KEMPELEN, W. von. *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine*. Frommann-Holzboog, 1791. (Grammatica universalis). Disponível em: <https://books.google.com.br/books?id=JAthAQAAMAAJ>. Citado na página 28.

KENT, R.; KENT, R.; READ, C. *The Acoustic Analysis of Speech*. Singular/Thomson Learning, 2002. ISBN 9780769301129. Disponível em: <https://books.google.com.br/books?id=e4tsAAAIAAJ>. Citado 2 vezes nas páginas 37 e 51.

KRATZENSTEIN, C. *Sur la naissance et la formation des Voyelles: (par M. Kratzensteinius, de l'Academie des Sciences de Saint-Petersbourg)*. [s.n.], 1780. Disponível em: <https://books.google.com.br/books?id=slfRcQAACAAJ>. Citado na página 28.

KRIEF, M. *Learning DevOps*. [S.l.: s.n.], 2019. ISBN 9781838642730. Citado na página 65.

KTH. *WaveSurfer — sourceforge.net*. 2022. <https://sourceforge.net/projects/wavesurfer/>. [Accessed 31-Jul-2022]. Citado na página 21.

LAPLANTE-LÉVESQUE, A.; PICHORA-FULLER, M. K.; GAGNÉ, J.-P. Providing an internet-based audiological counselling programme to new hearing aid users: A qualitative study. *International Journal of Audiology*, Informa UK Limited, v. 45, n. 12, p. 697–706, jan 2006. Disponível em: <https://doi.org/10.1080/2F14992020600944408>. Citado na página 20.

LATHROP, D. P.; KOSTELICH, E. J. Characterization of an experimental strange attractor by periodic orbits. *Physical Review A*, American Physical Society (APS), v. 40, n. 7, p. 4028–4031, oct 1989. Disponível em: <https://doi.org/10.1103/2Fphysreva.40.4028>. Citado na página 61.

LIEBERMAN, P.; BLUMSTEIN, S. E. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge University Press, 1988. Disponível em: <https://doi.org/10.1017/2Fcb09781139165952>. Citado na página 28.

LIEV. *Ouvindo Vozes*. 2022. Disponível em: <https://www.ouvindovozes.com.br/>. Citado na página 21.

LIMA, F. S. *Caracterização e análise de sinais obtidos pela vibração da pele do pescoço para avaliação da qualidade vocal*. Dissertação (Dissertação de Mestrado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, 2022. Citado na página 84.

LOIZOU, P. C. *Speech enhancement*. Boca Raton, FL: CRC Press, 2005. Citado 6 vezes nas páginas 29, 30, 38, 43, 44 e 45.

LOPES, L. W.; CAVALCANTE, D. P.; COSTA, P. O. da. Severity of voice disorders: integration of perceptual and acoustic data in dysphonic patients. *CoDAS*, FapUNIFESP (SciELO), v. 26, n. 5, p. 382–388, out. 2014. Disponível em: <https://doi.org/10.1590/2317-1782/20142013033>. Citado na página 27.

LOPES, L. W. *et al.* Acoustic assessment of the voices of children using nonlinear analysis: Proposal for assessment and vocal monitoring. *Journal of Voice*, Elsevier BV, v. 28, n. 5, p. 565–573, sep 2014. Disponível em: <https://doi.org/10.1016%2Fj.jvoice.2014.02.013>. Citado 2 vezes nas páginas 20 e 55.

LOPES, L. W. *et al.* Effectiveness of recurrence quantification measures in discriminating subjects with and without voice disorders. *Journal of Voice*, Elsevier BV, v. 34, n. 2, p. 208–220, mar 2020. Disponível em: <https://doi.org/10.1016%2Fj.jvoice.2018.09.004>. Citado 3 vezes nas páginas 27, 55 e 63.

MAKRIS, A. *et al.* MongoDB vs PostgreSQL: A comparative study on performance aspects. *GeoInformatica*, Springer Science and Business Media LLC, v. 25, n. 2, p. 243–268, jun 2020. Disponível em: <https://doi.org/10.1007%2Fs10707-020-00407-w>. Citado na página 70.

MANN, D. M. *et al.* COVID-19 transforms health care through telemedicine: Evidence from the field. *Journal of the American Medical Informatics Association*, Oxford University Press (OUP), v. 27, n. 7, p. 1132–1135, may 2020. Disponível em: <https://doi.org/10.1093%2Fjamia%2Focaa072>. Citado na página 19.

MARTINEZ, C.; RUFINER, H. Acoustic analysis of speech for detection of laryngeal pathologies. In: *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143)*. [S.l.: s.n.], 2000. v. 3, p. 2369–2372 vol.3. Citado na página 27.

Marwan, N. *Encounters with neighbours : current developments of concepts based on recurrence plots and their applications*. Tese (Doktor der Naturwissenschaften) — Institut Für Physike - Fakultät Mathematik und Naturwissenschaften, University of Potsdam, Germany, set. 2003. Citado na página 58.

MARWAN, N. HOW TO AVOID POTENTIAL PITFALLS IN RECURRENCE PLOT BASED DATA ANALYSIS. *International Journal of Bifurcation and Chaos*, World Scientific Pub Co Pte Lt, v. 21, n. 04, p. 1003–1017, apr 2011. Disponível em: <https://doi.org/10.1142%2Fs0218127411029008>. Citado na página 62.

MARWAN, N. *et al.* Recurrence plots for the analysis of complex systems. *Physics Reports*, Elsevier BV, v. 438, n. 5-6, p. 237–329, jan. 2007. Disponível em: <https://doi.org/10.1016/j.physrep.2006.11.001>. Citado na página 55.

MARWAN, N.; THIEL, M.; NOWACZYK, N. R. Cross recurrence plot based synchronization of time series. *Nonlinear Processes in Geophysics*, v. 9, n. 3/4, p. 325–331, 2002. Disponível em: <https://npg.copernicus.org/articles/9/325/2002/>. Citado na página 61.

MARWAN, N. *et al.* Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, American Physical Society (APS), v. 66, n. 2, aug 2002. Disponível em: <https://doi.org/10.1103%2Fphysreve.66.026702>. Citado na página 62.

MINDLIN, G. M.; GILMORE, R. Topological analysis and synthesis of chaotic time series. *Physica D: Nonlinear Phenomena*, v. 58, n. 1, p. 229–242, 1992. ISSN 0167-2789. Disponível em: <https://www.sciencedirect.com/science/article/pii/016727899290111Y>. Citado na página 61.

MONGODB. *MongoDB: The Developer Data Platform | MongoDB* — [mongodb.com](https://www.mongodb.com). 2022. <https://www.mongodb.com/>. [Accessed 06-Aug-2022]. Citado na página 70.

MULLER, J. *The Physiology of the Senses, Voice, and Muscular Motion, with the Mental Faculties*. Taylor, Walton & Maberly, 1848. Disponível em: <https://books.google.com.br/books?id=inuqHAAACAAJ>. Citado na página 28.

NAIK, N. Migrating from virtualization to dockerization in the cloud: Simulation and evaluation of distributed systems. In: *2016 IEEE 10th International Symposium on the Maintenance and Evolution of Service-Oriented and Cloud-Based Environments (MESOCA)*. [S.l.: s.n.], 2016. p. 1–8. ISSN 2326-6937. Citado 2 vezes nas páginas 65 e 70.

NAIK, N.; JENKINS, P. Web protocols and challenges of web latency in the web of things. In: *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*. [S.l.: s.n.], 2016. p. 845–850. Citado na página 21.

NETGURU. *Top 10 Django Apps — netguru.com*. 2022. <https://www.netguru.com/blog/top-django-apps>. [Accessed 05-Aug-2022]. Citado na página 70.

NEWMAN, S. *Building Microservices*. [S.l.: s.n.], 2022. ISBN 9781492034025. Citado 3 vezes nas páginas 22, 64 e 83.

OPPENHEIM, A. *Sinais e sistemas*. Prentice-Hall, 2010. ISBN 9788576055044. Disponível em: <https://books.google.com.br/books?id=ZOg9bwAACAAJ>. Citado na página 42.

OPPENHEIM, A. V. Superposition in a class of nonlinear systems. MIT Research Laboratory of Electronics, 1965. Citado na página 45.

OPPENHEIM, A. V. *Discrete-time signal processing*. [S.l.]: Pearson Education India, 1999. Citado 4 vezes nas páginas 32, 42, 43 e 45.

OPPENHEIM, A. v.; SCHAFER, R.; STOCKHAM, T. Nonlinear filtering of multiplied and convolved signals. *IEEE transactions on audio and electroacoustics*, IEEE, v. 16, n. 3, p. 437–466, 1968. Citado 2 vezes nas páginas 45 e 46.

PACKARD, N. H. *et al.* Geometry from a time series. *Phys. Rev. Lett.*, American Physical Society, v. 45, p. 712–716, Sep 1980. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevLett.45.712>. Citado na página 56.

PETERSON, G. E.; BARNEY, H. L. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, Acoustical Society of America (ASA), v. 24, n. 2, p. 175–184, mar 1952. Disponível em: <https://doi.org/10.1121%2F1.1906875>. Citado na página 28.

POINCARÉ, H. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, Springer Netherlands, v. 13, n. 1, p. A3–A270, 1890. Citado na página 54.

PRAAT. *Praat: doing Phonetics by Computer — fon.hum.uva.nl*. 2022. <https://www.fon.hum.uva.nl/praat/>. [Accessed 31-Jul-2022]. Citado na página 21.

QUEIROZ, M. R. G. *et al.* Terapia vocal no contexto da telefonaudiologia em pacientes disfônicos: revisão integrativa. *Audiology - Communication Research*, FapUNIFESP (SciELO), v. 26, 2021. Disponível em: <https://doi.org/10.1590%2F2317-6431-2021-2534>. Citado na página 19.

RABINER, L. R.; SCHAFER, R. W. *Digital processing of speech signals*. Upper Saddle River, NJ: Pearson, 1978. (Prentice-Hall signal processing series). Citado 3 vezes nas páginas 27, 35 e 72.

RABINER, L. R.; SCHAFER, R. W. *An Introduction to Digital Speech Processing*. Hanover, MD: now, 2007. (Foundations and Trends (R) in Signal Processing). Citado 13 vezes nas páginas 18, 29, 33, 35, 42, 43, 44, 46, 47, 49, 50, 51 e 52.

RAWALD, T. *Scalable and Efficient Analysis of Large High-Dimensional Data Sets in the Context of Recurrence Analysis*. Tese (Doktor-Ingenieur em Informatik). Disponível em: <https://edoc.hu-berlin.de/handle/18452/19518>. Citado na página 71.

RAWALD, T.; SIPS, M.; MARWAN, N. PyRQA—conducting recurrence quantification analysis on very long time series efficiently. *Computers & Geosciences*, Elsevier BV, v. 104, p. 101–108, jul 2017. Disponível em: <https://doi.org/10.1016%2Fj.cageo.2016.11.016>. Citado na página 55.

RAWALD, T.; SIPS, M.; MARWAN, N. Pyrq—conducting recurrence quantification analysis on very long time series efficiently. *Computers Geosciences*, v. 104, p. 101–108, 2017. ISSN 0098-3004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0098300416307439>. Citado na página 71.

ROSS, M. *et al.* Average magnitude difference function pitch extractor. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 22, n. 5, p. 353–362, 1974. Citado na página 34.

SAPIENZA, C. M.; RUDDY, B. H.; HOFFMAN-RUDDY, B. *Voice Disorders*. [S.l.: s.n.], 2022. ISBN 9781597567183. Citado 4 vezes nas páginas 20, 24, 25 e 26.

SAVI, M. *Dinâmica Não-linear E Caos*. [S.l.: s.n.], 2017. ISBN 9788576505549. Citado na página 56.

SAYFAN, G. *Mastering Kubernetes - Third Edition*. [S.l.: s.n.], 2020. ISBN 9781839211256. Citado na página 22.

SCHAFER, R. W. Echo removal by discrete generalized linear filtering. MIT Research Laboratory of Electronics, 1969. Citado 2 vezes nas páginas 45 e 46.

SERVILHA, E. A. M.; PENA, J. Tipificação de sintomas relacionados à voz e sua produção em professores identificados com ausência de alteração vocal na avaliação fonoaudiológica. *Revista CEFAC*, FapUNIFESP (SciELO), v. 12, n. 3, p. 454–461, apr 2010. Disponível em: <https://doi.org/10.1590%2Fs1516-18462010005000035>. Citado na página 20.

SEVERINI, R. da S. G. *et al.* Fast, cheap and feasible: Implementation of pediatric telemedicine in a public hospital during the covid-19 pandemic. *Jornal de Pediatria*, Elsevier BV, v. 98, n. 2, p. 183–189, mar 2022. Disponível em: <https://doi.org/10.1016%2Fj.jpmed.2021.05.007>. Citado na página 19.

SILVA, I. F. *Análise de em sensor piezoelétrico na aquisição de sinais de Vibração da Pele do Pescoço (VPP)*. Dissertação (Dissertação de Mestrado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, 2020. Citado na página 84.

- SIMBERG, S. *et al.* Exploring genetic and environmental effects in dysphonia: A twin study. *Journal of Speech, Language, and Hearing Research*, American Speech Language Hearing Association, v. 52, n. 1, p. 153–163, fev. 2009. Disponível em: [https://doi.org/10.1044/1092-4388\(2008/07-0095\)](https://doi.org/10.1044/1092-4388(2008/07-0095)). Citado na página 25.
- SPURLOCK, J. *Bootstrap*. [S.l.: s.n.], 2013. ISBN 9781449343910. Citado na página 70.
- STEMPLE, J. C.; ROY, N.; KLABEN, B. *Clinical Voice Pathology*. [S.l.: s.n.], 2018. ISBN 9781635500288. Citado 5 vezes nas páginas 20, 24, 25, 26 e 34.
- STEVENS, K. N.; HOUSE, A. S. Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America*, v. 27, n. 3, p. 484–493, 1955. Disponível em: <https://doi.org/10.1121/1.1907943>. Citado na página 28.
- SVEC, J. G.; GRANQVIST, S. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, American Speech Language Hearing Association, v. 19, n. 4, p. 356–368, nov 2010. Disponível em: [https://doi.org/10.1044/1058-0360\(2010-F09-0091\)](https://doi.org/10.1044/1058-0360(2010-F09-0091)). Citado na página 35.
- TAKENS, F. Detecting strange attractors in turbulence. In: RAND, D.; YOUNG, L.-S. (Ed.). *Dynamical Systems and Turbulence, Warwick 1980*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981. p. 366–381. ISBN 978-3-540-38945-3. Citado 2 vezes nas páginas 55 e 56.
- TEAGER, H. Some observations on oral air flow during phonation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 28, n. 5, p. 599–601, 1980. Citado na página 54.
- TEAGER, H.; TEAGER, S. Evidence for nonlinear sound production mechanisms in the vocal tract. *Speech production and speech modelling*, Springer, p. 241–261, 1990. Citado na página 54.
- TEAGER, H. M.; TEAGER, S. M. A phenomenological model for vowel production in the vocal tract. *Speech Science: Recent Advances*, College-Hill, p. 73–109, 1983. Citado na página 54.
- TEIXEIRA, A. B. *et al.* Otimização da terapia de reperfusão no infarto agudo do miocárdio com supradesnível do segmento ST por meio de telemedicina baseada no WhatsApp®. *Arquivos Brasileiros de Cardiologia*, Sociedade Brasileira de Cardiologia, dec 2021. Disponível em: <https://doi.org/10.36660/2Fabc.20201243>. Citado na página 19.
- THIEL, M. *et al.* Influence of observational noise on the recurrence quantification analysis. *Physica D: Nonlinear Phenomena*, v. 171, n. 3, p. 138–152, 2002. ISSN 0167-2789. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167278902005869>. Citado na página 59.
- THIEL, M.; ROMANO, M. C.; KURTHS, J. *Analytical description of Recurrence Plots of white noise and chaotic processes*. arXiv, 2003. Disponível em: <https://arxiv.org/abs/nlin/0301027>. Citado na página 59.
- THONES, J. Microservices. *IEEE Software*, Institute of Electrical and Electronics Engineers (IEEE), v. 32, n. 1, p. 116–116, jan 2015. Disponível em: <https://doi.org/10.1109/2Fms.2015.11>. Citado na página 63.
- TUTORIALSPPOINT. *Django - Overview* — [tutorialspoint.com](https://www.tutorialspoint.com/django/django_overview.htm). 2022. [https://www.tutorialspoint.com/django/django\\_overview.htm](https://www.tutorialspoint.com/django/django_overview.htm). [Accessed 05-Aug-2022]. Citado na página 70.

VIEIRA, V. J. D. *Avaliação de distúrbios da voz por meio de análise de quantificação de recorrência*. Dissertação (Dissertação de Mestrado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, 2014. Citado 7 vezes nas páginas 18, 26, 54, 57, 67, 70 e 73.

VIEIRA, V. J. D. *Análise de Variações Acústicas Não Estacionárias e seu Efeito na Detecção de Múltiplas Emoções e Condições de Estresse*. Tese (Doutorado em Engenharia Elétrica) — Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Campina Grande, Brasil, 2018. Disponível em: [https://drive.google.com/file/d/1\\_t3yDDkUem4HfVVRQutcLSeil-tYWw4T/view](https://drive.google.com/file/d/1_t3yDDkUem4HfVVRQutcLSeil-tYWw4T/view). Acesso em: 26 jul. 2022. Citado na página 18.

VIEIRA, V. J. D. *et al.* Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 28, n. 8, p. 085709, 2018. Disponível em: <https://doi.org/10.1063/1.5024948>. Citado 3 vezes nas páginas 18, 55 e 58.

VILLAMIZAR, M. *et al.* Infrastructure cost comparison of running web applications in the cloud using aws lambda and monolithic and microservice architectures. In: IEEE. *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*. [S.l.], 2016. p. 179–182. Citado na página 21.

WASEEM, M.; LIANG, P.; SHAHIN, M. A systematic mapping study on microservices architecture in DevOps. *Journal of Systems and Software*, Elsevier BV, v. 170, p. 110798, dec 2020. Disponível em: <https://doi.org/10.1016%2Fj.jss.2020.110798>. Citado 2 vezes nas páginas 22 e 63.

WEBBER, C. L.; MARWAN, N. (Ed.). *Recurrence Quantification Analysis*. Springer International Publishing, 2015. Disponível em: <https://doi.org/10.1007/978-3-319-07155-8>. Citado 6 vezes nas páginas 55, 57, 60, 61, 62 e 63.

WEBBER, C. L.; ZBILUT, J. P. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, American Physiological Society, v. 76, n. 2, p. 965–973, feb 1994. Disponível em: <https://doi.org/10.1152%2Fjappl.1994.76.2.965>. Citado na página 58.

YADAV, M. P.; PAL, N.; YADAV, D. K. A formal approach for docker container deployment. *Concurrency and computation*, Wiley Subscription Services, Inc, Hoboken, v. 33, n. 20, p. n/a, 2021. ISSN 1532-0626. Citado 2 vezes nas páginas 65 e 70.

YORKSTON, K. M.; BEUKELMAN, D. R. Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders*, American Speech Language Hearing Association, v. 46, n. 3, p. 296–301, ago. 1981. Disponível em: <https://doi.org/10.1044/jshd.4603.296>. Citado na página 66.

ZBILUT, J. P.; MARWAN, N. The wiener–khinchin theorem and recurrence quantification. *Physics Letters A*, v. 372, n. 44, p. 6622–6626, 2008. ISSN 0375-9601. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0375960108014102>. Citado na página 61.

ZBILUT, J. P.; WEBBER, C. L. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, Elsevier BV, v. 171, n. 3-4, p. 199–203, dec 1992. Disponível em: <https://doi.org/10.1016%2F0375-9601%2892%2990426-m>. Citado na página 58.



ZBILUT, J. P.; WEBBER, C. L. RECURRENCE QUANTIFICATION ANALYSIS: INTRODUCTION AND HISTORICAL CONTEXT. *International Journal of Bifurcation and Chaos*, World Scientific Pub Co Pte Lt, v. 17, n. 10, p. 3477–3481, oct 2007. Disponível em: <https://doi.org/10.1142%2Fs0218127407019238>. Citado na página 58.

ZHANG, S. *et al.* Fundamental frequency estimation combining air-conducted speech with bone-conducted speech in noisy environment. In: *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. [S.l.: s.n.], 2017. p. 244–247. Citado na página 34.

ZHANG, X. *et al.* Acoustic characteristics of normal and pathological voices analysis and recognition. In: *2019 6th International Conference on Systems and Informatics (ICSAI)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109%2Ficsai48974.2019.9010561>. Citado na página 20.

ZONG, Y. *et al.* Pitch detection using EMD-based AMDF. In: *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*. IEEE, 2013. Disponível em: <https://doi.org/10.1109%2Ficicip.2013.6568144>. Citado na página 34.